

User Traffic Characteristics Study and Network Security Implications in PWLAN

Kailash Balaji



School of Computer Science
McGill University
Montreal, Canada

November 2018

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Science.

© 2018 Kailash Balaji

Dedication

To my family for their unconditional love.

Acknowledgements

I wish to convey my gratitude, first and foremost, to my supervisor Professor Xue Liu for his complete support and guidance that has inspired me to give my best for this research. I am also thankful to my coursework professors for sharing their knowledge and expertise through lecture discussion that has immensely influenced my thesis.

My gratitude extends to the *School of Computer Science*, McGill University and *Graduate and Postdoctoral Studies*(GPS), McGill University for providing me this platform through the Masters program. I am humbled by the constant moral support and belief in my work shown by my dearest family and friends. I would like to express my most profound appreciation to Jackson Joseph-Eugene, Technical Co-ordinator, Ile Sans Fil, for providing network routers for my experiments. Lastly, my heartfelt thanks to my dearest friends Sayee and Dimple.

Abstract

Public Wireless Local Area Network(PWLAN) or public WiFi is increasingly popular in coffee shops, airports, hotels and other public areas where people can access internet services. Currently, PWLAN poses a security threat to sensitive user data such as user IDs, passwords, email IDs, etc.. Hackers who exploit the *Open* nature of a majority of public WiFi cause such threat to meet their nefarious agenda. Therefore, our research is set to examine this shortfall by studying user traffic characteristics to identify user behavior in PWLAN and carrying out network security experiments in the interest of protecting user information.

User traffic characteristics study provides insight into various network operation factors such as *Quality of Service*(QoS), performance, resource management, and network security anomalies. We find that this study helps our research to identify PWLAN user behavior. We postulate that PWLAN user behavior immensely influences public WiFi popularity. To carry out this study, we collect PWLAN user traffic in different public venues at different times of a day. We analyze the traffic flow, network packets, application, and protocol composition to identify PWLAN user behavior.

Network security plays a critical role in protecting user information. Hence we focus our research primarily on PWLAN network security. We initially carry out a security assessment of the network to understand the current security measures in place. We then construct the security trace report to document different security attack traits present in the network. We analyze communicating protocols, encryption standards, clients and web servers accessed through PWLAN to construct the security trace report. Finally, we simulate systematic *Phishing* attacks to intentionally breach confidentiality and acquire sensitive user data such as user IDs, passwords, email IDs, etc. from the network. Based on the results obtained through our research, we showcase the absolute need for PWLAN users to take appropriate measures to protect user information.

Résumé

Les réseaux locaux publics sans fil(PWLAN) ou publics WiFi sont de plus en plus populaires dans les cafés, les hôtels, les aéroports et dans les espaces publics où les gens peuvent se connecter pour accéder à des services Internet. Actuellement, les PWLAN constituent une menace sérieuse pour l'information sensible comme les noms d'utilisateur, les mots de passe, les codes de messagerie électronique et autres types de renseignements que les pirates tentent d'obtenir en exploitant la nature ouverte de tels réseaux publics. Par conséquent, notre recherche est conçue pour examiner cette lacune en étudiant les caractéristiques du trafic d'utilisateurs et pour effectuer des expériences sur la sécurité des réseaux, afin de protéger l'information des utilisateurs.

Les études sur le trafic d'utilisateurs permettent de mieux comprendre des facteurs importants comme la Qualité de Service(QoS), le comportement du réseau, les ressources du réseau et les anomalies du réseau, afin de permettre une gestion efficace des activités du réseau. Nous avons documenté cette étude sur les PWLAN en mesurant le trafic d'utilisateurs à différents moments de la journée, dans différents lieux publics. En mesurant le taux de clics ou le pourcentage de connexions d'utilisateurs à différents serveurs ou sites web via le PWLAN, nous avons tracé leur indice de popularité. En résultante, ceci nous a permis de créer un ensemble intéressant de données sur le trafic, qui nous permettent d'analyser les aspects relatifs à la sécurité de ces sites populaires qui peuvent traiter n'importe quel type d'information sensible. Pour résoudre le problème de sécurité des PWLAN, nous avons d'abord effectué une évaluation de sécurité et nous avons simulé des attaques de réseau afin de violer les propriétés de sécurité de manière intentionnelle. Ces expériences de sécurité nous ont apporté des résultats qui démontrent la gravité des menaces justifiant les contre-mesures proposés aux utilisateurs de PWLAN dans le cadre de cette étude.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research Objective	3
1.3	Thesis Contributions	3
1.4	Thesis Layout	4
2	Background	5
2.1	Wireless Networks	5
2.1.1	Wireless Local Area Network(WLAN)	6
2.1.2	Public Wireless Local Area Network(PWLAN)	7
2.1.3	IEEE 802.11 Frame Format	9
2.2	User Traffic Characteristics Study	11
2.2.1	Need For Traffic Study	12
2.2.2	Methods to Study User Traffic Characteristics	12
2.2.3	Traffic Models	14
2.3	Network Security and Privacy	15
2.3.1	Security Assessment	16
2.3.2	Security Trace Methods	20
2.3.3	Security Attacks	21
2.4	Related Work	22
3	User Traffic Characteristics Study	24
3.1	Data Collection	24
3.1.1	Methodology	25
3.2	Data Analysis	26

3.3	Flow Level Study	27
3.3.1	Connection Arrival Time	28
3.3.2	Packet Burst	28
3.4	Packet Level Study	33
3.4.1	Connection Duration	33
3.4.2	Bandwidth Utilization	33
3.5	Application and Protocol Composition	37
3.5.1	Hit Rates of Websites	37
4	Network Security Implications	41
4.1	Security Assessment	41
4.1.1	Web Traffic Composition	42
4.1.2	Encryption Scheme Analysis	43
4.1.3	Firewall Rules	44
4.2	Security Trace Report	45
4.3	Security Attack	55
4.3.1	Open WiFi versus Encrypted WiFi	56
4.3.2	Phishing	57
4.4	Mitigating Security Risks in PWLAN	61
4.4.1	Internet User Community	61
4.4.2	Software Programmers	62
4.4.3	Business and Network Providers	63
5	Conclusion and Future Work	64
5.1	Conclusion	64
5.1.1	Current Trend and Opportunity	64
5.1.2	User Traffic Characteristics Study	65
5.1.3	Network Security Implications	65
5.2	Future Work	67
	References	68

List of Figures

2.1	WLAN - Operation Modes, Source : TLDP.org	7
2.2	PWLAN - Network Design	8
2.3	IEEE 802.11 MAC Frame Format, Source : itcertnotes.com	10
3.1	DATASET A (Morning) - Connection Arrival Time	29
3.2	DATASET B (Afternoon) - Connection Arrival Time	30
3.3	DATASET C (Evening) - Connection Arrival Time	30
3.4	DATASET A (Morning) - Packet Burst	31
3.5	DATASET B (Afternoon) - Packet Burst	31
3.6	DATASET C (Evening) - Packet Burst	32
3.7	Connection Duration Graph	34
3.8	DATASET A (Morning) - Bandwidth Utilization	35
3.9	DATASET B (Afternoon) - Bandwidth Utilization	35
3.10	DATASET C (Evening) - Bandwidth Utilization	36
3.11	DATASET A (Morning) - Application Composition	38
3.12	DATASET B (Afternoon) - Application Composition	38
3.13	DATASET C (Evening) - Application Composition	39
3.14	DATASET A (Morning) - Protocol Composition	39
3.15	DATASET B (Afternoon) - Protocol Composition	40
3.16	DATASET C (Evening) - Protocol Composition	40
4.1	HTTP-POST Traffic (Unencrypted)	48
4.2	DNS query response	50
4.3	TCP Flow Graph for HTTP Client-Server Handshake	53
4.4	TCP Initial Sequence Number in HTTP Server	53

4.5	ICMP Error Message	55
4.6	HTTP Cookie	58
4.7	Phished Data	59
4.8	PWLAN - User Login Page	61

List of Tables

3.1	PWLAN User Traffic Data (Attribute - Value Pair)	25
3.2	PWLAN - Hit Rates of Websites	37
4.1	PWLAN - Web Traffic Composition	42
4.2	PWLAN - HTTP Return Status Code Group Statistics	45

List of Acronyms

AP	Access Point
DNS	Domain Name System
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
ICMP	Internet Control Message Protocol
IP	Internet Protocol
ISP	Internet Service Provider
MIM	Man-in-Middle
PWLAN	Public Wireless Local Area Network
SMTP	Simple Mail Transfer Protocol
SPF	Sender Policy Framework
SSL	Secure Socket Layer
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
WEP	Wired Equivalent Privacy
WiFi	Wireless Fidelity
WLAN	Wireless Local Area Network
WN	Wireless Node
WPA	WiFi Protected Access

Chapter 1

Introduction

Public Wireless Local Area Network(PWLAN) or public WiFi has become prevalent amongst internet users in coffee shops and other public spaces. Due to its popularity, it has attracted the attention of hackers who exploit certain network security aspect of PWLAN to obtain sensitive user data such as user IDs, passwords, email IDs, etc. from the network. We find various news articles, for example. [1, 2], on increasing popularity of public WiFi accompanied by numerous network security incidents and precautions-to-be-taken in PWLAN. Now, the call to place PWLAN network security under scrutiny to protect user information has become a priority. Through multiple sightings on the same topic, we set our course to research PWLAN user traffic and network security. In this course, we identify PWLAN user behavior and establish its relation with public WiFi popularity. We then focus primarily on network security to expose the vulnerabilities in PWLAN and understand the severity of security threat. In this chapter, we present thesis motivation, research objective, and thesis contributions. The thesis layout is furnished in the end for reader reference.

1.1 Motivation

The internet is undeniably essential in our daily life for tasks that require communication and information services. Technology business giants such as Google, Facebook, and Amazon have initiatives to spread internet access across the world. Several networking technologies can be utilized to bootstrap public internet access. For example, LCDNet [3] is a technology that facilitates internet sharing with the public. It delivers by utilizing unused network bandwidth from home internet connection or through participating *Internet*

Service Providers(ISP). *What inspires the increasing popularity of public WiFi?* By providing free WiFi in public areas, we infer that most businesses promote their product and statistically experience results of increased sales. The referred news articles [1, 4] report this trend of providing free public WiFi at businesses to promote sales. The customers or users view public WiFi as a complimentary add-on that allows them to stay connected to the internet and save cost on mobile data. We suppose that this is one of the many reasons that inspire the increasing popularity of public WiFi amongst internet users. To check e-mails, IM(*Instant Messaging*) friends from a coffee shop or while grueling through long hours of transit in airports; public WiFi has become one of the critical comfort factors. This current trend of public WiFi motivates our research to identify user behavior and expose network security vulnerabilities in PWLAN.

The internet traffic is classified into different types in-order to formulate network management strategies used for implementing *Quality of Service*(QoS) check, evaluating network performance, efficient resource management, and handling network anomalies. Researchers have created various internet traffic analytical models to aid network management. These models are extensively based on traffic characteristics study of various heterogeneous network links. Therefore, we take the first step to study PWLAN user traffic characteristics to identify user behavior.

PWLAN is termed *Open* WiFi when provisioned without a network password key phrase (WEP or WPA). The network password key ideally provides security by encrypting user traffic. The *Open* nature of the network raises several security concerns. The fact that anyone can join the network enables hackers to collect user traffic by setting up *rogue Access Points*(AP) or using network sniffer applications. Internet traffic flows unencrypted in *Open* WiFi allowing easy access to sensitive user data such as user IDs, passwords, email IDs, etc.. The referred news article [5] reports security issues in *Open* WiFi and the risk of using such networks. The fundamental reasons for providing PWLAN as *Open* WiFi are, minimizing cost spent on provisioning network security infrastructure, reducing network maintenance and support overheads and ensuring ease of use through simplification of user login procedures. These reasons may seem acceptable from a facilitator's perspective; however, it is not acceptable from an end user's perspective. Our research focuses primarily on PWLAN network security in-order to shed light on the vulnerability of sensitive data posted on such *Open* networks.

1.2 Research Objective

Our research has a two-part objective,

- Studying user traffic characteristics in PWLAN in-order to obtain relevant network operation information and identify user behavior that influences network popularity amongst internet users;
- Addressing the network security issues in *Open* WiFi by carrying out a security assessment of the network, constructing a security trace report and simulating *Phishing* attacks to intentionally breach confidentiality. This experiment demonstrates the level of security risk in *Open* WiFi. Our study is solely intended to spread awareness amongst internet users and network providers about the security vulnerabilities in PWLAN. These vulnerabilities are exploited by hackers to acquire sensitive user data such as user IDs, passwords, email IDs, etc. from the network.

1.3 Thesis Contributions

In this research, we present the results of user traffic characteristics study, and security procedures followed to expose security vulnerabilities in PWLAN. Our thesis contributions are as follows,

- We collect internet traffic from different smart devices in different public venues at different times of a day to analyze traffic flow, packet information, application, and protocol composition to sketch user traffic characteristics in PWLAN;
- We carry out a security assessment of the network by determining the percentage of web traffic, analyzing encryption schemes and studying *Firewall* rules;
- We construct a security trace report by analyzing user packets with attempted security attack traits and simulate *Phishing* attacks to breach confidentiality.

By performing these experiments, we successfully create a strong case for mitigating PWLAN security risks in the interest of protecting user information.

1.4 Thesis Layout

Chapter 2 provides background information on WLAN and PWLAN, user traffic characteristics study and network security and privacy. Chapter 3 sketches the user traffic characteristics study in PWLAN. Chapter 4 documents the security assessment of the network and reports the procedure followed for simulating *Phishing* attacks on PWLAN. Chapter 5 concludes our research and outlines future work.

Chapter 2

Background

In this chapter, we present background information on,

- WLAN and PWLAN;
- User traffic characteristics study;
- Network security and privacy on the internet;
- Related work.

2.1 Wireless Networks

Wireless networks use radio waves to establish wireless communication between network nodes that are capable of wireless connectivity. Laptop, smartphone, tablet, etc. are a few examples of network nodes or *Wireless Nodes*(WN). WLAN is one of the most common types of wireless networks that provides internet access to users at home and business. Mobile, Space, and *Wireless Wide Area Networks*(WAN) are other types of wireless networks in use. WLAN and PWLAN are tight-knit regarding their network design and operation, but, with fine-drawn differences. In the following sections, we provide the general network design and operation modes of WLAN and PWLAN. We also define the critical qualities of public WiFi imperative to our research.

2.1.1 Wireless Local Area Network(WLAN)

2.1.1.1 Network Design

Wireless Local Area Network(WLAN) is defined under the IEEE 802.11 standard and commonly termed as *WiFi*. A typical WLAN connects one or more *Wireless Nodes*(WN) to the internet through AP. AP is one of the critical components in WLAN that acts as a bridge between WN and the internet. In a typical home WiFi network, the wireless router or modem connected to the ISP by a distribution system (example. Ethernet LAN cable) acts as the internet access point. AP is also referred to as *Hotspot* installed in public venues for internet access. The key architectural components of WLAN referred in [6] are as follows,

- *Stations - Access Point*(AP) and *Wireless Nodes*(WN);
- *Basic Service Set*(BSS) - It is a set of communicating stations. Each set is identified by a unique *BSS identifier*(BSSID) that enables connection between APs across sets;
- *Extended Service Set*(ESS) - It is a set constructed by connecting two or more BSS with a communication link. A unique id called SSID identifies ESS. The SSID is the WiFi network's name that any user sees when their smart device shows a list of available WiFi networks to connect to;
- *Distribution System* - It connects APs in ESS for increased network coverage, for example, backbone satellite links.

2.1.1.2 Operation Modes

WLAN operates in two modes, *Infrastructure* mode and *Ad-hoc* mode. In infrastructure mode (typical WLAN), WNs connect to ISP for internet service through AP with a distribution system. Therefore, the BSS can communicate with each other to form ESS in this mode of operation. In ad-hoc mode, WNs communicate directly with each other to form *Peer-to-Peer*(P2P) network depending on the in-range signal strength of devices. Since there are no APs, BSS sets are independent in the ad-hoc mode of operation and communication with other BSS is not established. Fig. 2.1 shows Infrastructure mode (left) and Ad-hoc mode (right) of operation in WLAN.

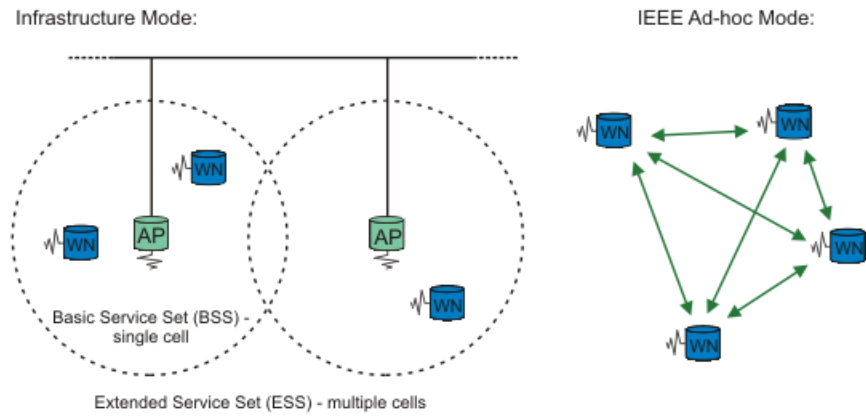


Fig. 2.1 WLAN - Operation Modes, Source : TLDP.org

2.1.2 Public Wireless Local Area Network(PWLAN)

2.1.2.1 Network Design

PWLAN is a derivative of WLAN - Infrastructure mode. It is implemented in public areas such as coffee shops, airports, hotels, etc. where people can access internet services. PWLAN corresponds to different names that are characterized by the nature of AP. WiFi network with secured AP is termed *Encrypted* WiFi and an unsecured AP is termed as public or *Open* WiFi. The provisioning of a network password key phrase (WEP or WPA) to secure an AP depends on the network provider's payment model for PWLAN. PWLAN network is designed based on four types of payment model referred in [7],

- *Subscription Model* - Users pay and subscribe in advance for network access;
- *On Site Model* - Users pay on site, example. people in the airport trying to connect to available WiFi networks;
- *User Fairness Model* - The payment for using the network depends on the network traffic load;
- *Open Fi* - Free WiFi service.

Since the first three payment model requires the users to pay for network access, the AP is usually secured with a WEP or WPA network password key phrase for encrypting

user traffic. Unfortunately, majority of the public WiFi that are based on *Open* Fi model have unsecured or *Open* AP. Network providers use *Open* Fi model to avoid network maintenance, reduce security and support overhead, promote sales, enable easy login, etc.. In consequence, user traffic in public WiFi flows unencrypted or in plain-text form. Other users on the same public WiFi network can see traffic information sent over the network and anyone can join the network. Most public *Hotspots* use *wireless user isolation* for providing security. We define *wireless user isolation* in the next section - Key Qualities of Public WiFi. Fig. 2.2 shows the generic network design of PWLAN.

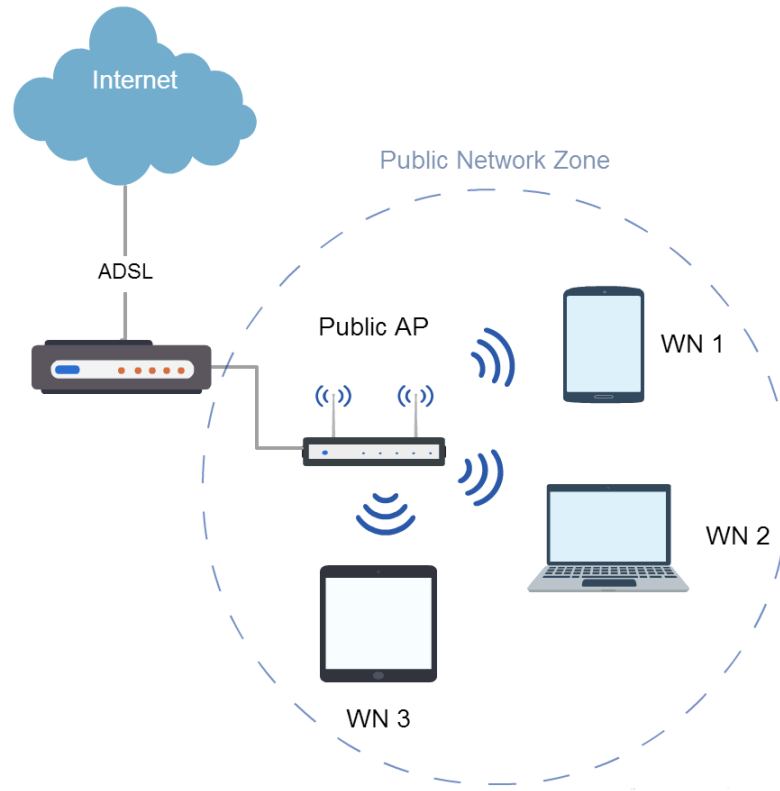


Fig. 2.2 PWLAN - Network Design

2.1.2.2 Key Qualities of Public WiFi

Public WiFi network available in public areas such as coffee shops, airports, hotels, etc. have the following key qualities,

- Any user can connect to the network since public WiFi has *Open* AP;
- Internet traffic flows unencrypted or in plain-text form, unless, the accessed website uses HTTPS standard for securing user connections or use of VPN;
- Network is vulnerable to sniffer applications, for example. Wireshark. It is a network monitor application used to capture user data packets;
- Most public WiFi use *wireless user isolation*. *Wireless user isolation* refers to the isolation of each user connected to the *hotspot* in-order to prevent hackers from getting into user devices. Most APs support *wireless user isolation* in their router settings;
- Hackers capture user traffic by creating rogue APs and using sniffer applications after joining the network as a legitimate user;
- Marketing public WiFi is possible through payment models and networking technologies (example. LCDNet);
- Growing acceptance rate due to increasing smartphone usage, saving cost on mobile data plans and user behavioral interests in *Web browsing, Online Social Networking(OSN), E-mail, Video Sharing, etc..*

2.1.3 IEEE 802.11 Frame Format

The 802.11 frame format referred in [8] applies to both WLAN and PWLAN network packets. The only difference in PWLAN is the flag bit set in the *Protected* field. If the network is provisioned without WEP or WPA network password key phrase, then it is set to bit "0" indicating data not protected. It is set to bit "1" only if users connect to HTTPS supported websites. In WLAN, the *Protected* field is set to bit "1" by default. The packet frame is classified into three types, control, data, and the management frame. The control frame is set to 30 bytes of information and serves as the MAC header. Fig. 2.3 shows the IEEE 802.11 MAC frame format.

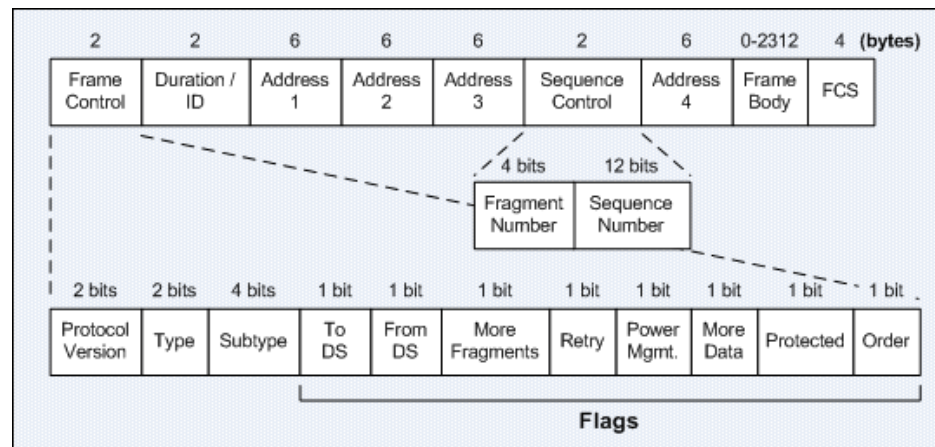


Fig. 2.3 IEEE 802.11 MAC Frame Format, Source : itcertnotes.com

In the MAC header,

Frame Control field contains the following,

- *Protocol Version* - indicates the version of 802.11 used for the transmission;
- *Type and Subtype* - the type of frame (control, data or management) for which the given subtype does a specific action;
- *To DS or From DS* - indicates if the packet is entering or exiting from a *Distribution System*(DS);
- *More Fragments* - indicates if there are frames to follow a fragmented frame;
- *Retry* - indicates if the frame will be re-transmitted following a transmission error;
- *Power Management* - indicates if the sending station is in *Power Save*(PS) mode of operation;
- *More Data* - signals the receiving station if the sending station has more frames to send;
- *Protected* - indicates if a password is used for encryption and authentication;
- *Order* - indicates if the frame is received in sent "order".

Duration field - indicates the duration till transmission of the next frame;

Address field - The address field includes the following addresses,

- *BSSID* - Unique identifier for the *Base Service Set*(BSS);
- *Destination Address* - The end receiver address;
- *Source Address*- The original sender address;
- *Receiver Address* - The address of the next station in the network path;
- *Transmitter Address* - The address of the current station in the network path that is ready to send the frame.

Sequence Control field contains the following,

- *Sequence Number* - indicates the sequence number of the frame;
- *Fragment Number* - If the frame is fragmented, this field indicates the fragment number of each subframe. Note: Sequence number of each frame in a fragmented frame remains the same until the next new frame is received.

Frame Body field - contains the information of the data or management frame (max. of 2312 bytes);

FCS field - *Frame Check Sum*(FCS) is a 4 byte field that contains check sum bits for checking data consistency and the presence of transmission errors.

2.2 User Traffic Characteristics Study

Internet traffic is continuously changing with the advent of new web applications. To plan and manage any network it is essential to analyze traffic that flows through it. In the following sections, we describe the need for the traffic study, techniques to determine traffic characteristics that we have adopted in our research and finally, we outline common traffic models created based on characterization studies.

2.2.1 Need For Traffic Study

Traffic Study of any network link is vital for network management purposes. Essentially for the following,

- *Checking Quality of Service(QoS) guarantees* - For example, The paper [9] emphasizes on checking QoS guarantees for prioritized internet traffic with factors such as low transmission delay, increased bandwidth, etc.;
- *Network performance evaluation* - For example, network performance evaluation based on the routing process;
- *Network resource management* - Since resource management policies vary with the type of internet traffic, traffic characteristics study helps operators to prioritize resources and emphasize management policies for different traffic types. For example, providing buffer allocation for resource sensitive traffic such as video conferencing, web browsing, etc.;
- *Handling network anomalies* - Intrusion detection, i.e., taking countermeasures against network anomalies such as *Denial of service*(DoS), spam traffic, packet floods, etc..

2.2.2 Methods to Study User Traffic Characteristics

Traffic characterization is carried out for three classes of internet traffic referred in [10],

- *Sensitive Traffic* - This type of traffic is sensitive to time, security measures, etc.. It has priority and must be served with necessary QoS guarantees and resource allocation. For example, VoIP calls, video conferencing and web browsing traffic;
- *Best Effort Traffic* - Network operators view this type of traffic as normal traffic that does not need prioritizing for QoS checks. It gets processed after the *sensitive traffic* class. For example, P2P traffic and email applications traffic;
- *Undesired Traffic* - This type of traffic is network anomaly, example, TCP - Syn packet floods that cause *Denial of Service*(DoS), spam traffic, etc.. Network operators block them or suppress their activities through security policies.

Given the traffic classes that require network management support, we now describe the methods to study user traffic characteristics. User traffic characteristics study helps network operators to formulate network and security policies.

Flow Level Study

Traffic flow level study is essential to address network bottlenecks such as speed limit and congestion control. The connection arrival times and packet bursts are analyzed to identify significant and short transfers. The author of [11] presents the importance of traffic flow level study. In this paper, network management techniques for congestion control and increased network throughput are developed based on flow level studies that capture traffic flow for a given distribution model (Poisson model) at an arrival process. Also, the developed traffic model assumes that the bandwidth is allocated dynamically for the traffic flow based on the total available bandwidth in the network path.

Packet Level Study

The packet level approach to study any network is more accurate than any other method. It is accurate because traffic propagates as packets after injecting into the network. We analyze the user connection duration and bandwidth utilization depending on user application behavior. As presented in [12], we understand that the core network elements such as routers, switches, and APs operate on a packet by packet basis, thus, making them a critical part of traffic characterization technique. The packet level study provides useful insight into effective resource management, determining network propagation delays for traffic flow and detection of network anomalies.

Application and Protocol Composition

It is important to note that application and protocol composition in internet traffic provides essential background on user behavior on the internet. We profile user application traffic (for example, *web browsing*, *email*, *online gaming*, *video streaming*, *VoIP*, *etc.*) that triggers necessary transport layer protocols (for example, TCP and UDP) to connect with respective servers. We use flow and packet level approach on top of this method to study PWLAN user traffic characteristics.

2.2.3 Traffic Models

Traffic characterization in any network is a step to help researchers develop traffic models based on mathematical or statistical approach. The objective of traffic modeling is to take either traffic flow level or packet level (sometimes both) characteristics and provide network operators with a useful framework for network management. Thorough understanding of traffic characteristics enables network engineers to design networks out of an experience and make predictions of performance on future requirements, as stated in [13]. Traffic modeling is also helpful in traffic generation based on different mathematical distribution models. The decision to choose a model depends on the type of network. We outline a few essential models such as the Poisson model, the Markov model, the Pareto model, and the Weibull model. Poisson and Markov methods model Non-Self-Similar traffic, whereas the Pareto and Weibull methods model Self-Similar traffic.

Self-Similar and Non-Self-Similar Traffic

Internet traffic is characterized as self-similar or non-self-similar depending upon the occurrence and nature of its burstiness. Burstiness is the peak value reached at a short time interval (example. consider a network download link at an average speed of 50 Mbit/s reaching a short-lived peak of 100 Mbit/s). If the traffic has burstiness at all times, then it is self-similar and if not, non-self-similar.

Poisson Distribution Model

The Poisson distribution model is used for modeling non-self-similar traffic generated in networks with different traffic streams, referred in [14]. The traffic flow at the arrival process is modeled based on the Poisson distribution function. For example, modeling incoming packets of VoIP calls.

Markov Model

The traffic modeling of non-self-similar traffic is done based on the Markov property which states that a behavior of a future state depends only on the present state and not the sequence of states that preceded it. Markov method tries to model the activities at the traffic source by valuing only its current state. For example, traffic flow activity.

Pareto Distribution Model

The Pareto distribution function is used to model long tail traffic (traffic characterized by large file size distribution). This model helps network operators to keep the load factor in check by implementing necessary load balancing policies.

Weibull Distribution Model

Weibull model is applied only over a fixed rate traffic flow in long tail traffic (self-similar) generated by a source that works on ON/OFF model, presented in [15]. In ON/OFF model, a source is considered ON state if it generates traffic flow in the network. The source is in OFF state when it has no traffic flow to generate.

2.3 Network Security and Privacy

Generally, security is defined by three core properties - Confidentiality, Integrity, and Availability. Security solutions and technologies exist (and continue to develop) to protect and maintain these properties.

Confidentiality

Confidentiality is built on the "need-to-know" principle that states that any sensitive data is restricted for use at all times unless the subject (user or computer program) is authorized to access them. In network security, encryption is a standard technique to protect the confidentiality of data and only an authorized program with the cryptographic key can access the data.

Integrity

Integrity is the property that defends user data against unauthorized change or modification meant for corrupting the data. Example, computing checksum to detect data corruption for a frame transmission is one of the methods for integrity check.

Availability

Availability property makes sure that a system or data is available to the authorized user at any point in time. In any network, *Denial of Service*(DoS) attack disrupts services. Traffic classification techniques help to detect and suppress them to protect availability property.

User Privacy

Internet user privacy accounts for methods and policies implemented for storing, distributing and displaying user information on the web. In recent times, user privacy in the web is declining at an alarming rate due to the breach of security properties and unauthorized distribution of user information to third parties. The most common threat to user privacy is the website or HTTP cookies. Cookies are text files containing name-value pair sent by a website (when accessed) to the user system. This name-value pair is usually a unique id generated for the user, and the value is the website's name. Certain websites add additional information such as session information and time of access to the cookies file. Cookies are used for tracking user activities so that websites (in business) can promote their product based on user preferences. The distribution of user information to unauthorized third party is a common threat resulting in *Spam* e-mails or *Rogue* advertisement pop-ups while web browsing. Cookies are mainly of two types, HTTP cookies (as defined earlier in this section) and flash cookies. Flash cookies are similar to HTTP cookies, except it uses flash content (for example, advertisement videos) for promoting business.

2.3.1 Security Assessment

The assessment of existing security elements of any network serves as an initial step for developing security solutions and formulating policies. We prepare a security assessment profile for PWLAN by,

- Determining the percentage of web traffic (HTTP/HTTPS) with the help of user traffic characteristics study;
- Analysing encryption schemes used by the network;
- Studying Firewall rules;
- Tracing packets with security attack traits.

The author of [16] proposed traffic analysis on the data link layer to infer users' online activities. In our research, we conduct security experiments in the IP and Application layer. In the following sections, we provide a brief description about HTTP and HTTPS protocol (web traffic), network encryption schemes - WEP and WPA (also termed as network password key phrase), firewall and security trace methods.

2.3.1.1 Hypertext Transfer Protocol (HTTP)

HTTP is one of the most popular application layer protocol designed for websites hosted by a web server. HTTP implements a request-response messaging format for client-server communication. The request message is sent through *Uniform Resource Locator*(URL) that typically consists of HTTP request methods such as,

- *GET* - Requests data from the server by locating the resource through URL;
- *POST* - Posts data on the internet with the help of *Uniform Resource Identifier*(URI) to identify resources (for example, posting user data through web forms). In this type of request method, the server treats the posted data as a new resource on the web;
- *HEAD* - Works as HTTP GET method, but it is only used to request meta-data (for example, content length, encoding technique, etc.);
- *PUT* - Works as HTTP POST method but if a resource already exists in the server, then PUT modifies it by updating with the new resource provided through URI.

HTTP generally uses TCP (Port 80) as its default transport layer protocol since it offers reliable communication. HTTP over any other transport layer protocol that provides reliable communication is possible. It is important to note that HTTP traffic flow is *plain-text* (unencrypted) and therefore it depends on the network's encryption scheme (WEP or WPA) for providing security to client-server communication.

2.3.1.2 Hypertext Transfer Protocol Secure (HTTPS)

HTTPS is the implementation of HTTP over SSL/TLS protocol to provide secured client-server communication. The following points describe a few important HTTPS characteristics,

- HTTPS uses asymmetric cryptography for providing secured client-server communication. The client encrypts user traffic with the public key of the server (this process also verifies the client through its digital signature), and the server uses the client's secret key to decrypt the user traffic;
- HTTPS uses security certificates (for example, X.509) to verify the authenticity of the hosting server and access control in-order to restrict unauthorized client traffic;
- HTTPS generally uses TCP as its default transport layer protocol, and it operates on TCP Port - 443;
- HTTPS is supported by websites that converse and handle sensitive user data (for example, online payment or banking websites that handle user bank account information).

Since HTTPS is the secure version of HTTP, the underlying protocol operation remains the same.

2.3.1.3 Encryption Schemes

WLAN network uses encryption scheme for two primary purposes, protecting user data by encryption algorithms and providing user authentication. WEP and WPA are two well-known security standards that provide network encryption.

Wired Equivalent Privacy (WEP)

WEP is the security standard that uses RC4 encryption algorithm for data encryption and CRC-32 (*Cyclic Redundancy Code*) checksum algorithm for providing data consistency. The following steps describe the WEP encryption referred in [17],

- In the first step, WEP computes a checksum value (4 bytes) for the user data to protect integrity property;
- For RC4 encryption, WEP creates the RC4 key (64 bits) by appending the base key (40 bits) and *Initialization Vector*(IV) of 24 bits. This RC4 key encrypts the data unit from step 1.

Wi-Fi Protected Access (WPA)

WPA uses message integrity checks to determine packet disruption by any unauthorized attacker between AP and the authorized user. WPA also uses *Temporal Key Integrity Protocol*(TKIP) that takes the RC4 cipher (of WEP) and adds a 128-bit key generated for each packet. Therefore the resistance to security attack is increased by using 256 bit key in WPA.

WPA2, the successor of WPA is the most common encryption scheme used in home and business WiFi. It offers stronger encryption and authentication than WEP or WPA. It uses *Advanced Encryption Standard*(AES) encryption algorithm along with CCMP. *Counter Mode Cipher Block Chaining Message Authentication Code Protocol*(CCMP) referred in [18] replaces TKIP (of WPA) as the new encryption protocol. The following steps briefly describe WPA2 encryption using CCMP,

- First, the data units are extracted from the 802.11 frame and *Message Integrity Check*(MIC) value is computed for keeping track of data consistency;
- In the next step, the data units and MIC are encrypted to form the cipher unit with CCMP 256-bit key to append to the CCMP header;
- Finally the cipher unit along with CCMP header is appended to the MAC header and prepared for transmission.

2.3.1.4 Firewall

Firewall is a software program that blocks unwanted traffic from public internet entering into a private network. It works on rules established for constructing filters to block traffic. The following methods are used by the firewall to inspect inbound and outbound traffic,

- *Packet Filtering* - In this method, packets are checked against filters (rules). Firewall decides to either allow or block traffic from reaching the node in the private network. This method is commonly used to filter spam e-mail traffic;
- *Proxy Service* - Firewall acts as the proxy recipient of traffic from the public internet;
- *Stateful Inspection* - This method is commonly used to detect anomaly packets. Instead of full packet analysis, only certain key components are checked against already

documented information. If the firewall detects a pattern or signature, then it takes the appropriate traffic handling procedure to either allow or block traffic.

Firewall filters are created either by specifying port numbers to block or by defining rules with words or phrases. In the former method, any traffic on a specified port (say, FTP port 20) is blocked. In the latter method, words or phrases are used to inspect traffic. For example, if a firewall rule is defined by specifying a word - "XYZ" then firewall inspects only the traffic that contains the exact word. The firewall allows users to configure all major protocols such as HTTP, FTP, TCP, IP, Telnet, etc. with filters or rules to check their traffic.

2.3.2 Security Trace Methods

Security trace methods are useful in classifying packets with attempted attack traits. These packets may contain sensitive user data that interests hackers. By analyzing certain key elements (as discussed below), we can infer the possible scheme developed for attacking user traffic from the hacker's perspective. In our research study, we observe traces of attempted attacks from,

- *HTTP Return Codes* - In HTTP client-server communication, the web server acknowledges the client with a return status code to inform about the status of the connection. The return status code is classified into five general categories. Any code that belongs to 1xx (one hundred) group is informational. Any code that belongs to 2xx (two hundred) group indicates successful connections. Any code in 3xx (three hundred) group is used for redirection. Any 4xx (four hundred) code group is an error code on the client side, and any 5xx (five hundred) code group is an error code on the server side. We document the return status codes in our collected traffic to observe packets with error response that might indicate an attack trait;
- *Unencrypted User Credentials* - HTTP, telnet, FTP, POP3, SMTP, IMAP are some of the well-known protocols without an encryption algorithm defined in their protocol suite. They solely rely on the network encryption scheme (WEP or WPA) to provide security for the user traffic. We analyze packets with these protocols triggered in public WiFi to find unencrypted user information;

- *Web Servers* - By analyzing the web servers that are accessed by user requests, we document known security attack possibilities and understand the hacker's motive;
- *Source Port Randomization* - We observe the source port on DNS queries and check the randomization factor in-order to determine resistance against *Domain Name Space*(DNS) spoofing. DNS spoofing is a technique to attack DNS tables to return malicious IP addresses upon querying;
- By analyzing TCP sequence numbers, we observe traces of sequence number prediction;
- By analyzing ICMP *echo* packets, we expose the network vulnerabilities to conduct reconnaissance, *Denial of Service*(DoS) and covert channel attacks.

2.3.3 Security Attacks

In recent times, we have observed various instances of security attacks ranging from targeted individuals to corporate companies and government offices [19, 20, 21]. From these frequent occurrences, we can infer that user data, user identities, and confidential information is not completely secure on the web. Often, we find that security attacks orchestrated by hackers have an objective of either gaining access control to a system or exploiting software vulnerabilities. The following factors influence the success of any security attack,

- Lack of educational awareness amongst internet users;
- Insecure coding practices while developing software;
- Negligible or poor implementation of security policies.

To demonstrate the level of security risk in public WiFi, we carry out the most common security attack known as *Phishing*.

Phishing

Phishing is classified as one of the most common security attacks where hackers try to acquire sensitive user data by employing both social engineering and technical subterfuge, as presented in [22]. Electronic mails that masquerade a trusted entity are often used as

one of the *phishing* techniques through social engineering. In technical subterfuge, human-engineered computer virus programs and malware are used to acquire sensitive user data such as user IDs, passwords, email IDs, etc.. In our research we carry out two main *phishing* attacks,

- *Man-in-Middle Phishing(MIM)* - In a typical client-server communication, the hacker strategically operates in the middle to intercept communication between both entities. Eventually, the hacker can impersonate as any one of them to acquire sensitive user data;
- *E-mail Phishing* - In this attack, a visually deceptive email is sent to the user(s) from a *phishing* web host that looks identical to a trusted website (for example, *Online Social Networking(OSN)* websites, banking, shopping, travel websites, etc.). Once the user believes the email, then the person is *phished* to give up sensitive user data.

2.4 Related Work

Traffic characterization in any network is a step to help researchers develop traffic models based on mathematical or statistical approach. The objective of traffic modeling is to take either traffic flow level or packet level (sometimes both) characteristics and provide network operators with a useful framework for network management. Thorough understanding of traffic characteristics enables network engineers to design networks out of the experience and make predictions of performance on future requirements, as stated in [13]. The author of [23] took the first step towards modeling internet traffic in public WiFi and opened a new research direction to track changes in traffic patterns and network intrusion. The objective of this research was to develop a common modeling framework for the number of simultaneously present customers in public WiFi venues. The author of [24] sketched specific user behavior in PWLAN by capturing bursty traffic. By analyzing large and short transfers (packet burst), user behavior analysis in PWLAN was initiated. The paper [25] carried out user traffic characterization by combining user traffic from different public venues. In our research, we analyze consolidated traffic data collected from different devices in different public venues at different times of a day. Our study aims to establish a common framework for modeling all types of user traffic in PWLAN. We characterize PWLAN traffic with the same network factors as [23] and validate results from the collected user traffic.

We believe that *User traffic characteristics study* not only helps to obtain valuable insight into factors that influence PWLAN popularity but also facilitates in taking a step further into exploring network security aspect of *Public* or *Open* WiFi.

Network security solutions and technologies exist to protect and maintain confidentiality, integrity, and availability of data and system. It is to be noted that there is no "one size fits all" solution to secure a network. The balancing of acceptable risks and countermeasures to mitigate those risks is the challenge to provide security in any network [26]. We find that PWLAN suffers from security vulnerabilities and threats due to improper deployment strategies. Hackers study these technical loopholes along with certain key factors to program a scheme to acquire sensitive user data. The paper [27] analyzed security attack strategies to document the type of strategy that works and the reasons for its success. In this research, the key inference was that hackers take advantage of user naivety and employ visual deception techniques to acquire sensitive user data. The author of [28] conducted a case study on *phishing* attacks and inferred that significant changes in the internet are required to tackle such attacks. The paper sighted a few countermeasures for *phishing* in-order to protect user information. The web browser tool developed in [29] aimed at offering novel solution for *phishing* attacks. The idea of the tool was to display a warning message whenever users tried to submit sensitive user data such as user IDs, passwords, e-mail IDs, etc. to an untrusted website. The tool used a server certificate validation procedure to determine trustworthiness of a website. In our research, we constantly refer to the papers mentioned above in our security experiments with *phishing* attacks in-order to demonstrate the security risk in PWLAN. Combining both *User Traffic Characteristics Study and Security Experiments* in PWLAN, we prove the need for research and the opportunities to acquire new knowledge in this area.

Chapter 3

User Traffic Characteristics Study

In this chapter, we present the results of PWLAN user traffic characteristics study according to the following steps,

- Data collection;
- Data analysis;
- Flow level study;
- Packet level study;
- Application and protocol composition;
- Hit rates of popular websites.

3.1 Data Collection

Traffic data collection is the first and foremost step in PWLAN user traffic characteristics study. This step is critical as our study is based on a live user traffic dataset generated in different public areas. We collect user traffic from different devices in different public venues at different times of a day. We use Wireshark, a network protocol analyzer to accomplish data collection from laptops, and tablets connected to public WiFi in the coffee shop, public library, and hotel. We tag the collected traffic datasets as follows,

- *Dataset A* - Represents traffic dataset collected in the Morning, 10:00 - 13:00;
- *Dataset B* - Represents traffic dataset collected in the Afternoon, 13:00 - 16:00;
- *Dataset C* - Represents traffic dataset collected in the Evening, 16:00 - 19:00.

The following table (Table. 3.1) tabulates the PWLAN user traffic data attributes and respective values.

Attribute	Value
Venue	Coffee Shop Public Library Hotel
Trace Duration	4 Weeks
Time of Day	Dataset A - Morning Dataset B - Afternoon Dataset C - Evening
WiFi Capable Devices	Laptop, Smart Phone, Tablet
Software	Wireshark Network Analyser
Total No. of Connections	Dataset A - 262,809 Dataset B - 178,061 Dataset B - 316,659

Table 3.1 PWLAN User Traffic Data (Attribute - Value Pair)

3.1.1 Methodology

The methodology used for the data collection process is as follows,

- *Joining PWLAN* - We connect our laptop to public WiFi in the targeted coffee shop, public library, and hotel;
- *Traffic collection* - We choose the right network interface and run Wireshark to collect live user traffic;

- *Packet filtering* - We filter user data packets by targeting connected user IPs. The filter phrases are presented in the data analysis section.

3.2 Data Analysis

We analyze traffic datasets using Wireshark extensive functions available under *Statistics* menu such as I/O graph, TCP stream, IP endpoints, packet filter phrases, and HTTP request-response statistics [30].

I/O Graph

The "IO" graphs window provides configurable graphs of the collected user traffic. It allows up to five graphs created with specific packet filter phrases, the time interval in X-axis and units in the Y-axis. For example, we sketch the bandwidth utilization in packet level study by providing the filter phrase - "tcp.port==80" to measure HTTP bandwidth, time of day in X-axis and bandwidth in byte per second units in Y-axis.

TCP Stream

The "Follow TCP Stream" function in Wireshark offers comprehensive details about a selected TCP packet. For example, we identify unencrypted user credentials by filtering "http.post" TCP packets and using the "Follow TCP Stream" from Wireshark tools menu to get a detailed pop-up dialogue that identifies sensitive user data.

IP Endpoints

The "Endpoints" window in Wireshark generates a detailed dialogue of communicating client-server endpoints with IP and transport layer information. For example, we calculate the hit rates of visited websites by analyzing the IP endpoints to resolve and locate web servers.

Packet Filtering

Packet filtering is the most important function in Wireshark. It allows users to filter network packets for focused analysis with set phrases or rules. The packet filter is classified

into two types *capture filter* and *display filter*. The capture filter phrases are used to reduce the size of *live* traffic in-order to save disk space. The display filter phrases are used *after* collecting traffic to display specific packets. For example, we use the display filter:

```
http.request.method == "POST" || http.request.method == "PUT"
```

to filter HTTP POST or PUT packets to analyze resources posted on the web server after traffic data collection. We use the capture filter - "ip" and "port 53 and not arp" to capture IP and DNS traffic and filter out ARP packets.

HTTP Request-Response Statistics

The "HTTP - Statistics" window in Wireshark provides detailed statistics on HTTP request-response messages. For example, we use this option to tabulate HTTP user connections, HTTP Return Status Code Group statistics and identify web servers (refer to chapter 4).

3.3 Flow Level Study

Traffic flow level study is essential to address network bottlenecks such as speed limit and congestion control. The connection arrival times and packet bursts are analyzed to identify significant and short transfers. The author of [24] took the first step to sketch specific user behavior in PWLAN by capturing bursty traffic. The paper [25] provides user traffic characterization by combining all different public venues and the author of [23] attempts to model large-scale WiFi traffic in *Hotspots*. In our research, we analyze user traffic data collected from different device types in different public venues and consolidate with respect to different times of a day. Our study aims to establish a common framework for modeling all types of user traffic in PWLAN. The following graphs illustrate the traffic flow level study,

- *Connection Arrival Time* - This graph is with respect to time of day;
- *Packet Burst* - This graph is with respect to packet flow at inter-arrival time.

3.3.1 Connection Arrival Time

Traffic flow is initiated at the arrival time of user connections in PWLAN. We observe that the connection arrival time at different times of a day has variations with respect to user activity. Each day our trace collection starts at 10:00 and ends at 19:00, thus covering the significant part of a day. In DATASET A (Fig. 3.1), average connection arrival time is close to 2 seconds for the increased arrival rate that reflects the number of users entering a coffee shop to start their day. In DATASET B (Fig. 3.2), average connection arrival time increased to 5 seconds given the slight decrease in the arrival rate of users during lunch time. In DATASET C (Fig. 3.3), we observe average arrival time close to 1 second given the maximum arrival rate of users returning from work.

3.3.2 Packet Burst

Packet burst is the occurrence of a consecutive group of packets with shorter inter-arrival time. This packet burst represents the TCP flow of short or long transfers and has a considerable influence in characterizing network architecture. In our study, we observe that the length of the burst is maximum in the morning (Fig. 3.4) and evening (Fig. 3.6) due to maximum user activity at those times of a day. In the afternoon, we observe reduced packet bursts (Fig. 3.5). Packet bursts reflect the characterization of packet stream in which the number of packets in the burst (minimum 2), length of the burst and size of the burst determine the traffic flow.

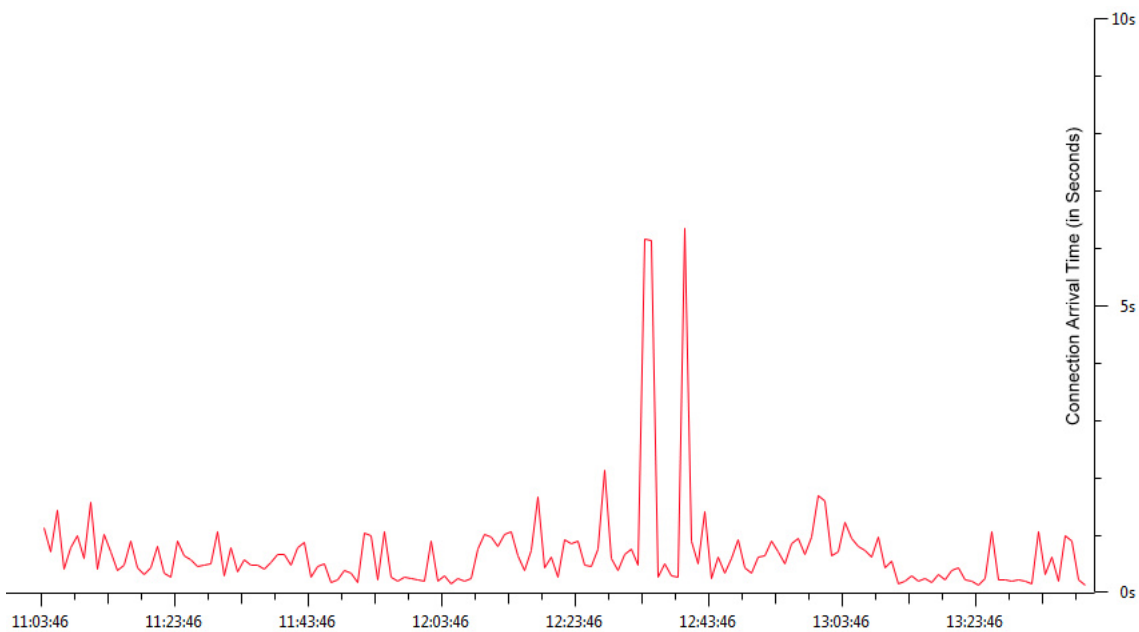


Fig. 3.1 DATASET A (Morning) - Connection Arrival Time

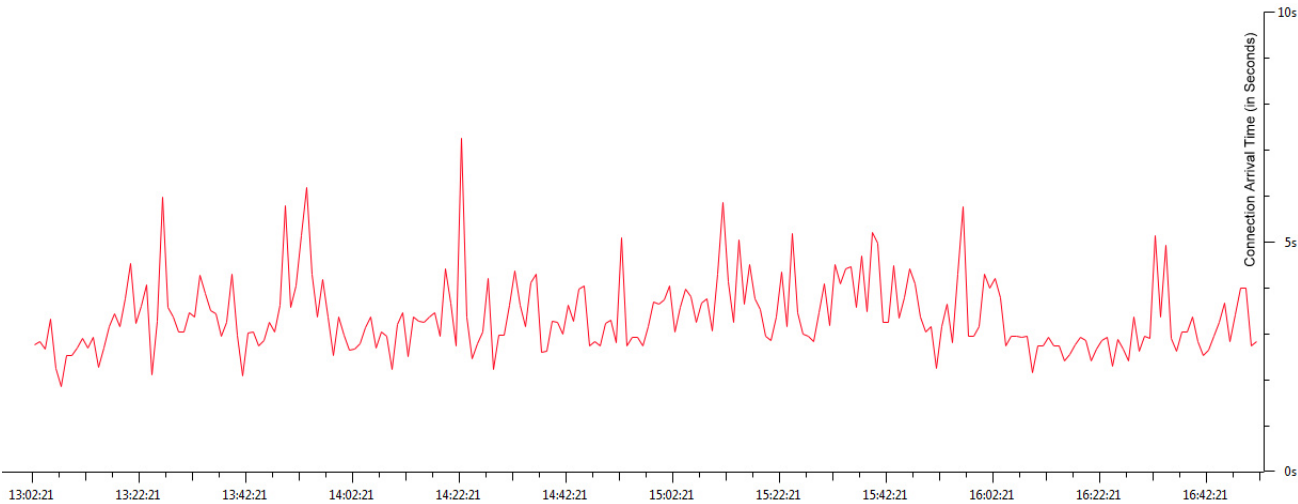


Fig. 3.2 DATASET B (Afternoon) - Connection Arrival Time

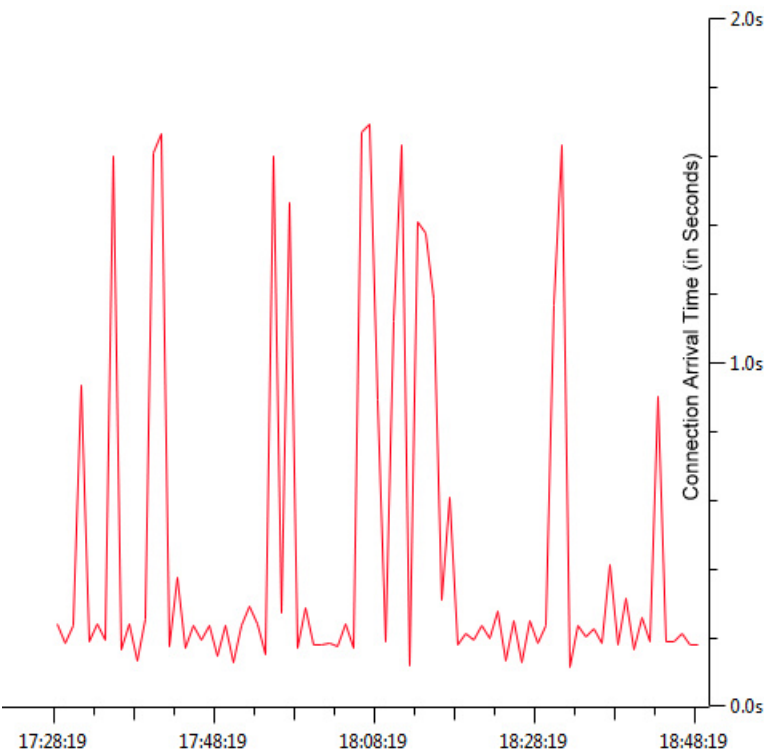


Fig. 3.3 DATASET C (Evening) - Connection Arrival Time

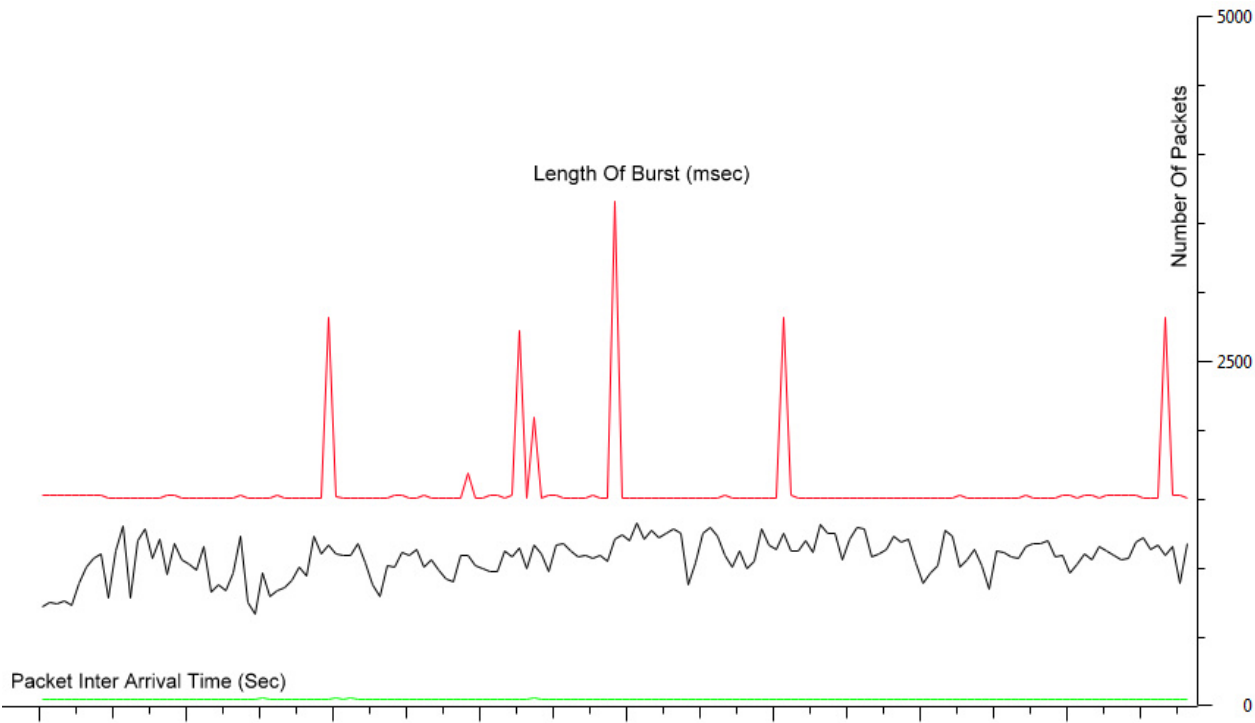


Fig. 3.4 DATASET A (Morning) - Packet Burst

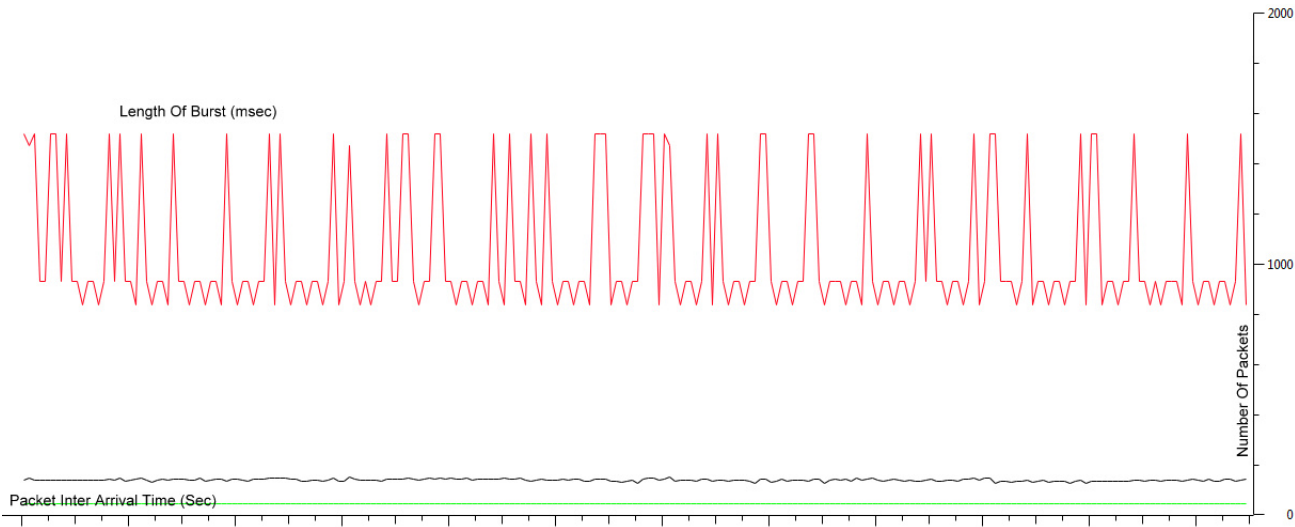


Fig. 3.5 DATASET B (Afternoon) - Packet Burst

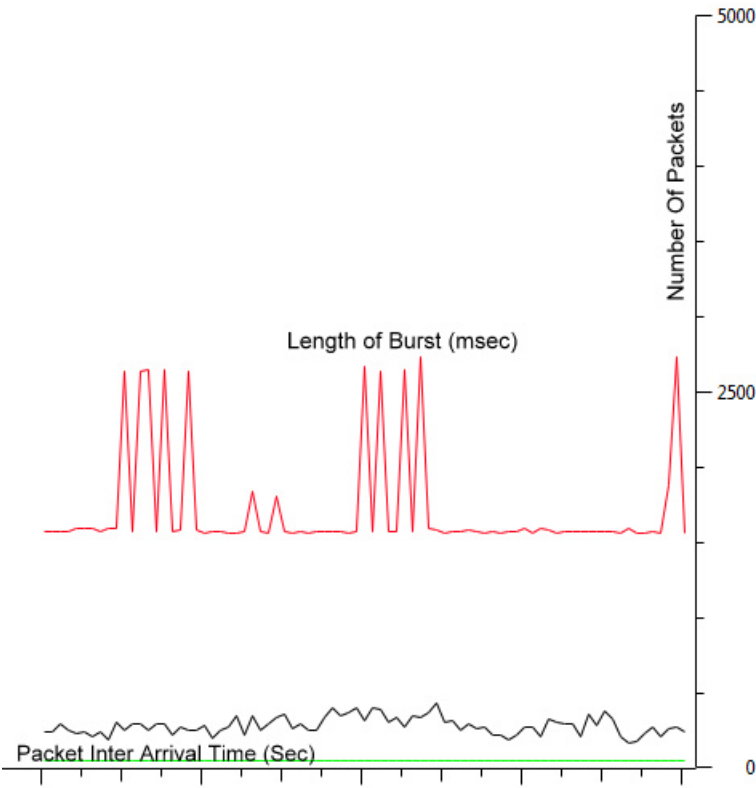


Fig. 3.6 DATASET C (Evening) - Packet Burst

3.4 Packet Level Study

The packet level approach to study any network is more accurate than any other approach. It is accurate because traffic propagates as packets after injecting into the network. We analyze user connection duration and bandwidth utilization depending on user application behavior that provides useful insight into PWLAN network operation. The extension of this study is illustrated in determining application and protocol composition. The following graphs (Fig. 3.7 - 3.10) illustrate the packet level study,

- *Connection Duration* - The graph of connected *Wireless Nodes*(WN) with respect to connection time;
- *Bandwidth Utilization* - This graph is with respect to PWLAN network bandwidth.

3.4.1 Connection Duration

A user connection is defined by association and disassociation of user device with an AP. We collect traffic from different devices such as laptops, smartphones, and tablets. The collection of such devices is Wireless Nodes (WN). The connection duration graph (Fig. 3.7) presented in this report is a sample set generated at a hotel that provides *Open* WiFi. At a given point in time (especially in the evening), approximately 90 Wireless Nodes (WN) associated with the *Open* AP. The peak connection duration observed is 3 hours with an average of 1.5 hours indicating the expected result, since, guests staying at the hotel log in to the network to access internet services in the evening. From a network security perspective, this connection analysis enables hackers to identify "long-lasting" targets. Long-lasting targets are users who have an active connection to the AP for an average of 20 minutes.

3.4.2 Bandwidth Utilization

Bandwidth defines connected network's potential to facilitate fast internet services through data upload and download speeds. Our study documents the user traffic behavior at different times of a day against network bandwidth in-order to determine utilization peaks. User traffic behavior helps network providers to manage network resources and serve user needs. The bandwidth utilization graph presented in this report accounts for 10 Mbps

network link provided at a coffee shop. DATASET A (Fig. 3.8) averages 200 Kbps that indicates users have enough bandwidth for web browsing. DATASET B (Fig. 3.9) averages the same bandwidth utilization as DATASET A that confirms web browsing as the popular application. In DATASET C (Fig. 3.10), we observe peak values ranging from 500 Kbps to 600 Kbps that indicates the presence of other application types. For example, video calls. Theoretically, Skype video calls are possible at such upstream bandwidth with reduced video quality.

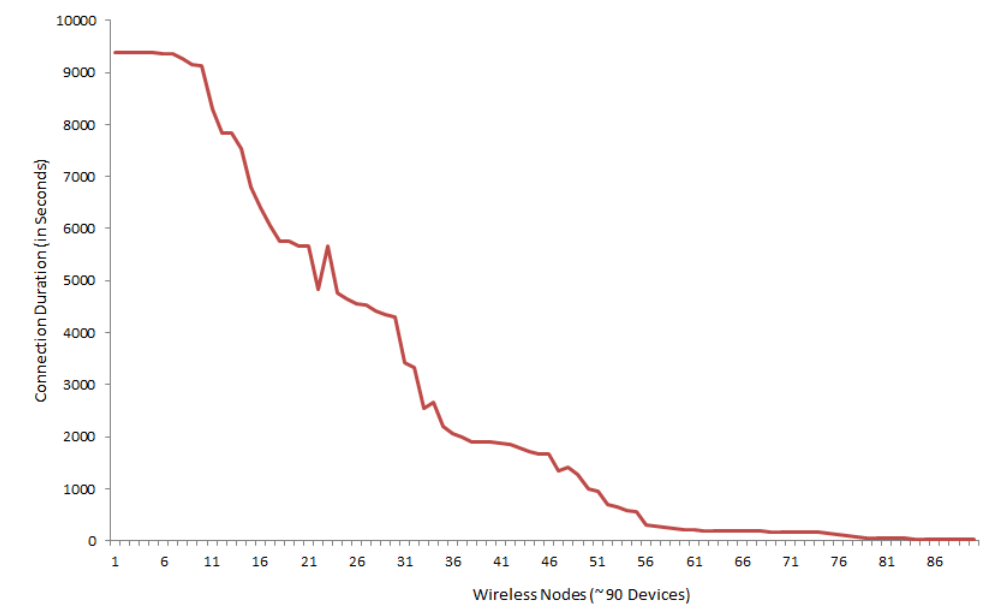


Fig. 3.7 Connection Duration Graph

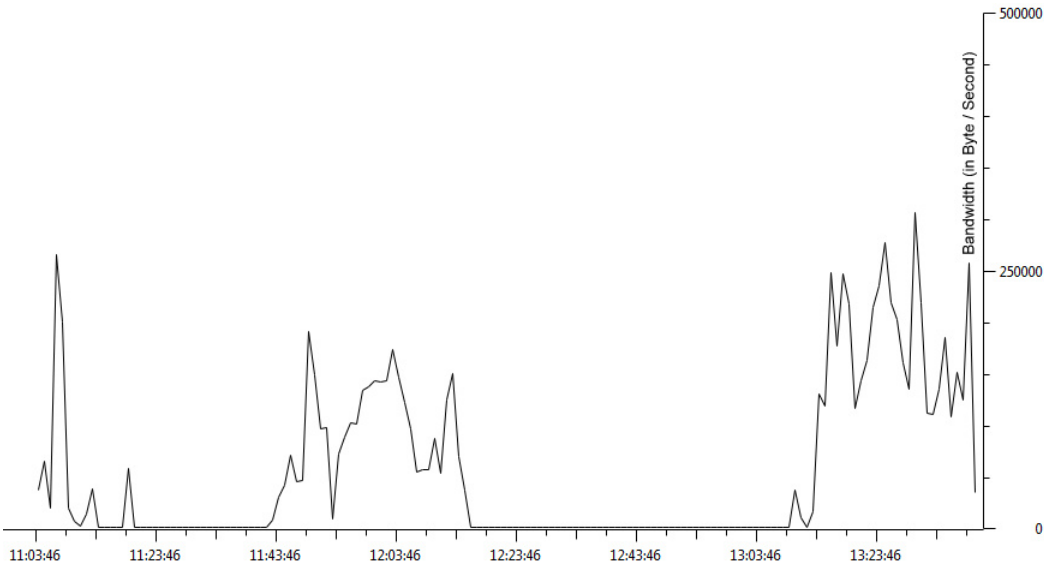


Fig. 3.8 DATASET A (Morning) - Bandwidth Utilization

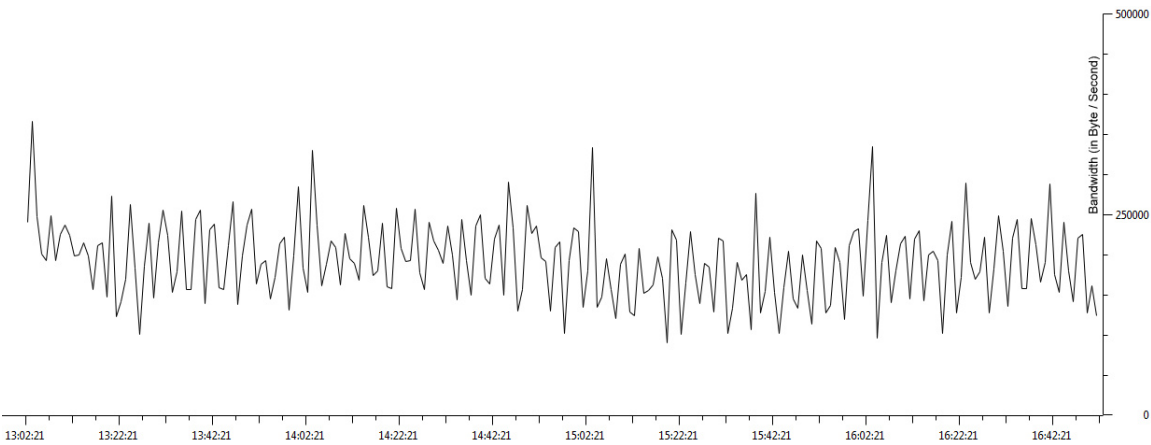


Fig. 3.9 DATASET B (Afternoon) - Bandwidth Utilization

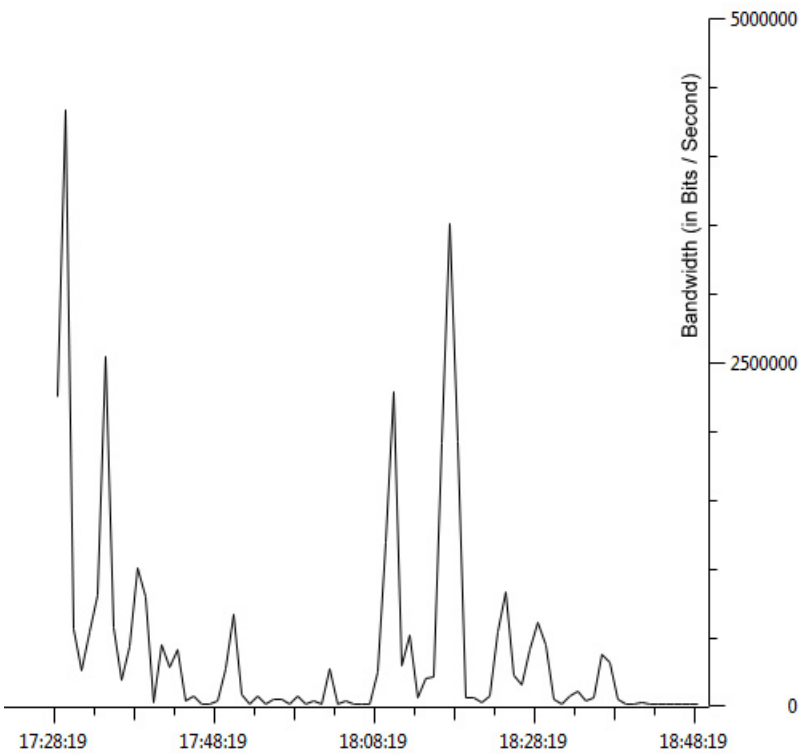


Fig. 3.10 DATASET C (Evening) - Bandwidth Utilization

3.5 Application and Protocol Composition

In today's internet, we find new applications and protocols (to support) deployed quite frequently. It is essential for us to keep up with the ever-changing internet traffic. This study is an extension of the packet level approach that provides resourceful insight into application types and triggered protocols.

Though there are many application types, we observe that *Web Browsing*, *OSN*, and *e-mail* are quite popular in PWLAN. There are other application types such as *P2P*, *Gaming*, and *Video Streaming* but their percentage is minimal in our dataset. The following graphs (Fig. 3.11 - 3.13) sketches the composition of popular application types such as *Web Browsing*, *OSN*, and *E-mail* in DATASET A, B and C. The presence of such application types is evident from the bandwidth utilization measure. From the protocol composition, we observe that TCP is still the dominant transport layer protocol, followed by UDP and ICMP. For most application types, TCP and UDP are default transport layer protocols. TCP dominates the mix because of its properties such as reliable communication, congestion control, retransmission and error checking. The following graphs (Fig. 3.14 - 3.16) sketches the composition of TCP, UDP and other protocols triggered in PWLAN.

3.5.1 Hit Rates of Websites

The following table (Table. 3.2) tabulates the hit rates or connection arrival rates of popular websites. With the help of this analysis, in chapter 4, we implement security assessment procedures and simulate network security attacks to demonstrate the level of security risk in public WiFi.

Application	Website	Connection Arrival Rate
Web Browsing	Google, Bing, Yahoo, etc.	36.421
Social Networking (OSN)	Facebook, Twitter, Pinterest, etc.	21.857
E-mail	Gmail, Yahoo, etc.	9.671
Video Streaming	Youtube, Vimeo, CrunchyRoll, etc.	2.084
P2P File Sharing	BitTorrent, YTS, Torrentz, etc.	0.812

Table 3.2 PWLAN - Hit Rates of Websites

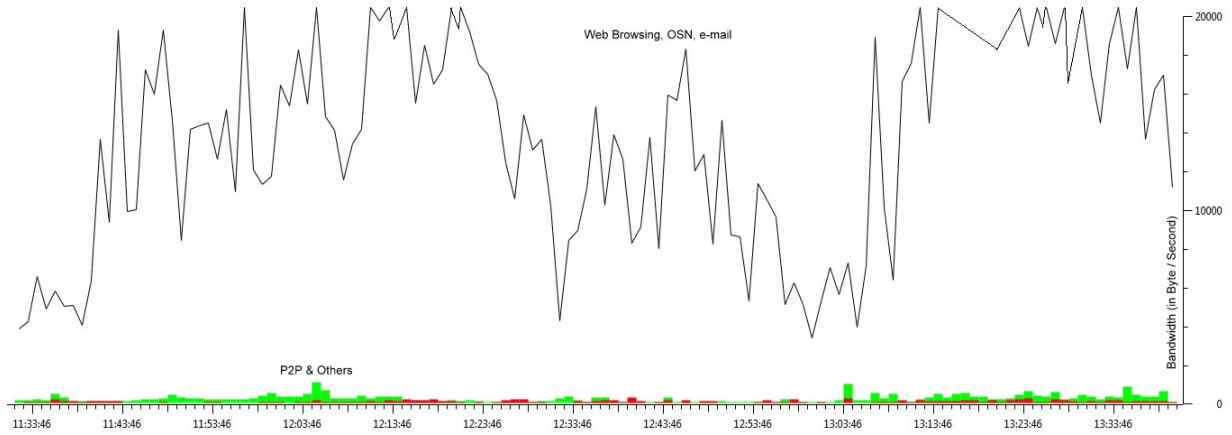


Fig. 3.11 DATASET A (Morning) - Application Composition

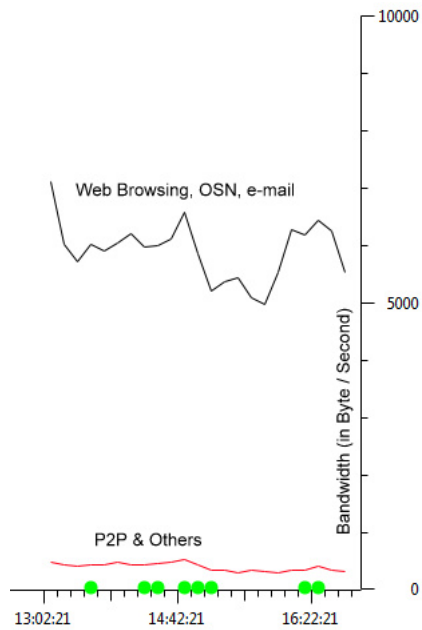


Fig. 3.12 DATASET B (Afternoon) - Application Composition

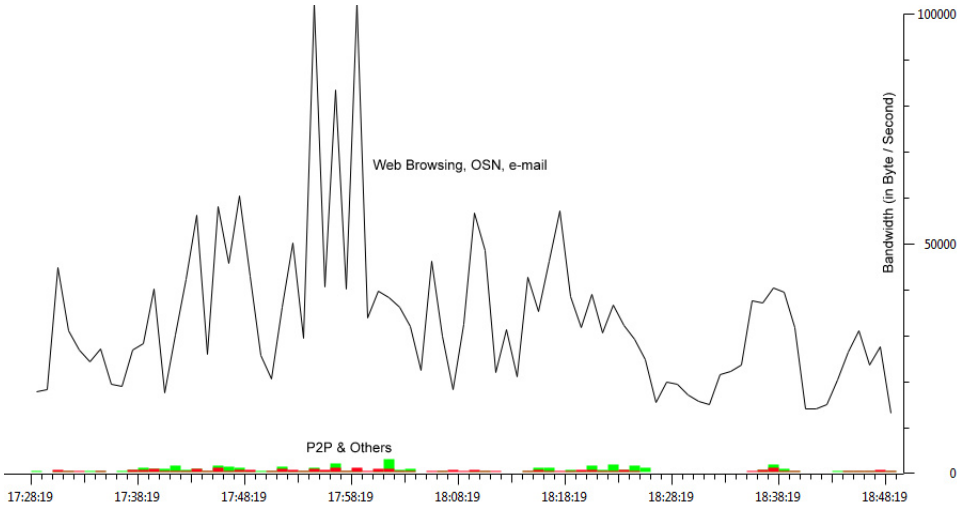


Fig. 3.13 DATASET C (Evening) - Application Composition

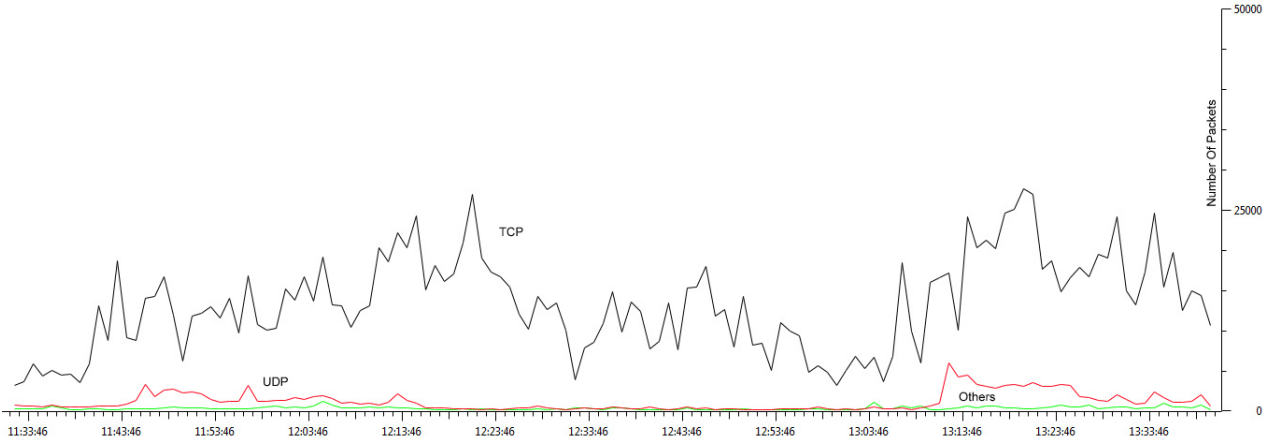


Fig. 3.14 DATASET A (Morning) - Protocol Composition

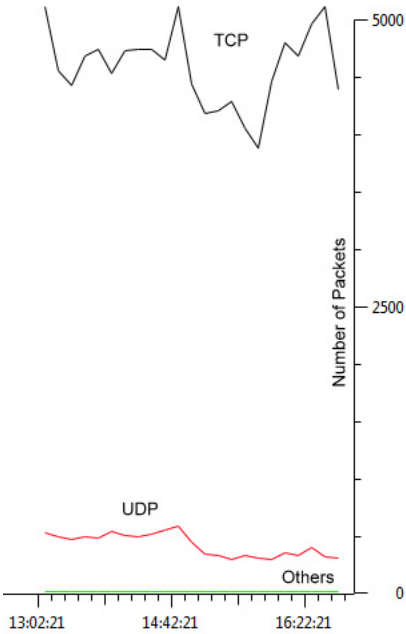


Fig. 3.15 DATASET B (Afternoon) - Protocol Composition

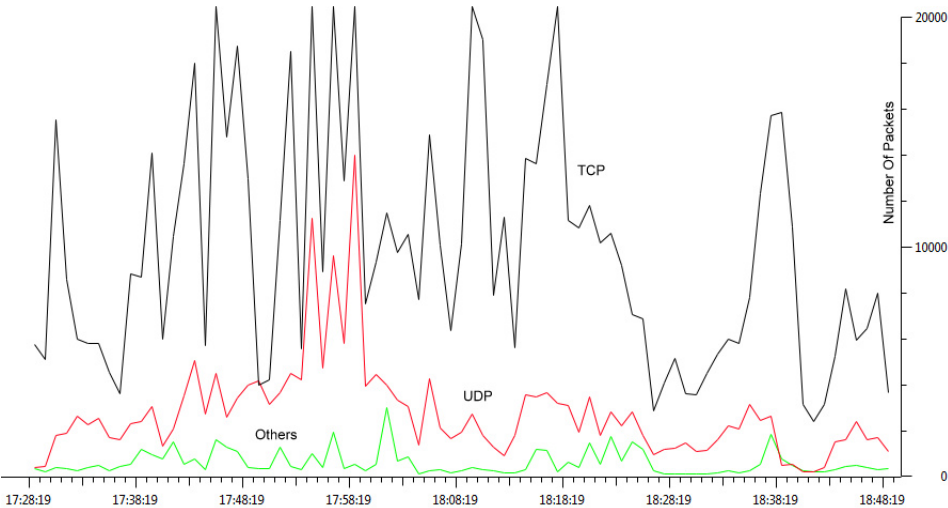


Fig. 3.16 DATASET C (Evening) - Protocol Composition

Chapter 4

Network Security Implications

In this chapter, we present the followed procedure to expose network security vulnerabilities in PWLAN. First, we carry out a security assessment of the network by determining web traffic composition (HTTP/HTTPS), analyzing the presence and use of encryption schemes (WEP/WPA) and studying firewall rules. Secondly, we construct the security trace report to analyze packets with security attack traits. Finally, we simulate two common *Phishing* attacks to intentionally breach confidentiality and demonstrate the level of security risk in public WiFi.

4.1 Security Assessment

The primary objective of security assessment is to identify factors that promote weakness in network security. This step is vital to our study for two important reasons,

- Identifying security policies that are implemented in the network;
- Understanding possible schemes developed by hackers to acquire sensitive user data.

In our research, we carry out a security assessment for the entire traffic dataset (morning, afternoon and evening). The following sections provide the different assessment strategies and analyzed results in PWLAN.

4.1.1 Web Traffic Composition

In this step, we filter HTTP and HTTPS web traffic to determine their composition in the network traffic to facilitate security insight into accessed websites and user behavior. We filter HTTP traffic on well-known TCP port 80 and HTTPS on TCP port 443 as it carries the majority of internet traffic. There are other transport layer protocols (example. UDP) that also support HTTP or web traffic. It is important to note that HTTP does not provide secured client-server communication through any encryption mechanism of its own. HTTP relies on the network's encryption scheme (WEP or WPA) for protecting user data, without which, web traffic is available in *plain-text* or unencrypted form. HTTPS supported websites offer secured client-server communication. Hackers are always keen to attack HTTPS packets that contain sensitive user data such as, user IDs, passwords, email IDs, etc.. These credentials are used for *On-line Social Networking*(OSN), email, banking, and other internet accounts. Hackers employ brute force, session hijacking, rogue AP and SSL strip attacks to hack HTTPS packets. Table. 4.1 tabulates the HTTP and HTTPS composition in PWLAN.

Protocol	Composition (%)
HTTP	54.72
HTTPS	45.28
HTTP GET	98.0
HTTP POST	2.0

Table 4.1 PWLAN - Web Traffic Composition

From the above table, HTTP dominates the web traffic composition. Many websites that render HTTP version are starting to support HTTPS version too. In the transparency report published by Google audits [31], it is evident that 80% of top websites in the world do not deploy HTTPS by default. Therefore, HTTP user connection is more prevalent than HTTPS. The small percentage of HTTP-POST indicates that user credentials such as user IDs, passwords, email IDs, etc. posted on web servers are available in *plain-text* or unencrypted form. We postulate the use of same plain-text user credentials for other internet accounts.

4.1.2 Encryption Scheme Analysis

By analyzing the presence and use of encryption schemes, we study the general scenario of network encryption in PWLAN. WEP and WPA are two common security standards that offer data encryption and client authentication in WiFi network. Following the safety critical reasons stated in [17], WEP is obsolete. The weakness of WEP is as follows,

- We recall the WEP encryption steps from Chapter 2, WEP computes *Integrity Check Value*(ICV) on the data units of 802.11 frame to keep track of data consistency. The ICV value is not encrypted. Therefore it opens a possibility of tampering user data by changing the payload bits and consequently changing the ICV. At this point, we lose data integrity;
- Hackers can masquerade as authorized users to redirect user traffic towards their malicious websites;
- Data replay attacks are widespread. By re-transmitting collected WEP packets at a later point in time reveals information about the encryption key and user data;
- WEP uses the encryption key derived from *Initialization Vector*(IV) and a shared base key for its RC4 encryption algorithm. Since the IV is of short length (24 bits), the same IV could appear in another packet after a certain time interval to expose the encryption key;
- Modern brute force or key cracking tools can hack WEP in few minutes or even seconds.

WPA2 security standard is widely used in home and business WiFi networks as it provides stronger encryption and user authentication. In our study, we find that out of 50 public AP (coffee shop, public library, and hotel) only 2 are provisioned with an encryption scheme. We also observe that secured APs have WPA-PSK security standard. WPA-PSK can be hacked by cracking the pre-shared key (small length password) using rainbow tables. WPA or WPA2 is susceptible to brute force and dictionary attacks. It is not that WPA/WPA2 is unbreakable but employing it to encrypt user traffic provides a better chance of resisting security attacks in public WiFi.

4.1.3 Firewall Rules

By studying firewall rules, we justify the significance of implementing a firewall on network switches and routers to provide network security. Firewall is a program and network device that blocks undesired internet traffic from public internet entering into a private network. It usually works on the application, transport, network, and data link layers. It is well known for its packet filtering and proxy services. Firewall is programmed to work on rules that are set and defined for the network traffic. These rules are classified into two types,

- *Inbound Rule* is defined to inspect traffic destined to a private network, sourced from the public internet;
- *Outbound Rule* is defined to inspect traffic sourced from a private network, destined to the public internet.

Firewall rules operate and process inbound traffic packets in order of precedence (as follows).

- *Bypass* - It is an option through which rules are set for authorized traffic to bypass "block" rules;
- *Force Allow* - To user preference or setup, certain known traffic is allowed inside the network. For example, traffic from users' remote work machine;
- *Block or Deny* - Blocks all inbound traffic if it matches a set rule;
- *Allow* - Allows inbound traffic into the network;
- *Default* - The default option is to deny all inbound traffic and allow all outbound traffic.

Firewall is installed with the default option - "block" but it allows users to configure rules and change the setup of major protocols such as, HTTP, FTP, TCP, IP, etc.. From the hacker's point of view, a firewall is a program that guards the network. Therefore, hackers analyze different open ports through port scanning methods and network probing techniques. Tools such as nmap, traceroute, hping (based on ICMP) allow hackers to detect firewall in the network path. nmap is a tool used for port scanning. It returns

either "open" or "closed" status for the port specified by the hacker. Traceroute is used for network probing. By sending ping or ICMP echo packets (by default, firewall blocks all traffic other than ping), hackers can locate a firewall in the network path. hping is used for sending TCP packets to return with information on the firewall rules. These tools allow hackers to get information on the number of active hops in the network, network topology, and the packet trace path. In our study, we use these tools to learn about the rules set up in PWLAN. To our surprise, a majority of public WiFi does not have a configured firewall. About 1 out of 50 APs implement firewall rule on the inbound traffic to block torrents (P2P traffic). We suppose that this rule is implemented to prevent users from consuming bandwidth for downloading substantial media content. It is also inferred that the software firewall becomes futile when anyone can join *Open* WiFi to collect user traffic through sniffer applications (Wireshark).

4.2 Security Trace Report

We analyze the collected traffic dataset for traces left by an attempted attack. This process provides us with the knowledge about packets containing sensitive user data and the hacker's point of view to execute security attacks. We construct the security trace report by the following methods,

1. **HTTP Return Status** - By determining the number of HTTP connections established, we build cases on HTTP request packets characterized as either success or erroneous based on the HTTP response codes sent by the web server. This method also establishes the nature of the user connection. The following table (Table. 4.2) tabulates the HTTP user connection statistics.

Total HTTP Client-Server Connections | 1,400,782

HTTP Return Code Group	Status	Composition (%)
1xx	Informational	0.17
2xx	Success	70.86
3xx	Redirection	26.28
4xx / 5xx	Error	2.69

Table 4.2 PWLAN - HTTP Return Status Code Group Statistics

HTTP return status is classified into five categories based on the code. Group 1xx codes inform the client to continue with the HTTP request, Group 2xx codes report success status for client connections, Group 3xx is used for connection redirection, Groups 4xx and 5xx codes are used for reporting errors in client and server respectively. We provide the composition of each return status code groups in the following study conducted in PWLAN,

- *Group 1xx* - The composition of code 100 is 90.27 % and code 101 is 9.73 %. Code 100 indicates that the server accepted the initial HTTP request and allowed the client to continue with the request. Code 101 is used for switching to a different protocol for better service to the client request;
- *Group 2xx* - The composition of code 200 (79.29 %), code 201 (0.15 %), code 204 (20.43 %) and code 206 (0.13 %) is observed in the collected traffic. Code 200 is returned to report that the server successfully served the client request. Code 201 indicates that the resource provided in the HTTP POST or PUT method is created successfully on the server. Code 204 is generally returned for a successful delete resource operation. Code 206 indicates that only a part of the requested resource is sent by the server (example, sending partial data after resuming an interrupted download). We observe that the HTTP-POST method is present in the collected traffic to create new web resources;
- *Group 3xx* - The composition of code 302 (34.62 %), code 304 (29.83 %) and code 307 (35.55 %) is observed in the collected traffic. Code 302 indicates that the requested resource could have been moved temporarily to a different location. In such case, the client should send the request again for redirecting to the new location. Code 304 indicates that the requested resource is not modified from the time since the client downloaded a copy earlier than its current request. Therefore, the server does not re-transmit data. Code 307 is similar to 302 except that when a client sends the request again for redirection, the request method is not changed;
- *Group 4xx* - The composition of code 400 (43.28 %), code 401 (5.12 %), code 403 (20.76 %) and code 404 (30.84 %) is observed in the collected traffic. Code 400 is generally returned for a bad request characterized by a syntax error in the request message. Code 401 is generated when the server requires the only

authenticated request to serve. Code 403 is used by the server when it is unable to serve client requests due to reasons undisclosed. Code 404 is used instead of 403 when the requested resource is not found;

- *Group 5xx* - The composition of code 500 (92.63 %) and code 502 (7.37 %) is observed in the collected traffic. Code 500 is returned due to an unexpected error on the server side, and therefore client requests are denied. Code 502 is returned when an upstream server returns an invalid response to a requesting server that acts as a proxy for serving client requests.

2. **Unencrypted User Credentials** - We analyze the collected traffic datasets for unencrypted user data such as user IDs, passwords, email IDs, etc.. Application protocols such as HTTP, FTP, Telnet, POP3, SMTP, IMAP, etc. are triggered as a result of user activity on the network. The following points describe the security implications in few of the protocols mentioned above.

- HTTP - As pointed out earlier, HTTP does not have an encryption algorithm defined in its protocol suite. Therefore it relies on the network's encryption scheme (WEP or WPA) to provide secured client-server communication. Since a majority of public WiFi is not provisioned with an encryption scheme, HTTP traffic is available in unencrypted or *plain-text* form to all users in the same network. Websites that render HTTP version converse in *plain-text* form of web traffic with the client;
- FTP and Telnet [32, 33] - *File Transfer Protocol*(FTP) is the most common protocol used for file transfers between host and requesting clients. Telnet enables users to communicate with a remote machine over TCP connection. These two protocols are considered highly insecure without a network encryption passphrase (WEP or WPA) as it is easy to breach important security properties (confidentiality, integrity, and authentication),
 - (a) *Confidentiality* - User credentials are available in *plain-text* form, therefore it is easy to hack sensitive user data;
 - (b) *Integrity* - User data can be tampered as there are no integrity checks;
 - (c) *Authentication* - Insecure user ID-password pair is used for authentication.

Note: These two protocols do not support any other authentication mechanism other than user ID-password.

- SMTP [34]- *Simple Mail Transfer Protocol*(SMTP) suffers from the *plain-text* communication of data between mail servers if the network is not provisioned with an encryption scheme (WEP or WPA). SMTP servers are built on trust, but hackers exploit them to e-mail malicious content to targeted users. Hackers employ IP spoofing techniques to hide their identity while communicating with the SMTP server in *plain-text*.

```
Stream Content
```

POST /mail_send.php HTTP/1.1
Host: [REDACTED]
Connection: keep-alive
Content-Length: 342
Cache-Control: max-age=0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Origin: [REDACTED]
User-Agent: Mozilla/5.0 (Linux; Android 4.3; Nexus 7 Build/JWR66V) AppleWebKit/537.36
Content-Type: application/x-www-form-urlencoded
Referer: http://[REDACTED]
Accept-Encoding: gzip,deflate,sdch
Accept-Language: en-US,en;q=0.8
Cookie: [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]&orgname=tulips&username=tulips&password=[REDACTED]&message=Hi+&[REDACTED]Xappy+Labor+Day%
[REDACTED]-HTTP/1.1 302 Found
Date: Sun, 01 Sep 2013 21:45:56 GMT
Server: Apache
X-Powered-By: PHP/5.3.14
Expires: Thu, 19 Nov 1981 08:52:00 GMT
Cache-Control: no-store, no-cache, must-revalidate, post-check=0, pre-check=0
Pragma: no-cache
Set-Cookie: login_at=1378071956; expires=Mon, 02-Sep-2013 21:45:56 GMT
location: view_profile.php?userid=204115
Vary: Accept-Encoding
Content-Encoding: gzip
Content-Length: 20
Keep-Alive: timeout=2, max=500
Connection: Keep-Alive
Content-Type: text/html; charset=ISO-8859-1

Fig. 4.1 HTTP-POST Traffic (Unencrypted)

In this experiment, we analyze the TCP flow stream of HTTP client-server communication after filtering packets with only the HTTP-POST method. From Fig. 4.1, we observe that the traffic contains user credentials and data posted on an HTTP website. It is readily available in *plain-text* or unencrypted form. This test proves the

HTTP security nature in public or *Open* WiFi. We also observe FTP user downloaded data by filtering traffic on well-known TCP Port 20.

3. **Web Servers** - Web server is used to host HTTP websites to deliver their content to requesting clients. In our study, we find commonly used web servers such as Apache, Microsoft IIS, and Lighttpd. This study allows us to pinpoint documented security vulnerabilities in these web servers that are exploited by hackers.

- Apache HTTP web server [35] - Security tests carried out by crawling websites that are hosted by Apache web server revealed that the server status page is *Open*. The server status page contains useful information such as request message, network path, IP address of the requester and information on server file structure. In our traffic, we find Apache version 2.2 (approximately 1%) that suffers from the above vulnerability;
- *Internet Information Services*(IIS) server [36] - Microsoft's IIS version 5.0 (2% of the trace in our traffic) has proven security risks that allow hackers to upload and execute malicious content on server's directory. Hackers can perform *Denial Of Service*(DoS) attack on vulnerable servers and compromise system privileges. IIS version 6.0 (commonly seen in our traffic) is an update with the security patch that resolved the earlier issue. However, hackers can acquire sensitive data from server's protected files by sending numerous request bits through WebDAV protocol. This protocol enables hackers to manipulate files stored in a web server;
- Lighttpd [37] - It provides the same function as any other web server. In recent years, numerous security vulnerabilities of Lighttpd have come to light. Allowing hackers to perform *Denial Of Service*(DoS) attacks, traversing server directories and elevating access permissions to retrieve protected files are few documented incidents. In our collected traffic, we find versions 1.4.19, 1.4.23 and 1.4.26 that exhibit the above vulnerabilities.

4. **DNS queries** - *Domain Name System*(DNS) is used to resolve a requested website name (for example, "www.google.ca") to its appropriate IP address (173.194.41.184). DNS enables the client to find the server location on the internet. Recording all the name-IP values in one DNS server is impossible. Generally, the DNS infrastructure

is created in such a way that many DNS servers communicate with each other to find the information to resolve a client request successfully. In the following points, we describe DNS operation and security vulnerabilities [38],

- The client has a stub process called the resolver that contacts the DNS server with a query to resolve. The DNS name server accepts the client request and checks if it has a resolved result. From Fig. 4.2, usually the request sets a flag called recursive query that tells the origin DNS server to query other DNS servers to find the resolved result (in-case the origin DNS server does not contain the result). When the server gets the result, it sends the response packet with the IP address of the host back to the resolver. DNS servers are implemented with a cache table to make this process more efficient. DNS cache table records successful resolved result to serve the next client request (if repeated) without querying other servers repeatedly. In home WiFi, ISP's name server acts as the origin DNS server. There are public DNS name servers available to resolve DNS queries (for example, OpenDNS);

```

Frame 8665: 236 bytes on wire (1888 bits), 236 bytes captured (1888 bits)
Radiotap Header v0, Length 29
IEEE 802.11 QoS Data, Flags: .....F.C
Logical-Link Control
Internet Protocol Version 4, Src: 10.128.128.128 (10.128.128.128), Dst: 10.158.72.185 (10.158.72.185)
User Datagram Protocol, Src Port: domain (53), Dst Port: 55116 (55116)
Domain Name System (response)
[Request In: 8210]
[Time: 0.669065000 seconds]
Transaction ID: 0x5562
Flags: 0x8180 Standard query response, No error
  1... .. = Response: Message is a response
  .000 0... .. = Opcode: Standard query (0)
  .... 0... .. = Authoritative: Server is not an authority for domain
  .... 0... .. = Truncated: Message is not truncated
  .... 1... .. = Recursion desired: Do query recursively
  .... 1... .. = Recursion available: Server can do recursive queries
  .... 0... .. = Z: reserved (0)
  .... 0... .. = Answer authenticated: Answer/authority portion was not authenticated by the server
  .... 0... .. = Non-authenticated data: Unacceptable
  .... 0000 = Reply code: No error (0)
Questions: 1
Answer RRs: 2
Authority RRs: 2
Additional RRs: 2
Queries
  www.facebook.com: type A, class IN
Answers
  www.facebook.com: type CNAME, class IN, cname star.c10r.facebook.com
  star.c10r.facebook.com: type A, class IN, addr 69.171.242.27
Authoritative nameservers
Additional records

```

Fig. 4.2 DNS query response

- If hackers can trick the DNS resolver to accept DNS response with an IP address to redirect client traffic to their malicious website then their attack succeeds. We expand on this statement with an example stated in [38]. We have Bob acting as the resolver, Alice as the DNS name server and Eve as the hacker. Now for every request made by Bob, there is a query identifier called qid associated with it. If Bob sees this qid returned in the response message, then he accepts the message and gets directed to that specific IP. For Eve to make Bob accept her response, first, she should be able to see the qid sent by Bob. Generally, this qid is a 16-bit value (not a strong value). If Eve guesses this qid and her response reaches Bob before Alice's response, then Eve wins, and her resolved result (name-IP) is cached in the DNS cache table (cache poisoning) for the next client query (if repeated).

Every cached result is stored in the DNS cache table with a TTL(Time To Live) which means that DNS stores the result for that TTL time value and once expired, the value is overwritten by resolved results following the process all over again. The resolver generates random qid to defend against such attacks. It increases resilience against hackers' guessed values of qid, and within that time, the original DNS response must reach the client. It is the concept behind source port randomization. In our collected traffic, we study the above phenomenon by examining the DNS packets. We carry out an experiment suggested in [39], a tool called dig is used to check the DNS name server's randomness. The degree of randomness is given a "Great," "Good" and "Poor" rating depending on the standard deviation values (Great - 3980 to 20,000, Good - 296 to 3980 and Poor 0-296) on the ports used by the client resolver. The below command executed in Linux OS tests the DNS server at the specified IP and the result obtained is - "GOOD" with a standard deviation of 3654.

```
$ dig @10.128.128.128 +short Porttest.dns-0arc.net TXT
```

5. **TCP Sequence Numbers** - This study is carried out to test the server-side sequence number generation. *Transmission Control Protocol*(TCP) is the widely used transport layer protocol by applications that require reliable transport of packets and establish a connection with the server. The client and server maintain sequence numbers to keep track of the packets exchanged. To send data, First, a TCP connection

is established between the client and server through a three-way handshake process. In this process, the client sends its *Initial Sequence Number*(ISN) to the server. On successful receive, the server responds with its ISN, and the client acknowledges back. The process is represented as follows, (C - Client, S - Server, SYN - Synchronized (TCP Flag), ACK - Acknowledgment)

C->S:SYN(ISN of C) — S->C:SYN(ISN of S), ACK (ISN of C) — C->S: ACK
(ISN of S) — C->S (or) S->C: data

We describe the following documented security vulnerabilities of sequence numbers used in TCP connections,

- *TCP Sequence Number Prediction* [40] - In the three-way handshake process, if hackers can guess the initial sequence number of the server, then they can send data to hack the server. Impersonating authorized user IP (*IP Spoofing*) accomplishes this attack. In the older protocol version, the ISN generated by the server incremented by a constant number (per second) thus allowing hackers to easily guess the next sequence by obtaining the "increment" value. To resist this attack, the server's ISN is randomized with a 32-bit value;
- *TCP RST (Reset Flag) attack* [41] - This attack is used to target a remote server to cause all established connections to shut down (*DoS*). It is carried out by forcing the server to accept the RST flag that has sequence numbers in the server-generated range. In a long lasting connection, hackers can guess multiple sequence number range to shut down the server. The solution to make the RST sequence number same as the server's expected sequence number increases the odds of resisting this attack;
- *TCP Spoofing Vulnerability* [42] - If the ACK (Acknowledgment) sequence number is too low or too big, then it is ignored by the client, followed by sending RST flag to the server. If hackers can guess the correct ACK sequence number then they can send RST to the server to close active connections (extension of TCP RST attack).

In our collected traffic datasets, we observe that the server sequence number is randomized (resists sequence number attack). Fig. 4.3. shows the flow graph for the

TCP based HTTP server response packet in the three-way handshake process. We find that the ISN of the server set to 0, but in fact, it is a random number as seen from packet ASCII code (Fig. 4.4.), 0x4a1c548e (decimal equivalent = 1243370638). We also infer that there are no traces of an unusual amount of RST packets sent to the connected server.

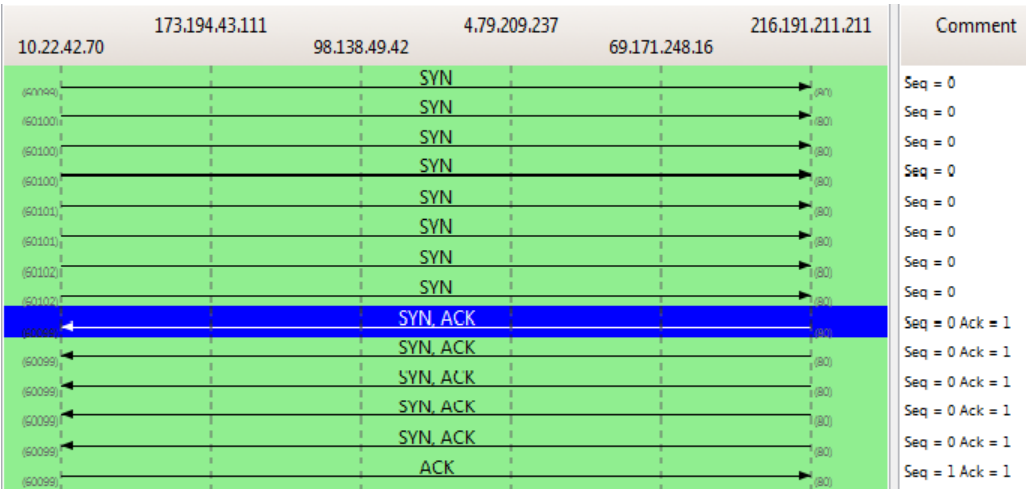


Fig. 4.3 TCP Flow Graph for HTTP Client-Server Handshake

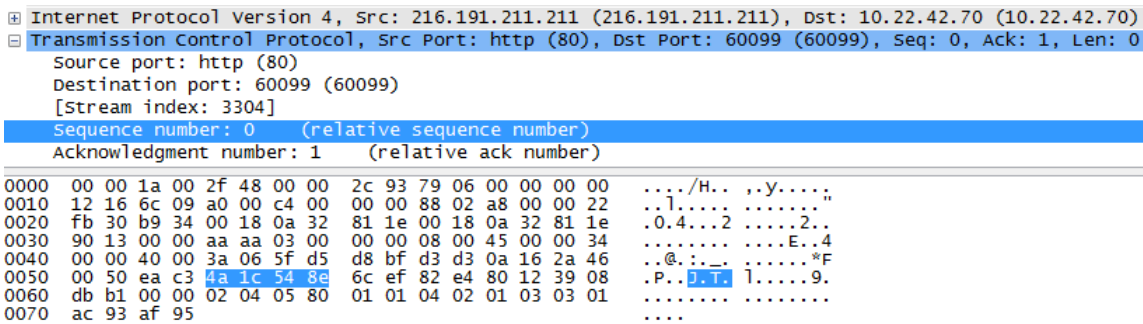


Fig. 4.4 TCP Initial Sequence Number in HTTP Server

6. **ICMP Echo Packets and Error Messages** - *Internet Control Message Protocol*(ICMP) occupies the core of IP. It is used for reporting error messages for any short-lived connection error and providing a sense of the network topology through its query messages. ICMP has the following documented security vulnerabilities [43],

- *Reconnaissance* - This method is used for acquiring network information by using ICMP query and error messages. Traceroute is a program (or tool) based on ICMP query (ICMP echo request) to detect network devices such as firewalls and routers in the network path. It provides a brief idea of the underlying network design. ICMP error messages show the information on protocols and ports used by the host. Other methods to acquire network information include *Access Control List*(ACL) detection for identifying ACL rules applied to packet filters (as seen in section 4.1.3) and OS fingerprinting;
- *DoS Attacks* - Smurf DoS and Tribe Flood Network (TFN) attacks are executed through ICMP messages. In smurf DoS, attackers send IP spoofed ICMP echo requests to broadcast address by which multiple hosts can listen and respond to the request. It causes a flood of response messages with a high degree of traffic congestion in the network. TFN is a distributed DoS attack suite for flooding attacks carried out by ICMP, SYN packets. Hackers executing this attack would normally enslave one or more compromised clients that would handle systems under it to simultaneously flood with ICMP, SYN packets to shut down a targeted server;
- *Covert Channels* - From section 4.1.3 (Chapter 4), firewall only allows ICMP echo messages (ping traffic) in and out of the network by default. ICMP messages contain a data field that can be set to record the network path or RTT (Round Trip Time) values for ICMP echo reply-response. It is possible to send a spoofed request with malicious content through the covert channel opened by the ICMP echo or ping traffic.

Fig. 4.5, shows ICMP error message packet that provides information about the port and protocols used by the host.

```

Internet Control Message Protocol
Type: 3 (Destination unreachable)
Code: 3 (Port unreachable)
Checksum: 0x9a5e [correct]
❑ Internet Protocol Version 4, Src: 10.76.215.240 (10.76.215.240), Dst: 123.201.124.38 (123.201.124.38)
  Version: 4
  Header length: 20 bytes
  ❑ Differentiated Services Field: 0x00 (DSCP 0x00: Default; ECN: 0x00: Not-ECT (Not ECN-Capable Transport))
  Total Length: 58
  Identification: 0x1841 (6209)
  ❑ Flags: 0x00
  Fragment offset: 0
  Time to live: 47
  Protocol: UDP (17)
  ❑ Header checksum: 0x9946 [correct]
  Source: 10.76.215.240 (10.76.215.240)
  Destination: 123.201.124.38 (123.201.124.38)
  [Source GeoIP: Unknown]
  [Destination GeoIP: Unknown]
❑ User Datagram Protocol, Src Port: 51413 (51413), Dst Port: 37709 (37709)
  Source port: 51413 (51413)
  Destination port: 37709 (37709)
  Length: 38
  ❑ Checksum: 0x0655 [unchecked, not all data available]

```

Fig. 4.5 ICMP Error Message

4.3 Security Attack

Hackers use various factors and tools to their advantage and carry out successful security attacks to acquire sensitive user data. It includes user IDs, passwords, email IDs, OSN accounts, credit cards, corporate files, government documents, media files, etc. Our security attack experiments are devised from the hacker's perspective.

Hacking Factors

The following are key hacking factors that influence successful security attacks,

- *Medium of communication* - In our study, *Open* WiFi acts as the medium for users to connect to the internet in different public venues;
- *Potential targets* - Hackers identify their targets by analyzing the network traffic through various tools (listed in the next section);
- *Nature of information* - We believe that the nature of information could be user login credentials, credit card information, confidential files, etc. that satisfies hackers' agenda;
- *Insecure software design* - Application or protocol design and implementation undermine security in the development process. Most developers do not follow secure

coding practices. In consequence, hackers take advantage of this flaw;

- *Visual deception and naivety* - Lack of educational awareness amongst internet users and naivety plays a major role in successful security attacks.

In our research, we use *Open* WiFi connection to target user login credentials and OSN accounts. Lack of educational awareness and visual deception serves as the base for executing our procedure.

Hacking Tools

We list a few of the major tools used by hackers to breach security properties,

- Linux OS - backtrack and kali;
- Network monitoring and sniffers - Wireshark, NAST, Cain and Abel;
- Web hosting services - GoDaddy, 000webhost, ripway;
- Website cloning - SET, cloneSite;
- Port Scanners - nmap, hping, traceroute.

Our experiment uses Wireshark, Cain and Abel, traceroute, PHP and MySQL for executing two types of *Phishing* attacks to target the confidentiality property.

4.3.1 Open WiFi versus Encrypted WiFi

In our research, we use the term *Open* WiFi for any WiFi network that has unsecured or *Open* AP. *Encrypted* WiFi has secured AP through WEP or WPA network encryption scheme (predominantly WPA2). *Wireless user isolation* is implemented in *Open* WiFi for providing security but not in *Encrypted* WiFi. *Wireless user isolation* is the process of isolating each user connected to the AP. Therefore, other users cannot be reached on the same network. It is implemented to prevent hackers from reaching client machines to acquire user data. In *Encrypted* WiFi, other users on the same network can be reached. For example, users share data by creating a network shared drive. It is allowed because of the provision of the network with an encryption scheme to provide security. Users connected

to *Open* WiFi do not know "who else is connected" since anyone can join the network at any time; therefore, users have less control over the network. Whereas, In *Encrypted* WiFi network (for example, at home), users have better control, and they can see "who else is connected" to their network. Security attacks in *Open* WiFi can be carried out easily using less time and resource. In *Encrypted* WiFi, hackers dedicate a lot of time and resource. *Open* WiFi limits user activity through controlling network bandwidth. In *Encrypted* WiFi (for example, at home), it depends on the personal internet connection plan with an ISP. *Open* WiFi is increasingly popular because of "internet for everyone, anywhere" motto of many network providers, store owners, businesses (Google, Facebook, Amazon), etc. Users save mobile data cost and stay connected to the internet anywhere. *Encrypted* WiFi is limited to location, unless, *Open* WiFi is provisioned with WEP or WPA for encrypting user traffic (which unfortunately is not the case for a majority of public WiFi). The cost of *Open* WiFi network maintenance is high compared to *Encrypted* WiFi, therefore, many network providers cut costs on network infrastructure and shy away from security.

4.3.2 Phishing

Phishing is the most common security attack carried out by hackers to acquire sensitive user data such as user IDs, passwords, email IDs, etc.. It can be engineered to give up the credit card or banking information. This attack can be executed on both *Open* and *Encrypted* WiFi networks. In *Open* WiFi, *Phishing* is more common because hackers dedicate less time and resource to acquire numerous user IDs, passwords, email IDs, etc. *Phishing* is classified into two types, *Man-in-Middle(MIM)* phishing and *E-mail* phishing.

4.3.2.1 Man-in-Middle(MIM) Phishing

Man-in-Middle(MIM) is the most common form of *Phishing* attack carried out in *Open* WiFi to hack *On-line Social Networking(OSN)* accounts. It uses the *ARP-Spoofing* technique along with specialized tools such as Wireshark and password sniffers. Hackers collect HTTP or web cookies of the user connection to OSN websites (for example, Facebook). The collected web cookies are later reused to hack user OSN accounts.

Experiment Procedure

- Join any *Open* WiFi (for example, coffee shop WiFi);

- Use Wireshark to collect user traffic;
- Filter HTTP traffic data;
- All the website domains visited by the user should be available;
- Filter HTTP cookie (type "http.cookie") of any visited OSN website (for example, Facebook);
- Use cookie injector scripts (EditThisCookie) to inject the collected HTTP cookie back into the OSN website through any web browser (for example, Google Chrome);
- User OSN account is hacked without cracking the user credentials;
- Note: The attack works a majority of the time when users are active;
- This attack consumes less time and resource;
- Identifying target IP is key. We use the results from user traffic characteristics study to identify long-lasting users (refer to chapter 3).

Using the above procedure we collected unencrypted user credentials, e-mail IDs, FTP downloads and web searches. This experiment is used to sketch user behavior in public WiFi. Fig. 4.6 and 4.7, illustrates the MIM attack in public WiFi.

```
> Internet Protocol Version 4, Src: 10.253.196.237, Dst: 216.115.110.119
> Transmission Control Protocol, Src Port: 39075, Dst Port: 80, Seq: 3300, Ack: 1, Len: 543
> [4 Reassembled TCP Segments (3842 bytes): #351307(1396), #351308(507), #351309(1396), #351310(543)]
v Hypertext Transfer Protocol
  > POST /w/ygo-mail/msgreply.bp?.ts=1378071446&.intl=ca&.lang=en-ca HTTP/1.1\r\n
    Host: m.yahoo.com\r\n
    Connection: keep-alive\r\n
    Content-Length: 1939\r\n
    Cache-Control: max-age=0\r\n
    Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8\r\n
    Origin: http://m.yahoo.com\r\n
    User-Agent: Mozilla/5.0 (Linux; Android 4.3; Nexus 7 Build/JWR66V) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/29.0.1547.59 Safari/537.36\r\n
    Content-Type: application/x-www-form-urlencoded\r\n
    Referer: http://m.yahoo.com/w/ygo-mail/reply.bp?f=Inbox&m=2_0_1_16615_AM92imIAAAx5UhtBkwAAAP0BpkM&srcp=message&i=12&.ts=1378071435&.intl=ca&..
    Accept-Encoding: gzip,deflate,sdch\r\n
    Accept-Language: en-US,en;q=0.8\r\n
  > truncated[Cookie: BP=v=1&r=422344ca&t=354; AO=o=0; ucs=bnas=0; YMDP=prod version=beachhead; YLS=v=1&p=1&n=0; V=v=0.75&cc=0&m=0; DK=v=2&d=m16&
    \r\n
    [Full request URI: http://m.yahoo.com/w/ygo-mail/msgreply.bp?.ts=1378071446&.intl=ca&.lang=en-ca]
    [HTTP request 1/1]
    File Data: 1939 bytes
```

Fig. 4.6 HTTP Cookie

```
"to" = "[REDACTED]@gmail.com"
"cc" = ""
"bcc" = ""
"subject" = "Don't You Ask Your G+4333"
"body" = "Hi [REDACTED] (5-11)
Sent you a few messages [REDACTED]"
```

Fig. 4.7 Phished Data

4.3.2.2 E-mail Phishing

We define the term *Owner's Attack*. We pointed out earlier that any user can share unused internet bandwidth with the public or create a *Open* AP to become the owner of any *Open* WiFi network. If the owner has malicious intent, then their AP is termed *rogue* AP. Such APs can be used to execute *E-mail* phishing attack on connected users. *E-mail* phishing is an attack where hackers send a visually deceptive e-mail to users to act upon. Usually, the user is "phished" for login credentials (user IDs, passwords, and email IDs) or credit card information.

Experiment Procedure

- Create *Open* WiFi with the help of mobile *Hotspots* to become an owner;
- Owner's WiFi network name or SSID is set as the public WiFi in any public venue. This step is executed to target both new and previously connected devices. The previously connected devices send beacon signals to find known networks to connect automatically. A majority of the users do not set "forget network" option in their smart device after exiting a coffee shop, and hackers take advantage of this lack of awareness when they return next time;
- The signal strength of the *hotspot* is kept maximum to attract users in and around the public venue;
- Owner creates a WiFi login page for users to submit their email address to connect to the network. It is an easy way of enumerating legitimate destination e-mail addresses;

- For successful *E-mail phishing*, the sender e-mail address must seem legitimate and the e-mail must pass spam filters;
- To create legitimate sender e-mail address, either the owner can buy a valid domain name through web hosts (for example, GoDaddy) and use their SMTP server or it can be spoofed to look legitimate;
- *E-mail Spoofing* - This process is easy to implement, hackers can buy SMTP server and mailing software (for example, PHPMailer). The mailing software creates *Phishing* mail. In the mail, the "from" address can be a legitimate sender address from a valid domain (for example, @cafe.com). *Phishing* URL can be included in the e-mail body and sent to the enumerated destination e-mail addresses;
- To make sure that the e-mail reaches the target, hackers must pass the SPF check. Email servers do the SPF check at the destination. The e-mail server matches the source IP address with the IPs in its SPF list. If a match is found, then it assumes that the e-mail is from an authorized user and allows it to the destination user's inbox else it flags the e-mail as *spam*. Currently, SPF failed messages are eventually delivered to the destination user address since e-mail servers (for example, Gmail) do not handle this exception;
- If the user doesn't check the source code of the e-mail to see SPF check failure, the attack is successful (visual deception and naivety);
- *URL Phishing* - When the targeted user clicks the sent URL link, they are redirected to the hackers' website. This website is created through the cloning technique. This technique enables hackers to clone a legitimate website and copy its "look and feel." SET is the most popular tool to carry out this technique. The HTML form action for the HTTP-POST method can be modified to any intention of the hacker to acquire sensitive user data such as user IDs, passwords, email IDs, credit card information, etc.

Through this procedure, we test "Owner's Attack" by creating a user login page to collect e-mail addresses and carry out *E-mail* phishing attack through our experimental routers that act as *Open* or public *hotspot*. We use PHP and MySQL to accomplish the task. Fig. 4.8, shows a snapshot of the created login page.

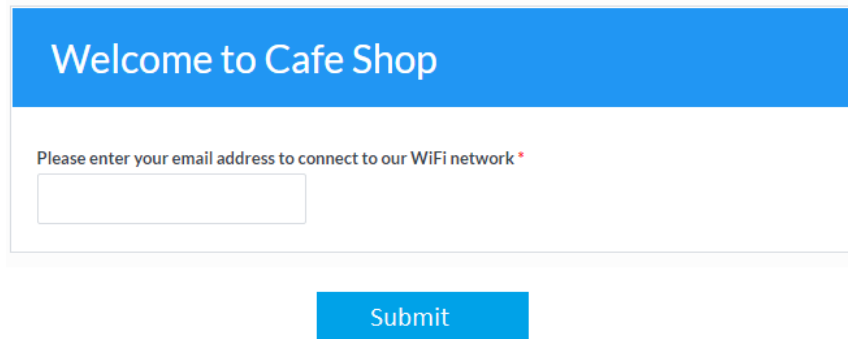
The image shows a web form titled "Welcome to Cafe Shop" in a blue header. Below the header, there is a text prompt "Please enter your email address to connect to our WiFi network *". Underneath this prompt is a rectangular input field for an email address. At the bottom of the form, there is a blue button labeled "Submit".

Fig. 4.8 PWLAN - User Login Page

4.4 Mitigating Security Risks in PWLAN

We present measures to mitigate security risks in public WiFi. It is intended for protecting user information in PWLAN. We strongly suggest internet users, software programmers, and network providers implement these measures in practice.

4.4.1 Internet User Community

- *Educational awareness* - Hackers take advantage of user naivety and lack of educational awareness. It is important to incorporate network security and user privacy in our education system. Conducting annual security campaign for the general public is a strong initiative [44] to spread awareness. Technical conferences, educational-groups, hackathon, and media are a few ways of keeping ourselves informed and updated with security policies and practices;
- *Limited actions* - In *Open* WiFi, we must take caution before, during and after accessing internet services. Before connecting to *Open* WiFi, we must take a moment to analyze if it is required to connect. When connected, limit internet actions to *web browsing*, exercise common sense while posting data on *OSN*, use of VPN and HTTPS supported websites is strongly recommended. After disassociating from *Open* WiFi, set "forget network" in the smart device and run anti-virus software to identify and eliminate malicious programs. The Stop.Think.Connect campaign [45] focuses on shaping user behavior in public WiFi;

- *Secured access* - For users connecting frequently to *Open* WiFi, use of VPN is highly recommended. *Virtual Private Network*(VPN) offers secured communication by acting as a physical barrier between the client and the internet. User traffic generated from a client is encrypted by VPN tunnel endpoint and directed to the internet from a single anonymous VPN server. Thus the hacker can not associate user traffic to individual users after collecting traffic from the network. Protocols such as *Point-to-Point Tunnelling Protocol*(PPTP), *Layer 2 Tunnelling Protocol*(L2TP) and OpenVPN offer VPN services. These protocols can be chosen based on cost and set-up procedures.

4.4.2 Software Programmers

- *Secure coding practices* - By exercising secure coding practices while developing software, we can eliminate most of the security vulnerabilities. Input data validation is an important step to check data sources, command line arguments and network interface. Hackers employ trial and error method to input data to analyze loopholes in a software program. The validation check eliminates such weakness. There is no concrete set of practices defined for secure coding. It is very much similar to the standard set of coding practices. Programmers must take actions heeding compiler errors and warnings, memory leak, unused variables, quality assurance, and user acceptance testing;
- *Threat modeling* - In this step, programmers isolate each module of software and anticipate threat factors. Once a threat is identified then it is documented and handled by the risk mitigation process. Threat modeling adheres to principles of risk assessment and mitigation in SDLC (Software Development Life Cycle) model;
- *SSDLC* - Identifying and handling security threats at early stages of software development reduces cost and maintenance overhead. *Secure Software Development Life Cycle*(SSDLC) is a well recognized and highly recommended model to address security threats during software development. In this model, security, threat, and risk factors are identified and documented at each step of software development. Therefore secured design, coding, and testing is realized at the end that saves cost, resources and especially eliminates security threats.

4.4.3 Business and Network Providers

- *Corporate policies* - For businesses, establishing strong corporate security policies and effectively communicating it to all employees is the first and foremost step. Implementing a firewall, *Intrusion Detection System*(IDS), anti-virus, content filtering, and strong authentication procedures are necessary. Weekly or monthly network security audits help to keep network security in check;
- *Secure AP* - Securing AP with an encryption scheme (WPA2 is recommended), identifying and eliminating *rogue* AP and deploying captive portal WiFi are few ways of securing public WiFi. The captive portal is time bounded WiFi provisioned through a secured login. Though it does not offer free unlimited WiFi, it is best to protect "long-lasting" users from hackers;
- *Firewall* - Configuring and maintaining a firewall is not costly, rather it is a smart investment to avoid numerous security incidents and cost associated with it. Checking and handling undesired traffic to and from a network helps users to stay safe inside the network;
- *Invest in security* - Investing in network and security infrastructure not only helps business to gain trust amongst customers but also keeps hackers at bay. Securing AP, configuring Firewall, network auditing and maintenance are few worth-while investments in the interest of protecting user information.

Chapter 5

Conclusion and Future Work

In this chapter we present,

- Inference of current trend and opportunity in PWLAN;
- Summary of *User Traffic Characteristics Study and Network Security Implications* in PWLAN;
- Future work in PWLAN research.

5.1 Conclusion

5.1.1 Current Trend and Opportunity

In the last decade, there has been a dramatic increase in the number of connected devices to access internet services. Public WiFi has now become an expectation in any public venue. From the user's perspective, staying connected online and saving mobile data are the two main reasons for connecting to public WiFi. From the facilitator's perspective, the promotion of sales and business through minimized network infrastructure (*Open* WiFi) is highly profitable. Hackers spend less time and resource to acquire sensitive user data in public Wifi compared to *Encrypted* WiFi. WLAN's inherent protocol flaws are open for devising sophisticated *Phishing by Spoofing* security attacks. In consequence, the security weakness of PWLAN has become prevalent, and we addressed it to an extent in this research.

We define PWLAN opportunity for research and business. PWLAN enables the research opportunity to study internet traffic, user behavior, and psyche. PWLAN is increasingly popular with expanding user base, and the security weakness presents a business opportunity. Users rely on VPN for secure connection in the absence of secured network infrastructure. VPN services may prove expensive, and the underlying encryption or tunneling protocol adds load to the network causing slow operating speeds. We suggest the development of a single and directed application for providing security by anonymity through the use of TOR [46].

5.1.2 User Traffic Characteristics Study

We analyzed the consolidated user traffic datasets from different smart devices collected at different times of a day in different public venues. We followed the flow level, packet level, and application and protocol composition to study the user traffic characteristics in PWLAN. The application and protocol composition helped us to take the first step to determine user behavior in PWLAN. By identifying application types and triggered protocols, we find that *Web Browsing*, *OSN*, and *e-mail* are the primary application types accessed through PWLAN. TCP dominated the transport layer protocol mix, followed by UDP and ICMP. In Flow level study, we captured the nature of PWLAN through user connection arrival rates and packet bursts pattern that reflected on user behavior. Finally, through packet level approach we used connection duration and bandwidth utilization measure to form a relationship between user behavior and traffic characteristics that influence public WiFi popularity. We identified "long-lasting" users whose information may be targeted by hackers. We tabulated hit rates of popular websites to design and implement network security experiments. The objective of user traffic characteristics study was to determine the reasons behind PWLAN popularity, aid network management activities and create a common framework for traffic modeling in PWLAN. We believe our study met that objective. We strongly suggest that the research community keeps up with the pace of PWLAN expansion to understand and tackle more challenges from its ever-changing user traffic.

5.1.3 Network Security Implications

In our research, we took an important and decisive step to explore network security aspect of PWLAN in the interest of protecting user information. We first carried out a security

assessment of the network by determining web traffic composition, analyzing encryption schemes, studying firewall rules and then constructed a security trace report. The level of security risk involved in PWLAN was demonstrated by simulating two common *Phishing* attacks. Through our experiments, we find that the majority of the network traffic composition is *web* traffic, i.e., HTTP/HTTPS traffic version v1.1. We suggest migrating web traffic from version v1.1 to version v2 for better performance and increased security. HTTP web traffic in *Open* WiFi is unsecured, and it is proved by the analysis of HTTP-POST packets to expose sensitive user data. We observed that out of 50 public APs, only 2 of them were provisioned with network encryption scheme (WEP or WPA). Majority of public WiFi does not have configured firewall. In our security trace report, we documented HTTP return code statistics, web server vulnerabilities, unencrypted user credentials and security vulnerabilities in most common protocols such as TCP, ICMP, DNS, and HTTP. The reported data indicates caution while connecting to public WiFi. Through *Phishing* attacks, we found that hackers use essential characteristics such as visual deception, trapping naive users, channeling their desire and passion for attacking communicating protocols to acquire sensitive user data. Overall, our security experiments proved successful in demonstrating security risk in *Open* WiFi. We suggest the immediate implementation of measures to protect user information in PWLAN.

5.1.3.1 Protecting User Information

We outline some of the measures for handling network security vulnerabilities and protecting user information in PWLAN,

- *Internet User Community* - We must cultivate educational awareness about internet security and privacy. Constantly exercising common sense and being vigilant while posting sensitive user data or connecting to *Open* WiFi helps in preventing security incidents. In most occasions, it is wise to avoid connecting to *Open* WiFi. If connected, we should limit ourselves to web browsing and access only HTTPS supported websites. It is safe to set our mobile or laptop devices to "forget" the network after exiting the coffee shop to avoid automatic connection to the same network name (refer to chapter 4 for more information). The use of *Virtual Private Network*(VPN) services for remote working or accessing highly sensitive user data is highly recommended. Maintaining anonymity in public internet is one of the best solutions at

present for tackling network security issues. The TOR project [46] has developed its web browser application for anonymous web browsing that guarantees privacy online;

- *Software Programmers* - Programmers must strongly exercise secure coding practices while developing software. Threat analysis and incorporating security (through *Secure Software Development Life Cycle*(SSDLC) model) at early stages of software development helps in preventing software hacks;
- *Network Providers* - Network providers must stop undermining security and invest more in network infrastructure to protect user data. They should offer basic network security by securing AP through encryption schemes and configuring a firewall. Enforcing strong corporate security policies at work, implementing content filtering, use of good quality anti-virus programs and implementing network *Intrusion Detection System*(IDS) offers strong resistance against security attacks.

5.2 Future Work

Our research is just a stepping stone towards understanding user traffic and network security in PWLAN. We believe that the research scope can be dramatically increased. Employing machine learning algorithms to analyze traffic patterns can devise a common framework for traffic modeling all types of traffic generated in PWLAN. This technique proves invaluable for network management. In our security experiments, we targeted the confidentiality property to expose network security vulnerabilities in PWLAN. The scope of security tests can be broadened by targeting integrity and availability properties in-order to expose new possible vulnerabilities. We live in a world where the internet for everyone, anywhere, is almost possible. Therefore, we must continue to tackle research challenges of ever-changing user traffic and stand for protecting user information on the web. We sincerely hope that our research appeals to the reader and triggers awareness while connecting to public WiFi.

References

- [1] “Round up on WiFi technology for business.” http://www.huffingtonpost.com/vala-afshar/50-incredible-wifi-tech-s_b_4775837.html.
- [2] “Increasing popularity of public hotspots.” <http://www.marketingcharts.com/wp/online/public-wi-fi-hotspots-grow-400-worldwide-10263/>.
- [3] Arjuna Sathiaselalan and Jon Crowcroft, “Lcd-net: Lowest cost denominator networking,” in *ACM SIGCOMM (Volume:43)*, April 2013.
- [4] “WiFi statistics indicating impact on business.” <http://www.purplewifi.net/best-wifi-statistics-2014/>.
- [5] “Security risks in public networks.” <http://www.usatoday.com/story/tech/2013/07/01/free-wi-fi-risks/2480167/>.
- [6] G. Held, *Deploying Wireless LANs : Concept, Operation and Utilization*. McGraw-Hill Professional Publishing, 2002.
- [7] “De-briefing public hotspots.” [https://de.wikipedia.org/wiki/Hot_Spot_\(WLAN\)](https://de.wikipedia.org/wiki/Hot_Spot_(WLAN)).
- [8] *IEEE 802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*.
- [9] K. Thomson, Miller G.J, and R. Wilder, “Wide-area internet traffic patterns and characteristics,” in *IEEE (Volume:11 , Issue:6)*, pp. 10–23, November 1997.
- [10] “Internet traffic Classification and Categories.” http://en.wikipedia.org/wiki/Traffic_classification.

-
- [11] T. Bonald and Laurent Massoulie, “Impact of fairness on internet performance,” in *SIGMETRICS '01 Proceedings of the 2001 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pp. 82–91, June 2001.
 - [12] A. Dainotti, A. Perscape, and Giorgio Ventre, “Packet level characterization of network traffic,” in *IEEE Computer-Aided Modeling, Analysis and Design of Communication Links and Networks, 11th International Workshop*, pp. 38–45, 2006.
 - [13] B. Chandrasekar, “Survey of network traffic models,” tech. rep., Washington University, St. Louis, 2006.
 - [14] “Traffic generation model.” http://en.wikipedia.org/wiki/Traffic_generation_model.
 - [15] Ju Fang, Jie Yang, and Heng Liu, “Analysis of self-similar traffic based on the on/off model,” in *IEEE, Chaos-Fractals Theories and Applications, 2009. IWCFTA '09. International Workshop*, pp. 301–304, November 2009.
 - [16] Fan Zhang, Wenbo He, Xue Liu, and Patrick Bridges, “Inferring users’ online activities through traffic analysis,” in *WiSec*, June 2011.
 - [17] “WEP encryption standard.” http://en.wikipedia.org/wiki/Wired_Equivalent_Privacy.
 - [18] “CCMP protocol.” <http://en.wikipedia.org/wiki/CCMP>.
 - [19] “Employee account hacked for credentials.” <http://krebsonsecurity.com/2015/04/sendgrid-employee-account-hacked-used-to-steal-customer-credentials/>.
 - [20] “ebay security breach.” <http://www.forbes.com/sites/gordonkelly/2014/05/21/ebay-suffers-massive-security-breach-all-users-must-their-change-passwords/>.
 - [21] “US state department hacking.” <http://www.cnn.com/2015/03/10/politics/state-department-hack-worst-ever/index.html>.
 - [22] “Phishing attacks statistics report,” tech. rep., APWG, 2014.

-
- [23] A. Ghosh, Rittwik Jana, Ramaswami, Jim Rowland, and Shankaranarayanan, “Modelling and Characterization of Large Scale Wi-Fi traffic in Public Hotspots,” in *IEEE, Infocom*, 2011.
 - [24] Balachandran, Voelker, Bahl, and Rangan, “Characterizing user behavior and network performance in a public wireless lan,” in *SIGMETRICS*, pp. 195–205, June 2002.
 - [25] Divgi and Chlebus, “User and traffic characteristics of a comercial nation wide wifi hotspot network,” in *PIMRC*, pp. 1–5, September 2007.
 - [26] M. Waliullah and D. Gan, “Wireless LAN security threats and vulnerabilities,” in *IJACSA*, 2014.
 - [27] Dhamija, Tygar, and Hearst, “Why Phishing Works,” in *Human Factors in computing systems*, 2006.
 - [28] C. Dahiya and R. Garg, “Phishing and Anti-Phishing techniques : Case Study,” in *IJARCSSE*, 2013.
 - [29] Engin Kirda and Christopher Kruegel, “Protecting Users against phishing attacks using AntiPhish.”.
 - [30] “Wireshark user documentation.” <https://www.wireshark.org/docs/>.
 - [31] “HTTPS encryption on the web.” <https://transparencyreport.google.com/https/top-sites>.
 - [32] “Security implications in FTP.” http://en.wikipedia.org/wiki/File_Transfer_Protocol#Security.
 - [33] “Security implications in Telnet.” http://www.sans.org/security-resources/idfaq/telnet_rlogin.php.
 - [34] “Security implications in SMTP.” http://www.tcpipguide.com/free/t_SMTPSecurityIssues.htm.
 - [35] “Security vulnerability in Apache server.” http://httpd.apache.org/security/vulnerabilities_22.html.

-
- [36] “Security threat to IIS server.” http://www.theregister.co.uk/2009/05/18/iis6_file_pilfering_bug.
 - [37] “Documented Lighttpd web server vulnerabilities.” http://www.cvedetails.com/vulnerability-list/vendor_id-2713/Lighttpd.html.
 - [38] “DNS forgery.” <http://linuxgazette.net/153/misc/moen/chargen.txt>.
 - [39] “DNS randomness test.” <https://www.dns-oarc.net/oarc/services/porttest>.
 - [40] “TCP sequence number prediction.” http://insecure.org/stf/tcpip_smb.txt.
 - [41] “Defending against tcp sequence number attacks.” <http://tools.ietf.org/html/rfc1948>.
 - [42] “Linux kernel security vulnerability.” <http://www.securityfocus.com/bid/580/discuss>.
 - [43] “ICMP echo security exploits.” http://www.sans.org/security-resources/idfaq/icmp_misuse.php.
 - [44] “Cyber security awareness month.” <https://www.publicsafety.gc.ca/cnt/ntnl-scrt/cbr-scrt/csm-en.aspx>.
 - [45] “The stop.think.connect campaign.” <https://www.dhs.gov/stopthinkconnect>.
 - [46] “TOR for privacy online.” <https://www.torproject.org>.