Smartphone traffic characteristics and context dependencies

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Shahriar Kaisar

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

Abstract

Smartphone traffic contributes a considerable amount to Internet traffic. The increasing popularity of smartphones in recent reports suggests that smartphone traffic has been growing 10 times faster than traffic generated from fixed networks. However, little is known about the characteristics of smartphone traffic. A few recent studies have analyzed smartphone traffic and given some insight into its characteristics. However, many questions remain inadequately answered. This thesis analyzes traffic characteristics and explores some important issues related to smartphone traffic. An application on the Android platform was developed to capture network traffic. A user study was then conducted where 39 participants were given HTC Magic phones with data collection applications installed for 37 days. The collected data was analyzed to understand the workload characteristics of smartphone traffic and study the relationship between participant contexts and smartphone usage.

The collected dataset suggests that even in a small group of participants a variety of very different smartphone usage patterns occur. Participants accessed different types of Internet content at different times and under different circumstances. Differences between the usage of Wi-Fi and cellular networks for individual participants are observed. Download-intensive activities occurred more frequently over Wi-Fi networks.

Dependencies between smartphone usage and context (where they are, who they are with, at what time, and over which physical interface) are investigated in this work. Strong location dependencies on an aggregate and individual user level are found. Potential relationships between times of the day and access patterns are investigated. A time-of-day dependent access pattern is observed for some participants. Potential relationships between movement and proximity to other users and smartphone usage are also investigated. The collected data suggests that moving participants used map applications more. Participants generated more traffic and primarily downloaded apps when they were alone. The analyses performed in this thesis improve basic understanding and knowledge of smartphone use in different scenarios.

Acknowledgements

I would like to take this opportunity to thank and express my gratitude to the people who helped me and made the successful completion of this thesis possible.

First and foremost, I would like to express my genuine gratitude and sincere appreciation to my supervisors Dr. Derek Eager and Dr. Kevin Stanley who helped me all the way from the beginning of my study at U. of S. When I started my program at U. of S., I had very little idea about doing research. Both of my supervisors guided me in the right directions from the beginning of my program and helped me understanding the art of research by explaining everything in details and even answering my silly questions without being annoyed. They provided me with lots of support, encouragement, motivation and ideas. They extended their helping hands whenever I needed any suggestions. I also appreciate their efforts in correcting my thesis which took a lot of their valuable time and yet they were patient. I could not have imagined having a better guidance for my M.Sc. study.

Besides my supervisors I would like to thank the rest of the members of my thesis committee: Dr. Dwight Makaroff, Dr. Tony kusalik, Dr. Anh van Dinh for their suggestions and insightful comments.

My sincere thanks to my lab mates specially Mohammad S. Hashemian who helped me a lot during the whole period of my M.Sc. program. He along with Dylan knowles developed 'iEpi' application which accompanied data collection for the user study. They provided me with lots of support and suggestions during the development of TryTcp application for which I am very grateful to them.

Special thanks to Dr. Nathaniel Osgood for being one of the collaborators of the user study and providing necessary supports.

I gratefully acknowledge Martin Arlitt (University of Calgary) and Niklas Carlsson (Linkoping University) who provided me with lots of support, interesting ideas and thoughtful comments.

I am very thankful to my room mates and friends here in Saskatoon who provided unconditional support and encouragement for the successful completion of my thesis.

I am grateful to the participants who participated in our user study and helped us collecting data.

Last but not the least, I would like to express my sincere gratitude to my parents and siblings who were always there for me. I would have been lost without them.

Contents

Permission to Use i

Abstract ii

Acknowledgements iii

Contents iv

List of Tables vi

List of Figures vii

List of Abbreviations ix

1 Introduction 1
- 1.1 Smartphones . . . 1
  - 1.1.1 Smartphone Growth . . . 2
- 1.2 Smartphone Traffic . . . 3
- 1.3 Thesis Motivation . . . 5
- 1.4 Thesis Objectives . . . 6
- 1.5 Thesis Findings . . . 6
- 1.6 Thesis Organization . . . 7

2 Related Work 8
- 2.1 Internet Traffic . . . 8
- 2.2 Web Traffic . . . 10
- 2.3 Cellular Traffic . . . 11
- 2.4 Traffic Characteristics of Smartphones . . . 13
- 2.5 Application Usage of Smartphones . . . 15

3 Data Collection and Analysis 18
- 3.1 Participants and Duration of the Study . . . 18
- 3.2 Hardware Platform . . . 19
- 3.3 Software Platform . . . 20
  - 3.3.1 Healthlogger . . . 20
  - 3.3.2 TryTcp . . . 22
  - 3.3.3 LogActivity . . . 24
- 3.4 Limitations of Data Set . . . 25
- 3.5 Analysis Tools . . . 25

4 Workload Characteristics 27
- 4.1 Traffic Characteristics . . . 27
- 4.2 Application Use . . . 36
  - 4.2.1 Popular Applications . . . 37
  - 4.2.2 Top Users . . . 40
- 4.3 Popularity Analysis . . . 43
  - 4.3.1 Popular Websites . . . 43
  - 4.3.2 Top Users . . . 50

5 Context Dependencies 58

5.1 Network Dependency . . . 58
5.2 Location Dependency . . . 63
5.3 Time-of-Day Dependency . . . 72
5.4 Movement Dependency . . . 74
5.5 Proximity Dependency . . . 80

6 Summary and Conclusion **85**
6.1 Thesis Summary . . . 85
6.2 Thesis Contribution . . . 86
6.3 Discussion . . . 87
6.4 Future Work . . . 88
6.5 Conclusion . . . 89

References **90**

List of Tables

1.1 Smartphone Sales Categorized by OS (Thousands of Units)[13] 3

3.1 Specification of HTC Magic Phone(PVT 32B)[1] 20
3.2 Data Collected by Healthlogger Application 21

4.1 Highlights of the Study 28
4.2 Breakdown of Traffic by Ports 33
4.3 Breakdown of Wi-Fi Traffic by Content Type 36
4.4 Breakdown of Cellular Traffic by Content Type 37
4.5 Top Application Users 40
4.6 Top Ranked Users 51

5.1 Participants at Different Locations 64
5.2 Participants at Different State 75
5.3 Participants at Different Scenario 81

List of Figures

1.1 Smartphone Traffic by OS[18] 4

3.1 Components of TryTcp 22

4.1 Traffic at Different Hours of the Day 28
4.2 Traffic at Different Days of the Week 29
4.3 Wi-Fi Traffic in Different Weeks of the Study 29
4.4 Cellular Traffic in Different Weeks of the Study 30
4.5 Request Generation by Participants 31
4.6 Data Transfer by Participants 31
4.7 Traffic Generated by Individual Participants 32
4.8 Ratio of Wi-Fi to Collected Traffic 32
4.9 CDF of Transfer Size 33
4.10 HTTP Methods for Wi-Fi and Cellular Traffic 34
4.11 HTTP Content Type for Wi-Fi and Cellular Traffic 35
4.12 Relative Popularity of Applications 38
4.13 Relative Popularity of Application Categories 39
4.14 Applications Used Over Different Networks (Requests) 39
4.15 Applications Used Over Different Networks (Bytes) 40
4.16 Application Usage of Participant P31 (Application Usage Rank=1) 41
4.17 Application Usage of Participant P13 (Application Usage Rank=2) 41
4.18 Application Usage of Participant P24 (application usage rank=3) 42
4.19 Application Usage of Participant P30 (application usage rank=4) 42
4.20 Application Usage of Participant P29 (application usage rank=5) 43
4.21 Popularity Distribution of Web Pages 44
4.22 Top 10 Websites by Hits Accessed Using Wi-Fi Networks 44
4.23 Top 10 Websites by Request Byte(%) Accessed Using Wi-Fi Networks 45
4.24 Top 10 Websites by Response Byte(%) Accessed Using Networks 45
4.25 Categories of Websites by Requests Accessed Using Wi-Fi Networks 46
4.26 Categories of Websites by Request Byte (%) Accessed Using Wi-Fi Networks 46
4.27 Categories of Websites by Response Byte (%) Accessed Using Wi-Fi Networks 47
4.28 Top 10 Websites by Requests Accessed Using the Cellular Network 47
4.29 Top 10 Websites by Request byte(%) Accessed Using the Cellular Network 48
4.30 Top 10 Websites by Response Byte (%) Accessed Using the Cellular Network 48
4.31 Categories of Websites by Requests Accessed Using the Cellular Network 49
4.32 Categories of Websites by Request Byte (%) Accessed Using the Cellular Network . . 49
4.33 Categories of Websites by Response Byte (%) Accessed Using the Cellular Network . 50
4.34 Websites Accessed by Participant P24 Over Wi-Fi Networks(Wi-Fi Network Rank=1) 52
4.35 Websites Accessed by P23 Over Wi-Fi Networks (Wi-Fi Network Rank=2) 52
4.36 Websites Accessed by Participant P31 Over Wi-Fi Networks (Wi-Fi Network Rank=3) 53
4.37 Websites Accessed by Participant P30 Over Wi-Fi Networks (Wi-Fi Network Rank=4) 53
4.38 Websites Accessed by Participant P12 Over Wi-Fi Networks (Wi-Fi Network Rank=5) 54
4.39 Websites Accessed by Participant P30 Over the Cellular Network (Cellular Network Rank=1) 55
4.40 Websites Accessed by P25 Over the Cellular Network (Cellular Network Rank=2) . 55
4.41 Websites Accessed by Participant P7 Over the Cellular Network (Cellular Network Rank=3) 56
4.42 Websites Accessed by Participant P32 Over the Cellular Network (Cellular Network Rank=4) 56

4.43 Websites Accessed by Participant P16 Over the Cellular Network (Cellular Network Rank=5) . . . 57

5.1 Data Transfer for Top Hosts Accessed Using Wi-Fi Networks . . . 59
5.2 Data Transfer for Top Hosts Accessed Using the Cellular Network . . . 59
5.3 Data Transfer for Top Hosts for Participant P24 (Wi-Fi Rank=1, Cellular Rank=7) 60
5.4 Data Transfer for Top Hosts for Participant P23 (Wi-Fi Rank=2, Cellular Rank=25) 60
5.5 Data Transfer for Top Hosts for Participant P30 (Wi-Fi Rank=4, Cellular Rank=1) 61
5.6 Data Transfer for Top Hosts for Participant P12 (Wi-Fi Rank=5, Cellular Rank=28) 61
5.7 Data Transfer for Top Hosts for Participant P25 (Wi-Fi Rank=9, Cellular Rank=2) 62
5.8 Data Transfer for Top Hosts for Participant P7 (Wi-Fi Rank=15, Cellular Rank=3) 62
5.9 Data Transfer for Top Hosts for Participant P32 (Wi-Fi Rank=10, Cellular Rank=4) 63
5.10 Data Transfer for Top Hosts for Participant P16 (Wi-Fi Rank=12, Cellular Rank=5) 63
5.11 Data Transfer at Different Times of the Day . . . 64
5.12 Data Transfer at Different Days of the Week . . . 65
5.13 Content Type for Wi-Fi Traffic at Different Locations . . . 65
5.14 Top Websites for Wi-Fi Traffic at Different Locations . . . 66
5.15 Website Categories for Wi-Fi traffic at Different Locations . . . 66
5.16 Data Transfer for Top Hosts for Participant P24 (Wi-Fi Rank=1) . . . 67
5.17 Data Transfer for Top Hosts for Participant P23 (Wi-Fi Rank=2) . . . 68
5.18 Websites Accessed on Wi-Fi Networks by Participant P23 at Different Locations . . 68
5.19 Data Transfer for Top Hosts for Participant P31 (Wi-Fi Rank=3) . . . 69
5.20 Websites Accessed On Wi-Fi Networks by Participant P31 at Different Locations . . 69
5.21 Data Transfer for Top Hosts for Participant P30 (Wi-Fi Rank=4) . . . 70
5.22 Websites Accessed On Wi-Fi Networks by Participant P30 at Different Locations . . 70
5.23 Data Transfer for Top Hosts for Participant P12 (Wi-Fi Rank=5) . . . 71
5.24 Websites Accessed On Wi-Fi Networks by Participant P12 at Different Locations . . 71
5.25 Facebook Access Pattern of Participant P30 (participant accessing facebook on the most days) . . . 72
5.26 A Closer Look at Facebook Access Pattern of Participant P30 . . . 73
5.27 News Reading Pattern of Participant P23 (top visitor to news websites) . . . 74
5.28 A Closer Look at News Reading Pattern of Participant P23 . . . 74
5.29 Content Type for Cellular Traffic . . . 75
5.30 Top Website for Cellular Traffic . . . 76
5.31 Website Categories for Cellular Traffic . . . 76
5.32 Relative Popularity (response byte %) of Website Categories for Cellular Traffic . . . 77
5.33 Websites Accessed by Participant P30 on the Cellular Network (Cellular Rank=1) . 77
5.34 Relative Popularity of Websites (response byte %) for Participant P30 on the Cellular Network . . . 78
5.35 Websites Accessed by Participant P25 on the Cellular Network (Cellular Rank=2) . 78
5.36 Relative Popularity of Websites (response byte %) for Participant P25 on the Cellular Network . . . 79
5.37 Websites Accessed by Participant P7 on the Cellular Network (Cellular Rank=3) . . 79
5.38 Relative Popularity of Websites (response byte %) for Participant P7 on the Cellular Network . . . 80
5.39 Content Type for Wi-Fi Users . . . 81
5.40 Top Websites for Wi-Fi Users, Usage Categorized by Proximity . . . 82
5.41 Website Categories for Wi-Fi Users, Usage Categorized by Proximity . . . 82
5.42 Websites Accessed by Participant P24 (Wi-Fi Rank=1), Usage Categorized by Proximity . . . 83
5.43 Websites Accessed by Participant P23 (Wi-Fi Rank=2), Usage Categorized by Proximity . . . 83
5.44 Websites Accessed by Participant P31 (Wi-Fi Rank=3), Usage Categorized by Proximity . . . 84

List of Abbreviations

AP	Access Point
BSSID	Basic Service Set Identifier
CDF	Cumulative Distribution Function
FTP	File Transfer Protocol
GPS	Global Positioning System
GUI	Graphical User Interface
HTTP	Hyper Text Transfer Protocol
HTTPS	Hyper Text Transfer Protocol Secure
ICMP	Internet Control Message Protocol
IGMP	Internet Group Message Protocol
IM	Instant Message
IP	Internet Protocol
LAN	Local Area Network
MAC	Media Access Control
MB	Megabyte
MHD	Mobile Handhold Device
OS	Operating System
P2P	Peer-to-Peer
PDA	Personal Digital Assistant
RSSI	Received Signal Strength Indicator
SD Card	Secure Digital Card
SIM	Subscriber Identification Module
SMS	Short Message Service
SSID	Service Set IDentifier
TTL	Time To Live
U. of S.	University of Saskatchewan
URL	Uniform Resource Locator
USB	Universal Serial Bus
WAP	Wireless Application Protocol
WWW	World Wide Web

Chapter 1

Introduction

The popularity of smartphones is increasing among people all over the world. Basic mobile services are now used by 4 billion subscribers worldwide [46]. A significant percentage of these mobile subscribers already have a smartphone or are planning to get one. Recent studies have shown that a significant percentage of Internet traffic is generated from these smartphones [12, 13]. However, little is known about the characteristics of the traffic generated from smartphones. Some recent studies have tried to analyze the characteristics of smartphone traffic to shed light on the topic. However, unanswered questions in this area remain. This study investigates the basic characteristics of smartphone traffic, identifying and understanding the impact of context (where they are, who they are with, at what time, and over which physical interface) on smartphone usage, and focusing on what smartphone users do in different circumstances.

1.1 Smartphones

A smartphone is a cellular phone which facilitates Internet usage. Smartphones typically come with a pre-installed OS and some built-in applications. They have the capability of running a large number of different applications. They provide support for applications such as voice calling, text messaging, image and video capture and web browsing. Smartphones provide greater computing capability and connectivity than their contemporary feature phones. This section describes the growth and development of smartphones.

The first smartphone, the IBM Simon, was designed in 1992 and shown as a concept product at COMDEX in Las Vegas [39]. Although IBM Simon was the first smartphone, it had applications such as calendar, address book, clock, calculator, notepad and e-mail. It also facilitated sending and receiving of fax messages. The user interface was a touch screen which caught the attractions of the users.[1] Nokia released their first smartphone Nokia 9000 in 1996.[2] Ericsson used the term 'Smartphone' for the first time in 1997 when they introduced their concept phone GS88.[3] The

Nokia 9210 was the first color screen model introduced in 2000.[4]

The Ercisson R380 was the first device to be marketed as a 'Smartphone' in 2000.[5] Popular Science magazine declared Ericsson R380 as one of the most important advances in science and technology [41]. The Kyocera 6035 was introduced in the US market by Palm Incorporated in 2001. It contained features of a personal digital assistant (PDA) and supported limited web browsing.[6] The first Blackberry was introduced in 2002 by RIM and supported wireless email use.[7]

Apple introduced its popular iPhone in 2007. The first generation iPhones were very costly and could not execute multi-tasked native applications. However, the phones gained popularity because they provided an excellent web browser, outstanding music quality and camera. The first generation iPhones also supported a multi-touch screen and a virtual keyboard instead of a physical one. Apple Inc. introduced the second generation iPhone in 2008 and provided 3G cellular support. In 2009 Apple introduced the iPhone 3GS, with a compass, faster processor, high resolution camera and video recording. The iPhone 4 with more enhanced features was released in 2010.[8]

Google released the Android operating system for smartphones in 2008.[9] Android is a software stack which includes an operating system, middleware and some applications. Android was first used by HTC Dream phones which were distributed by T-mobile. The phones came with built-in applications such as maps, calendar, e-mail and a web browser. From the beginning, Android supported multi-tasking and execution of native applications.[10] Third-party applications for Android are also available through Android Market. By 2010 Android had become the second most popular smartphone OS worldwide.

1.1.1 Smartphone Growth

The popularity of smartphones is growing day by day. Many people all over the world are using these phones to communicate, obtain information and for entertainment. Since the first smartphones were released several new OSes have been introduced by different companies and the features of smartphones have become more attractive.

The primary reasons behind the increased popularity of the smartphones are the powerful processors and graphics processing units, significant amounts of flash memory, high-resolution screens with multi-touch capability and diverse applications. Currently, smartphones are receiving attention from handset manufacturers, network operators and application developers. Users are discov-

[1] http://web.archive.org/web/19990221174856/byte.com/art/9412/sec11/art3.htm, access date=13/10/2011
[2] http://press.nokia.com/1996/08/15/, access date=13/10/2011
[3] http://pws.prserv.net/Eri_ no_ m oto/GS88_ Preview.htm , access date=13/10/2011
[4] http://press.nokia.com/2000/11/21/, access date=13/10/2011
[5] http://www.mobilemag.com/2001/09/25/ericsson-introduces-the-new-r380e/, access date=13/10/2011
[6] http://www.palminfocenter.com/view_ story.asp?ID=1707, access date=13/10/2011
[7] http://us.blackberry.com/newsroom/news/press/release.jsp?id=640, access date=13/10/2011
[8] http://www.tekserve.com/service/mac-faq/other/iphone-model-generation.php, access date=13/10/2011
[9] http://www.webcitation.org/5wiy036ap, access date=13/10/2011
[10] http://android-developers.blogspot.com/2008/09/, access date=13/10/2011

ering that smartphones can act as a personal computing device enabling access to the web and large numbers of applications, besides voice and messaging services.

ComScore published a report in early 2010 stating that over 45.5 million people in the United States owned a smartphones out of 234 million total subscribers.[11] Berg Insight published a recent report stating that global shipments of smartphones increased 74% in 2010. In total, 295 million units were shipped during 2010. The global user base of smartphones increased by 38% in 2010 and by the end of 2010 there were 470 million active users worldwide. By 2015 the global user base of smartphones is forecast to reach 2.8 billion at an annual growth rate of 42.9%.[12] Gartner presented a report which categorized smartphone sales according to different OSes. Android accounted for 25.5% of the market share in the third quarter of 2010. Table 1.1 summarizes some statistics from the report.[13]

Table 1.1: Smartphone Sales Categorized by OS (Thousands of Units)[13]

Company/ OS	**3Q 2010 units**	**3Q10 Market Share(%)**	**3Q 2009 units**	**3Q09 Market Share(%)**
Symbian	29,480.1	36.6	18,314.8	44.6
Android	20,500.0	25.5	1,424.5	3.5
iPhone	13,484.4	16.7	7,040.4	17.1
RIM	11,908.3	14.8	8,522.7	20.7
Windows mobile	2,247.9	2.8	3,259.9	7.9
Linux	1,697.1	2.1	1,918.5	4.7
Others	1,214.8	1.5	612.5	1.5
Total	**80,532.6**	**100**	**41,093.3**	**100**

1.2 Smartphone Traffic

Network traffic generated from smartphones is a more recent addition to Internet traffic. Along with an increased number of smartphone subscribers, network traffic generated from smartphones is also increasing. Nokia Siemens published a report stating that cellular traffic is forecast to grow 10 times faster than fixed Internet traffic.[14] Ericsson confirmed that most of the cellular traffic is generated from smartphones and network traffic is actually surpassing voice traffic.[15]

[11] http://www.comscore.com/Press_ Events/Press_ Releases/2011/1/comScore_ Reports_ November_ 2010_ U.S._ Mobile_ Subscriber_ Market_ Share , access date=18/8/2011
[12] http://www.bgr.com/2011/03/10/berg-Smartphone-shipments-grew-74-in-2010/ , access date=18/8/2011
[13] http://www.gartner.com/it/page.jsp?id=1466313, access date=18/8/2011
[14] http://www.totaltele.com/view.aspx?id=448681 , access date= 18/08/2011
[15] http://www.ericsson.com/thecompany/press/releases/2010/03/1396928, access date=18/8/2011

AdMob is a mobile advertising company. AdMob offers advertising solutions for different mobile platforms, including Android, iOS, webOS, Flash Lite, Windows Phone 7 and all standard mobile web browsers. They serve ads for more than 23,000 mobile websites and applications around the world. They store and analyze data from ad requests and clicks.[16] They publish statistics of cellular traffic identifying information about the device that generated the request. A report published by AdMob in February 2009[17] categorized traffic generated from handsets into three categories: mobile Internet device, smartphone and feature phone. iPods and gaming devices were considered as mobile Internet devices and normal cell phones were considered feature phones. According to the report, smartphone traffic increased from 35% to 48% of the total traffic from hand-held devices within one year and feature phone traffic went down from 58% to 35% during the same period of time. Traffic generated from mobile Internet devices also increased from 7% to 17%. This report confirms the increasing popularity of smartphones over feature phones.

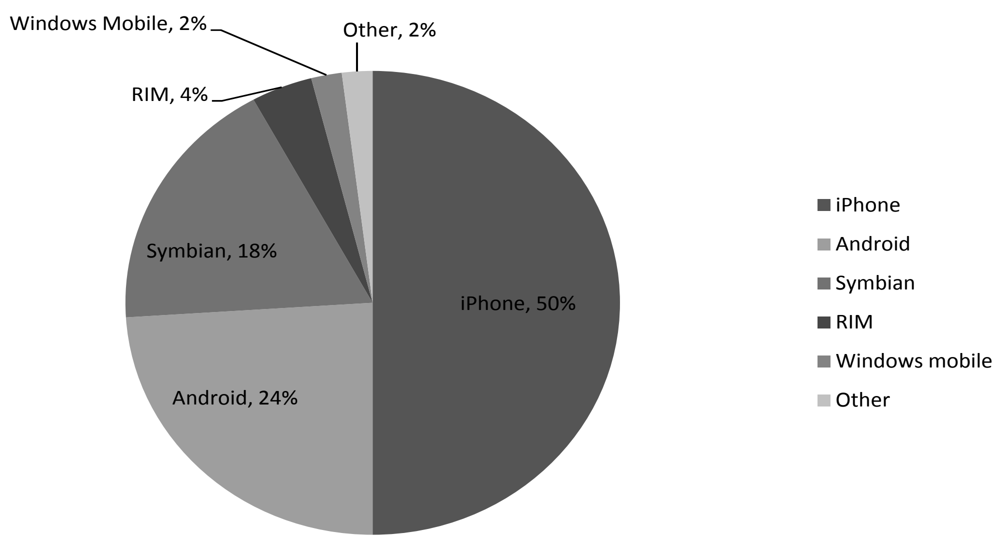


Figure 1.1: Smartphone Traffic by OS[18]

A report published by AdMob in May 2010[18] shows that smartphones generated 46% of the total traffic from hand-held devices. The report categorized smartphone traffic by operating system and showed that Apple's iPhone and Android generated around 74% of smartphone traffic. Figure 1.1 shows the breakdown of smartphone traffic by OS.[18]

[16]http://www.admob.com/,access date=18/8/2011
[17]http://tech.fortune.cnn.com/2010/03/25/Smartphone-traffic-is-up-193-in-a-year/,access date=18/8/2011
[18]http://metrics.admob.com/wp-content/uploads/2010/06/May-2010-AdMob-Mobile-Metrics-Highlights.pdf, access date=18/8/2011

1.3 Thesis Motivation

Smartphones have received significant attention from researchers recently. A number of studies have been performed to characterize usage of smartphones in different circumstances [13, 31]. However, there are many unanswered questions related to smartphone use.

Researchers have not tried to isolate Wi-Fi and cellular traffic and compare their usage patterns on an individual user basis in any of the previous work. This thesis compares traffic generated over Wi-Fi and cellular networks. Such comparisons are helpful for predicting usage patterns in a mixed Wi-Fi/cellular service context. Many cellular providers are planning to develop such mixed networks to address capacity limitations of cellular networks.

Caching has been largely investigated for desktop and laptop traffic [7, 23, 48]. One recent study has also investigated the idea of caching over 3G networks [10]. However, researchers collected data from backbone networks and did not analyze usage patterns on an individual level. For this thesis research, data was collected from an on-phone network logger and individual usage patterns were captured, so as to understand individual participant characteristics. Such analyses will facilitate implementation of better caching and prefetching decisions based on individual usage patterns.

One of the important motivations of this work is to understand the impact of location on smartphone usage. Of interest is whether participants change their usage patterns when they are in different locations. The possible existence of location dependencies could be exploited by prefetching or caching techniques at access points, gateways or proxy servers within Wi-Fi or cellular networks.

Another motivation of this work concerns the impact of times of the day on smartphone usage. Understanding time-of-day dependencies and access patterns could be useful for resource provisioning by the service providers. It would enable the possibility of developing intelligent techniques for reducing response times for individual users.

This thesis also investigates possible relationships between user proximity and smartphone usage patterns. Such analysis is useful for improving basic understanding about smartphone usage in different circumstances. Of interest is whether smartphones are primarily used when ‘alone’ or are they frequently incorporated as part of an in-person socializing context. Such basic understanding could help in understanding location dependencies in browsing (e.g. do participants browse more in locations where they are likely to be ‘alone’) and possibly understand the types of applications participants prefer in different scenarios.

Relationships between participant movement and smartphone usage are also studied, so as to understand usage patterns when moving or stationary. Such analysis helps to understand the impact of movement, for example commuting, on smartphone usage.

1.4 Thesis Objectives

This thesis analyzes the characteristics of smartphone traffic over Wi-Fi and cellular networks. The thesis explores the following aspects of smartphone usage.

- Differences between the usage of the Wi-Fi and the cellular network with respect to the amount of traffic, website diversity and media types particularly at an individual user level.
- Relationships between smartphone usage and locations by determining whether smartphone usage changes on aggregate and/or the individual user level when participants are in different locations.
- Dependencies on times of the day.
- Relationships between participants' movements and their usage patterns, in particular whether usage patterns change while commuting.
- Dependencies between smartphone usage and proximity to other people.

1.5 Thesis Findings

Important findings of this thesis include:

- Heterogeneity exists even in a small group of participants. Different participants preferred performing substantially different sets of activities over the Internet via smartphones.
- Network dependencies are observed in individuals' smartphone usage. Participants preferred performing some tasks over Wi-Fi networks and some others over the cellular network.
- Location dependencies are observed at both the aggregate and individual user levels. Participants did not prefer watching videos at work or school. Location-dependent usage patterns (e.g. accessing some set of websites at home and others at work) are also observed at the individual level.
- Participants preferred to access some sets of websites (e.g. social networking or news websites) at particular times.
- Moving participants used map applications more. However, participants generated more traffic while stationary.
- Overall, participants generated more traffic when they were alone compared to when they were in proximity of other people. Some participants changed their Internet usage patterns in the presence of other people while others did not.

1.6 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 describes related work providing background knowledge about Internet, web, cellular and smartphone traffic. Chapter 3 discusses the data collection procedure and components of the data collection applications. Chapter 4 focuses on the basic characteristics of smartphone traffic and application usage patterns. Chapter 5 explores the relationship between user context and smartphone usage. Finally, Chapter 6 summarizes the thesis and discusses some possible future directions.

Chapter 2

Related Work

The Internet is a key medium for personal and commercial communication. It has expanded exponentially over the last 20 years. The nature and characteristics of Internet traffic is also changing with the introduction of new technologies and applications. Initially, people used the Internet for communication and commercial purposes. However, the introduction of peer-to-peer technology changed the growth and trends of Internet traffic [3]. Recently, user-generated content has become very popular. Web 2.0 applications such as youtube, flickr and facebook have gained huge popularity and contribute significantly to Internet traffic [32]. The introduction of cell phones capable of accessing the Internet through Wi-Fi, 3G, 4G and Edge networks have added new dimensions to Internet traffic.

2.1 Internet Traffic

Kotz *et al.* [28] conducted a study at Dartmouth college to understand the characteristics of traffic generated in the campus-wide wireless network. They collected data from 476 access points starting in Fall 2001 and continuing for 77 days in total and showed the presence of 1706 distinct network cards. They classified traffic according to different days of the week and different times of the day. The amount of data was more or less equal for Tuesday through Friday and Saturday. However, there was a peak on Mondays and less usage on Sundays. The authors could not find any particular reason for the presence of a peak on Mondays. However, the lower traffic on Sundays was a weekend effect as described by the authors. The amount of traffic decreased during the night and increased during the day. The amount of traffic reached its peak around 10 AM in the morning, and then there was a slight decrease at 1 PM and it remained stable for the rest of the day. It started decreasing after midnight. The authors also categorized the IP traffic into TCP, UDP and other protocols. TCP contributed 97.5% of the traffic and UDP contributed 2.5% of the traffic according to bytes transferred. ICMP, IGMP and other protocols accounted for less than 1% of the traffic. Traffic was further categorized according to applications including HTTP (53%), netbios (3.2%), kazaa (2.6%), gnutella (1.8%), ftp (1.6%), AOL (1.5%) and icq (1%).

Li *et al.* [30] collected HTTP traffic for half an hour in late 2006 from a research institute's network and compared that with previously collected traffic in 2003. The authors found that over this 3 year period total traffic increased by 180%. Web browsing contributed 71.1% of the total traffic in 2003, while accounting for 52.71% of the total traffic in 2006. Other categories of traffic such as web apps, file download, advertising and e-mail rose sharply.

Internet traffic measurement and traffic characterization of broadband Internet connections were performed by Pries *et al.* [36]. Data was collected in 2008 which reflects the Internet usage of 250 households. The measurement lasted for 19 days and 400 GB of data was collected. The amount of traffic started decreasing at 1 am and reached a minimum at 6 am. It started to pick up and reached its maximum rate at 7 pm. The average session duration during weekdays was 129 minutes and 167 minutes during the weekend. Extensive usage of peer-to-peer traffic generated high session durations. Traffic was classified according to applications with peer-to-peer traffic contributing 40% of the total traffic, while web, streaming and instant messaging contributed 25%, 22% and 3% of the traffic respectively. Streaming traffic was further classified into HTTP-video (61%), flash video (31%) and HTTP audio (5%). Peer-to-peer traffic consisted of BitTorrent (56%), eDonkey (41%) and other torrent applications.

Borgant *et al.* [4] collected data for 15 minutes every day for seven years (2001-2008) on a trans-pacific backbone link. Over the 7 year period the authors found that TCP and UDP continuously generated more than 90% of the packets. ICMP generated around 5% of the packets during that period. Most of the traffic generated during that period was web traffic (40-60%). The second largest group was peer-to-peer traffic generating around 30% of the traffic. Common Internet services such as FTP, mail and news protocols accounted for around 5% of the traffic during that period and remained more or less stable.

Erman *et al.* [11] collected traffic from a US residential broadband provider with 100K subscribers from February 2007 to September 2008. They analyzed traffic from 550 million requests generated during that period with an average request size of 91 KB. The authors found that HTTP was the dominant protocol, accounting for the largest volume of traffic on the network. HTTP accounted for 41.5% of the traffic and HTTP multimedia (25.8%), file sharing (9%), multimedia (4.7%), games (1.4%), chat (0.6%) and mail (0.5%) were the other major contributors.

HTTP traffic from a research institute was collected from 2006 to 2009 and analyzed in [6]. The authors found that over this period, in most of the months the 'GET' method generated around 90% of the traffic. The 'GET' method is used by the client to request data from the server. The rest of the data were generated by the POST method which is used by clients for uploading data. The authors also found that across the study duration the distribution of requests per hostname was more or less the same.

2.2 Web Traffic

Web traffic contributes a major portion of Internet traffic. Researchers have performed extensive analysis to understand the characteristics and properties of web traffic to improve the response time of web requests and provide the users with a better experience of web surfing. Trends of web traffic have changed from time to time and with the introduction of new Web 2.0 technologies. Web traffic was originally dominated by text and image content. However, recent introduction of social networking, video sharing and file hosting sites have changed the characteristics and reshaped web traffic and access patterns. The introduction of the popular Ajax technologies has changed the way requests are being served and these technologies are responsible for a significant portion of web traffic.

Breslau *et al.* [5] performed extensive analysis on different datasets and found that the distribution of page requests follows a Zipf-like distribution where the relative probability of a request for the i^{th} most popular page is proportional to $1/\mathrm{i}^{\alpha}$ where α is less than 1. The authors measured different values of α for different traces. The authors also found a weak correlation between the access frequency of a web page and its size. They also reported a weak correlation between access frequency and a web page's rate of change. This research was done in 1999 and therefore the results are somewhat obsolete given that text-based websites are no longer the most popular websites.

Ihm *et al.* [22] collected traffic from a globally distributed proxy system in 2010 to shed light on the characteristics of modern web traffic. They found an increased popularity of javascript, css and xml content due to the extensive use of Ajax. The authors also found increased popularity of flash video. The popularity of non-flash videos was decreasing as reported by the authors. The authors also reported that search engines were the most popular sites used by people all over the world.

Differences between the traditional HTTP traffic and traffic generated from newly introduced Web 2.0 applications were investigated by Schneider *et al.* [40]. Unlike the traditional websites that support a request-response model where users generate requests and the server provide responses, Web 2.0 applications support asynchronous prefetching. Asynchronous prefetching refers to the technique where a user script automatically fetches data from the server and updates only a significant portion of the web page instead of reloading the whole web page. The authors found that Ajax applications including applications such as Google Maps and Google Mail generally transferred more bytes in a session than all other HTTP applications. They also found that transfer sizes were larger for Ajax applications than for other HTTP applications. Session lengths were also found to be longer for Ajax enabled applications.

Media content is contributing a significant portion of web traffic and has become popular among users recently. Distribution and access patterns of media content were analyzed by researchers in different settings. Gill *et al.* [16] collected video traffic from a campus network and reported that the access pattern follows a Zipf-like model. However, Cha *et al.* [8] performed a server-log based study and reported a significant deviation from Zipf. Guo *et al.* [17] also performed extensive analysis and reported that media access patterns followed a stretched exponential distribution instead of a Zipf-like distribution.

2.3 Cellular Traffic

The introduction of enhanced cellular phones has influenced the characteristics of Internet traffic. Currently, most of the cellular phones come with the ability to access the Internet through Wi-Fi, Edge, 3G or 4G networks. Customized websites have also been created for cellular access. Introduction of the Wireless Access Protocol (WAP) and WAP browsers have significantly improved the user experience. People can browse through websites from anywhere anytime or use different applications. Adoption and usage patterns of cellular phones have been investigated by researchers in different parts of the world [37, 18, 29].

Mobile browsing in Finland was investigated by Antero Kivi [27]. He investigated the early use of cell phones for accessing websites. He collected data with TCP/IP traffic measurements at the GSM/UMTS networks of three Finish mobile operators during 2005-2006. Data was also collected from 500-700 mobile subscribers who were using Symbian S60 mobile phones. He categorized hosts into different categories such as mobile operator, ICT company, News, Entertainment, Web Search/Portal and Adult content. The author concluded that in 2006 most of the traffic was generated by hosts in the mobile operator (28%), ICT Company (20%) and News (15%) categories. The rest of the traffic (37%) was generated by hosts in the Entertainment, Web search/portal, Adult and other content categories. A breakdown of traffic according to different days of the week and different hours of day was also presented. Traffic was more or less evenly distributed across different days of week. However, usage during Sundays was a little less than the other days of the week. Distribution of traffic at different hours of the day followed the natural intuitions. At night the usage was less than during the day. Usage increased around 8 AM in the morning, reached the peak around 1 PM and remained more or less stable for the rest of the day. It started decreasing after 11 PM. The author calculated the share of traffic generated by top websites. He showed that the top 5% of the domains accounted for 93% of the traffic and the top 10% of the domains were responsible for generating 96% of the traffic in 2006. The top domains included google.com (3.2%), mtv3.fi (3%), yle.fi (2.8%), nokia.com (2.7%) and nokia.fi (2.2%).

Cell phone Internet use in Japan was investigated by Ishii [24]. Japan possesses the highest diffusion rate of mobile Internet services. At the time of Ishii's srudy, several telecommunication companies were providing Internet access to mobile subscribers through 3G and other technologies. Around 40% of the internet users accessed the Internet through cell phones. The main use of mobile Internet was email. The diversity of the websites visited by people through cell phone was smaller compared to those visited through PCs. Similar conclusions have also been reported in [25]. Ishii also presented categories of popular websites accessed through cell phones [24]. According to the report, the most popular websites were search engines with 45.3% of accesses falling into this category. The rest of the popular website categories included weather, transportation information/maps, music and news which accounted for 38%, 29.6%, 29% and 25.5% of accesses respectively.

The ways people use the web on cell phones were investigated by Cui *et al.* [9]. They conducted a survey with 47 participants, and collected data using an in-device logger tool from 577 participants. From their analysis they concluded that people preferred to use Wi-Fi networks instead of the cellular network because Wi-Fi networks are cheaper and faster. The average session duration was much higher (4.5 minutes) for Wi-Fi traffic than for cellular traffic (2.4-2.6 minutes). The authors categorized activities performed using mobile device into 3 categories: information seeking, communication and content object handling. Information seeking referred to finding a particular piece of information about a product or place, collecting information from multiple sources for making decisions or gathering knowledge and casual web browsing for entertainment purposes. Communication included web mail or online communities. Content object handling referred to the act of manipulating digital content, such as obtaining ring tones, wall papers, sharing photos and other cellphone centric tasks. One of the interesting findings from their study is that people used mobile devices more to receive e-mails than to send replies (5%).

Adya *et al.* [1] studied and analyzed the user behavior on a particular web server specifically designed for cell phone users. The authors found that the distribution of document popularity did not follow a Zipf-like distribution but rather when graphed with a log-log plot exhibited three distinct linear regions. Most of the replies sent to clients were less than 3 KB. Sessions were relatively short-lived, and for 95% of the users the longest session length was less than 3 minutes. The browsing behavior for different categories such as stock quotes, news and yellow pages did not change between weekdays and weekends [1]. However, the amount of data accessed dropped by 45% during the weekends. Stock quotes, news and yellow pages were the most popular website categories among the users.

Researchers have also performed some comparisons between WAP traffic generated by cell phones and WWW traffic generated by desktop computers to understand their similarities and dissimilarities. Different conclusions have been made by different researchers. Halvey *et al.* [20] concluded from their study that mobile web surfing is similar to web surfing through desktop computers and both of them yield a Zipf-like distribution for web page popularity. However, Adya *et al.* [1] as mentioned previously stated that the distribution of document popularity does not follow a Zipf-like distribution for mobile web surfing. Kunz *et al.* [29] collected data from Bell

Mobility's access network for seven months. They also compared WAP traffic with WWW traffic and confirmed that some of the properties of WAP traffic were similar to those of WWW traffic: self-similarity and daily and weekly periodicity. However, there were some dissimilarities for session lengths and packet sizes which were generally smaller for WAP traffic.

2.4 Traffic Characteristics of Smartphones

Studies have been conducted to understand the characteristics of smartphone traffic in addition to cell phone traffic. Smartphones provide better Internet services through an enhanced interface and a diverse number of applications. Therefore, smartphone users usually generate more traffic than conventional cell phone users as discussed in Section 1.2. The nature and characteristics of the traffic generated from cell phones and smartphones are also possibly different as smartphones provide better user experience for Internet access.

Falaki *et al.* [14] collected traces from 255 smartphone users and performed a detailed traffic analysis. They used Android and Windows Mobile devices for collecting traces. The authors concentrated on the application usage, session characteristics and energy consumption. According to their analysis, users interacted with their smartphone 10 to 200 times a day on average with a mean session length of 10-250 seconds. The researchers concluded that although most of the sessions were short, long sessions also existed. Application usage was also reported. The authors presented the amount of traffic generated by and for smartphones during the study. The received traffic ranged from 1 to 1000 MB per day with a median value of 30 MB. The amount of traffic sent ranged from 0.3 to 100 MB with 5 MB being the mean. The study revealed many issues related to smartphone use. However, the users were given unlimited talk time, SMS and data plan, and therefore there is a possibility of bias in the results.

Network traffic, generated from smartphones, was also collected and analyzed by Maier *et al.* [31]. The authors collected traffic from 20,000 residential DSL lines. The collected traffic was filtered according to user-agent and IP TTL to obtain the traffic generated from mobile hand-held devices (MHDs). The collected traces are only able to capture the characteristics of smartphone use at home and while connected through Wi-Fi. Analysis revealed that most of the traffic generated from MHD's was HTTP traffic. HTTP contributed up to 80-97% of all the bytes generated from MHDs. The authors categorized HTTP traffic into multimedia, apps, browsing and xml. Multimedia consisted of the traffic generated from audio and video content. Apps referred to traffic generated from application download. Multimedia was the most voluminous content among all the categories and generated most of the traffic (around 40-60%). Application downloads generated around 15-20% of the traffic.

Falaki *et al.* [13] installed a custom logger tool in 43 phones and then distributed them among users to gather data and capture user behavior. One dataset, collected from 10 users, contained packet level traces of network traffic. A second dataset, collected from 33 users, contained application level traces of network traffic. They used Windows Mobile and Android OSes for performing their research experiments. They concluded from their experiment that people with Android phones tend to use their phone much more than people using Windows Mobile because of the user-friendliness of the Android platform. The authors found that more than half of the traffic was contributed by browsing while each of e-mail, media and maps contributed equally to the other half. The average traffic volume in their first data set was 2-20 MB per user per day and the average traffic volume in their second data set was 1-500 MB per user per day. The fraction of the traffic carried over Wi-Fi networks rather than cellular varied widely across users with a median of 0.5. However, 20% of the users did not use Wi-Fi at all, while another 20% of the users mostly used Wi-Fi to connect to the Internet. Therefore, smartphone studies based solely on measurement of Wi-Fi or cellular traffic will miss a significant portion of the traffic and its associated characteristics. The average downlink to uplink ratio was found to be 6:1. The authors presented a detailed breakdown of traffic according to ports and applications. Most of the traffic volume (bytes) was contributed by HTTPS (43.88%), HTTP (37.48%) and IMAP4S (15.21%). IMAP4S is primarily used for e-mail. An application-level breakdown of traffic was provided with browsing (58.02%), media (10.82%), messaging (10.33%) and maps (8.51%) contributing most of the traffic. The authors also measured the transfer size for smartphone traffic with the mean transfer size being 273 KB for downloads and 57 KB for uploads.

Challenges in deploying large scale smartphone user studies along with some of the results obtained from such a study were described by Oliver [33]. The study captured behavior of 17300 Blackberry users. On average, users interacted with their phone for 1.68 hours a day. However, the value was skewed by a small fraction of the users who interacted with their phone for several hours a day on average. The median interaction time was only 1.31 hours a day. Users interacted with their phone 86.9 times a day on average. This value was again skewed by people who interacted with their phone more often than others. The length of the sessions also varied widely among the results obtained. The median session length was 20 seconds and the mean session length was 68.4 seconds.

Gember *et al.* [15] collected traces from campus Wi-Fi networks to determine the differences between traffic from hand-held devices and that from other devices. They collected data from 1943 access points over two different networks for 3 days each during April-June, 2010. User-agent along with NIC MAC addresses were used to isolate hand-held device traffic from other traffic. The authors found several dissimilarities between hand-held device traffic and that from other devices. Most of the hand-held device traffic (97%) was web traffic. However, web traffic contributed to only 82% of the traffic generated by other devices. The amount of UDP traffic generated from hand-held device was very low (5.9%) as compared to that from other devices (25.7%). Traffic was categorized by protocols and ports. HTTP and HTTPS contributed to 91.1% - 97% of the traffic in two different traces for hand-held devices. E-mail contributed 1.51% of the traffic for the hand-held devices. TCP flow characteristics were also analyzed [15]. TCP flows were smaller (fewer bytes) and of shorter duration for hand-held devices than for other devices. Hand-held devices consumed proportionately more multimedia content with video accounting for 40% of the traffic for hand-held devices and 17% of the traffic for other devices. Googlevideo.com (35.4%), Pandora.com (18.2%), apple.com (10.57%) and facebook (2.45%) were the most popular hosts according to bytes transferred for hand-held devices.

Shafiq *et al.* [42] collected smartphone cellular traffic from a major USA state and found that the top 5% of the users generated almost 90% of the traffic and the top 10% of the applications accounted for more than 90% of the traffic. Schmiedl *et al.* [38] conducted a survey and concluded that 70% of the websites accessed through cell phones provided current information such as weather forecasts, news, transport schedules or general information (for example, Wikipedia), 17% of the websites were related to entertainment such as music or video and 13% of the websites were social networking sites such as facebook or twitter. However, another user study performed by Karlson *et. al.* [26] showed that people mostly used smartphones for reading e-mails.

Opera is a popular web browser for personal computers. Opera mini was launched for mobile phones in 2005. According to a report published by Opera Software in May 2011[1]the most popular websites among the cell phone users in India were google.com, facebook.com, youtube.com, vuclip.com and orkut.com. The same report showed that the most popular websites among the American users included google.com, facebook.com, youtube.com, wikipedia.org and my.opera.com. There were some variations between popular websites in different regions. However, in most of the countries the popular sites included google, facebook and youtube.

The previous work provides some intuition about smartphone traffic characteristics. However, some of the studies did not capture the behavior of traffic over both Wi-Fi and cellular networks. Falaki et al. [13] mentioned the percentages of Wi-Fi and cellular traffic present in their traces, but did not present any analysis to compare their characteristics. There could be significant differences in the ways that people use cellular and Wi-Fi networks. Identification and analysis of such differences could enable improved application or network design.

2.5 Application Usage of Smartphones

One of the main reasons behind the popularity of smartphones is the availability of diverse applications. iPhone and Android are more popular than other devices because of their app stores.

[1] http://media.opera.com/media/smw/2011/pdf/smw052011.pdf

Smartphone users download applications from the iPhone AppStore and Android Market for entertainment and other purposes [34]. Thousands of applications for both iPhone and Android devices are available. Software development kits (SDKs) have been developed and made available for both iPhone and Android so that application developers can easily create new applications. Researchers have performed some analyses to understand application use among smartphone users.

As described in Section 2.4, Falaki *et al.* [14] collected traces from 255 users and performed a detailed analysis. They used Android and Windows Mobile for collecting traces. The authors presented a detailed report about the application use of participants. 10 to 90 applications were used by each participant, with the mean being 50 applications throughout the whole study. Communication applications (applications used for messaging, SMS, and IM) were the most popular types of applications. Browsing applications and games were found to be the second most and third most popular application types, respectively.

Maier *et al.* [31] collected traffic trace from 20,000 residential DSL lines. They presented some results concerning the relative popularities of applications. The authors concluded that browsing applications were the most popular ones among users. Apple's web browser Safari was found to be the most popular application and 62% of all devices were using it. iTunes (up to 37%) and weather apps (up to 32%) were the second and third most popular applications respectively. For non-Apple devices they also found the browser to be the most popular application. The authors found that more than 70% of the applications were directly downloaded to mobile devices.

Shepard *et al.* [43] developed a custom logger tool for the iPhone. They were planning to deploy 25 phones for one year and collect information on usage. However, they have performed some initial deployment and published the outcome. They recruited 25 undergraduate students from Rice University to conduct the study. Participants were given an unlimited SMS and data plan and were asked to use these devices as their primary cell phone. The authors concentrated on application usage and performed detailed analyses to understand the patterns of application use of different users. The authors found that users preferred to use more applications initially and later restricted their use to only a small set of favorite applications. The authors also concluded that most of the popular applications were built-in applications such as safari, facebook, maps, and the app store. The researchers also mentioned that most of the TCP connections (around 48%) were short-lived and usually lasted 2 seconds or less.

Xu *et al.* [47] collected smartphone traffic from a tier-1 cellular network provider in the USA from August 24th, 2010 to August 30th, 2010. The authors used the HTTP user-agent field to identify the applications that generated smartphone traffic. A considerable number of popular applications (20%) were regional and served regional needs; such applications included news and radio applications. Usage of some applications was correlated with usage of other applications, which suggests that access to one application increases the probability of accessing related applications. Users were also frequently found to be using multiple applications for similar purposes (e.g. multiple news or bank apps). There was found to be considerable usage of social networking and games applications while users were moving. However, the authors used cell sector information for determining movement and locations which fail to provide precise information.

Hannu Verkassalo collected smartphone traffic during 2008-2009 from the MobiTrack global smartphone study and presented the results in [45]. The author found that 55% of all smartphone use was contributed by voice call and SMS. Internet browsing and multimedia applications contributed 14% and 15% of usage respectively. Popular applications downloaded from the app store included Adobe Reader, Navicore, Anti-Virus, Quick Office and Opera Mini.

Prior work has not addressed the question of whether there are differences in the applications people tend to use when Wi-Fi connected, versus cellular connected. As noted previously, identification and analysis of such differences could enable improved application or network design. Comparing application usage at home versus in the work place might also give some indication how users can be better served, for example through use of caching or prefetching policies.

Chapter 3

Data Collection and Analysis

Prior studies for analyzing smartphone usage have been conducted by various research groups in different environments, circumstances and with different groups of participants. However, most of the datasets were unable to capture activities over both cellular and Wi-Fi networks. Data collection through in-device logger tools captures network traffic over both networks. A dual purpose instrumentation and user study was conducted of which a part was for this thesis research and another part was for another project titled ‘iEPi’. Smartphone traffic collection was a part of the user study and a part of this thesis research. However, this thesis also made use of the data collected for the ‘iEPi’ project. The ‘iEPi’ project collected battery state, Bluetooth, Wi-Fi, GPS and accelerometer data from on-device sensors. An application was submitted to the behavioral research ethics board of the University of Saskatchewan presenting a summary and goals of the dual purpose user study. After the approval from the ethics board the study was begun.

A pilot study was conducted with 25 participants for 2 weeks before running the final user study. The pilot study helped to identify the potential bugs and problems of full-scale deployment of the system. No pilot study data is included in this analysis.

This chapter presents the data collection procedure, tools used for data collection, and tools used for analyzing data.

3.1 Participants and Duration of the Study

The user study began with 40 participants who were selected from the Computer Science Department of the University of Saskatchewan. However, one of the participants opted out after one week and the study continued with 39 participants. Therefore, data collected throughout the study includes 39 participants. Most of the participants were graduate students working in different labs. However, some participants were also selected from office and technical staff and undergraduate students. The study ran for 5 weeks starting from April 1st, 2011 and ending May 6th, 2011.

HTC Magic phones were provided to all participants with a package that contained a cell phone, USB cable, charger and headphones. Some of the participants used their personal SIM card and used the study phone as their primary phone. However, most of the participants were given ROGERS SIM cards with pay-as-you-go plans. The phones were loaded with $20 before distribution. Rogers pay as you go plans provided either unlimited data for $2 for a day or $7 for a week. The procedure of activating the data plan was demonstrated to the participants before the phones were distributed. The participants were requested to contact study organizers if additional funds were required. However, only 3 participants contacted the organizers for this purpose.

3.2 Hardware Platform

HTC Magic phones were used for conducting the user study. Two different hardware platforms exist for HTC Magic phones. The PVT32B platform which comes with a Qualcomm MSM7201A ARM11 processor and 192MB RAM was used for the study. The Qualcomm MSM7201A processor is an ARM-based, dual-core CPU/GPU and contains many built-in features. It supports 3G and provides a GPU capable of up to 4 million triangles per second. It supports hardware acceleration for Java. However, it does not accelerate execution of Android applications, as they are targeted to the Dalvik virtual machine, not the Java Virtual Machine.

HTC Magic phones have a micro SD card slot. An 8 GB memory card was installed in each phone before the study. When the phone is connected to a computer through a USB cable, it is possible to access the card without removing it from the phones. Media files are accessible if arranged in folders; however, folders can only be created from a third-party file management application or from a computer. HTC Magic phones have a 3.2 inch LCD flat glass touch-sensitive screen with a 480 by 320 pixel resolution. Users can interact with the phone through touch, tapping or touch-drag motions. The touch screen is capable of multi-touch gestures. HTC Magic phones do not have a physical keyboard. Text input is provided through an on-screen keyboard. The phones have a built-in accelerometer that rotates the keyboard between portrait and landscape view automatically based on the orientation of the phone.

HTC Magic phones come with a 3.2-megapixel camera that has autofocus functionality. It supports video recording and recorded videos can be directly uploaded to YouTube. Videos are recorded in 3GP format. HTC Magic phones are capable of playing H.264, 3GPP, MPEG4, and 3GP files. Table 3.1 summarizes the properties of HTC Magic phones[1].

HTC Magic phones use 3.7V, 1340mAh rechargeable lithium ion batteries that can be charged through a plug-in charger and USB cable. The phones have a built-in GPS receiver for location functionality including free turn-by-turn GPS (Google Maps Navigation). A phone can also use cell towers and Wi-Fi hotspots to help determine its location. It also has a built-in digital compass to help determine its orientation. HTC Magic supports Wi-Fi, 3G, Edge, and GPRS connectivity. The phones automatically uses Wi-Fi whenever it is available. If there is no Wi-Fi it uses the

[1]http://www.htc.com/ca/products/magic-rogers# tech-specs, access date=13/10/2011

Table 3.1: Specification of HTC Magic Phone(PVT 32B)[1]

Property	**Description**
Manufacturer	HTC Corporation
Dimensions	4.4 in H, 2.187 in W, 0.537 in D
Weight	116 gm
CPU	PVT32B 528 MHz Qualcomm MSM7201A ARM11
Memory	512 MB ROM
Storage	192 MB RAM
Removable storage	Micro SD (supports up to 16 GB)
Battery	1340 mAh, 3.7 V
Display	320 x 480 pixel, 3.2 in
Camera	3.2 Mega pixel

cellular network through 3G or Edge.

3.3 Software Platform

Android 2.1 was used as the phone operating system for the user study. Although Android 2.1 provided many required facilities, the operating system had to be modified to support the applications used for data collection. In particular, Bluetooth was required to be discoverable for an unlimited period of time for contact tracing. Therefore, minor modifications were made to the OS to enable indefinite discoverability. Google apps, compatible with Android 2.1, were also installed. Google apps provided applications such as gmail, google maps and Android Market. Different applications were installed on the phones for recording user behavior which include Healthlogger, TryTcp and Logactivity. Details of these applications are provided in later sections.

3.3.1 Healthlogger

Healthlogger is one of the applications installed in the phones for the 'iEPi' project which accompanied data collection for this thesis work and this thesis made use of the data collected by the Healthlogger application. An initial description of the Healthlogger application and its epidemiological utility can be found in Hashemian *et al.* [21]. However, the development and instrumentation of this application is not a part of this thesis research. A brief discussion about the 'Healthlogger' application is provided here to clarify its functionality and establish the source of much of the data described in Chapter 5.

Healthlogger collected the GPS location of the participants along with available Bluetooth and Wi-Fi node information. Accelerometer data was also collected to identify participants' movement.

Table 3.2: Data Collected by Healthlogger Application

Item	Information Collected	Duration
GPS	Id, Timestamp, Latitude, Longitude, Speed, Accuracy	2 min
Bluetooth	Id, Timestamp, Mac, Device Name, Device class , RSSI	1 min
Wi-Fi	Id, Timestamp, BSSID, Capabilities, SSID, Frequency, Level	5 s
Accelerometer	Id, Timestamp, Acceleration_ X, Acceleration_ Y, Acceleration_ Z	1 min

Table 3.3 summarizes the information collected through the healtlogger application. The collected information was uploaded to a SQL server whenever secured university Wi-Fi was available.

Healthlogger woke up the phone every 5 minutes to capture information. Each 5 minutes period is considered a duty cycle. When the phone wakes up, it tries to contact the GPS satellites to provide GPS location information. It waits 2 minutes to collect data. GPS locations can be easily obtained if the participant is outside. However, when the participant is inside a building or there exist some obstacles, GPS information is sometimes unavailable. GPS information contained the longitude and latitude of the participant along with the speed the participant is moving and an accuracy estimate of the data.

The Healthlogger application also collected information about discoverable Bluetooth devices. It tries to identify the devices in the immediate neighborhood that have Bluetooth in discoverable mood. The information contains timestamp, MAC address of the device identified, device class and the RSSI level. The name of the device is not collected according to ethical guidelines. Using the device class information it is possible to categorize devices into computer, phone, LAN/network access point or peripheral device and others. Minor device class information helps to further categorize the computer into subcategories such as desktop work station, server-class computer, laptop or hand-held PDA.

Healthlogger collected information about available Wi-Fi access points. Whenever the phone wakes up in each duty cycle period it tries to identify any available Wi-Fi access points around the participant. The participant's Id, timestamp information, BSSID and SSID of the available access point, level and frequency of the received signal and the security protocol of the access point was collected. SSID and BSSID of the access point helped to identify the location of the participant while indoors.

Accelerometer data was collected during the user study by the Healthlogger application. This information helped to identify participant movement patterns. Information about participants' movement along the x, y and z axis was collected to determine participants' activity level.

GPS and Wi-Fi data is used to obtain location information and Bluetooth data is used to determine participants' proximity to other people in this thesis work.

3.3.2 TryTcp

The TryTcp application was developed as a part of this thesis work. It is installed in the phones to record network activities of the participants. The Android operating system provides a shell to execute command line instructions. Tcpdump, a powerful command-line packet analyzer which contains a portable C/C++ library for network traffic capture, was used for capturing network packets[2]. Tcpdump helps to capture network traffic and also manipulate the properties of capture. The TryTcp application executes tcpdump commands using the shell interface of Android OS. The TryTcp application provides options for disabling recording of Internet usage for a specific period of time. This option was provided because of the privacy concerns of the participants. Figure 3.1 shows different components of the TryTcp application.

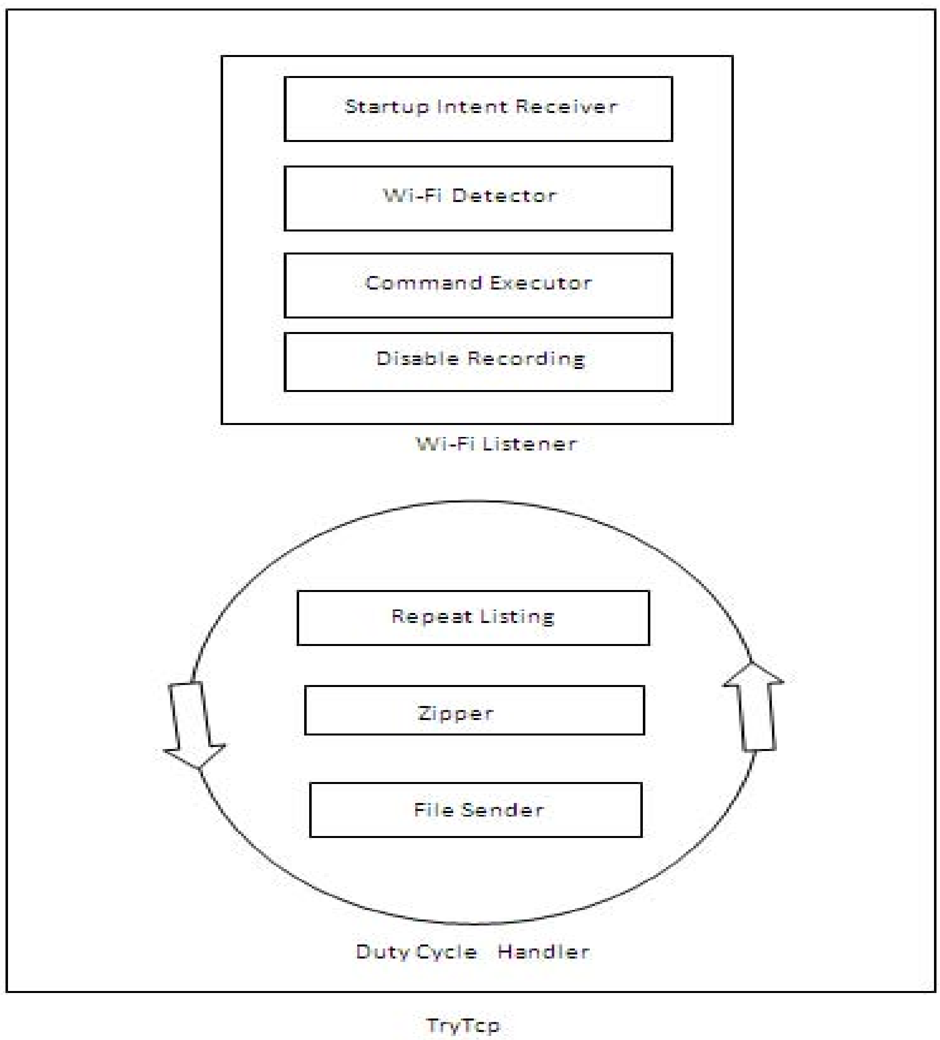


Figure 3.1: Components of TryTcp

[2]http://www.tcpdump.org/, access date=13/10/2011

Wi-Fi Listener

The Wi-Fi Listener handles activities related to changes in the Wi-Fi connection. Whenever there is a change in the Wi-Fi connection or some event of interest, the Wi-Fi Listener automatically detects it and takes appropriate action. Several functions are employed in the Wi-Fi Listener. They are the Startup Intent Receiver, the Wi-Fi Detector, the Command Executor and Disable Recording.

Startup Intent Receiver: The core components of Android applications such as activities, services and broadcast receivers are activated through messages known as Intents. An Intent object is a passive data structure that holds an abstract description of an operation to be performed, and particularly in the case of broadcasts, a description of something that is being announced. The Startup Intent Receiver is created in such a way that whenever the phone starts or restarts an Intent is created by the OS and sent to this receiver. After receiving such an intent this function checks whether the recording option was disabled earlier. If the recording option was not disabled, it sends a signal to Wi-Fi Detector to detect Wi-Fi connectivity.

Wi-Fi Detector: The Wi-Fi Detector monitors Wi-Fi connections to check whether there is a change in Wi-Fi connectivity. It receives a signal from the Startup Intent Receiver when the phone starts. It also receives a signal from Disable Recording function when recording can resume. The main purpose of this detector is to monitor whether the phone is using a Wi-Fi network or the cellular network. Whenever an Internet connection changes from cellular to Wi-Fi network or vice-versa, it sends a signal to the Command Executor function to indicate the change and redirects it to listen to the interface which is being used for network traffic.

Command Executor: The Command Executor executes tcpdump commands with appropriate options to capture network traffic. It receives signals from the Wi-Fi detector. The signals contain information about the interface that is being used for network traffic. Based on that information, the Command Executor decides whether to start monitoring cellular or Wi-Fi traffic. Some of the properties are set while executing the tcpdump command. The whole packet is captured during data collection including headers and payloads. Captured packets are saved in an SD card directory using a file name with timestamp and interface information. Later, files are compressed using the Zipper function to save space and uploaded to the server. The Command Executor also receives signals to disable recording for specific periods of time. In that case, it kills the tcpdump process.

Disable Recording: TryTcp GUI provides opportunity to disable recording for 30, 60 or 120 minutes. If recording is disabled the duration is saved in a file and a signal is sent to the Wi-Fi Detector component instructing it to disable detection. A signal is also sent to the Command Executor asking it to kill all running tcpdump processes. Whenever this period expired, a signal is sent to the Startup Intent Receiver asking it to start recording again. A signal is also sent to the

Wi-Fi Listener asking it to monitor the Wi-Fi connection again.

Duty Cycle Handler

The Duty Cycle Handler is one of the major components of the TryTcp application. As mentioned previously every 5 minute period is considered a duty cycle. Synchronization was performed between Healthlogger and TryTcp requiring both the applications to start their duty cycles at the same time to extend battery life. Several functions are executed by the Duty Cycle Handler during each duty cycle.

Repeat Listing: Repeat Listing is executed at the beginning of each duty cycle. This function performs several tasks. It records battery level and memory status in different files. It records the amount of free space in both internal memory and on the SD card. This function also identifies a list of processes running at that particular point of time which is saved using timestamp information in a file. All the files are saved on the SD card before being uploaded to the server.

Zipper: The Duty Cycle Handler executes the Zipper function to compress captured network traffic files into zip files. Compression techniques are used to optimize memory utilization. For compression the Java zip package was used providing up to 75% of compression. For most files, the compression technique provided a satisfactory performance of up to 30% of compression.

File Sender: The purpose of this function was to upload recorded files to the server. The File Sender function first communicated with the Wi-Fi Detector function to check whether Wi-Fi was available and connected. If Wi-Fi was connected it checked the SSID of the Wi-Fi connection. Files were only uploaded when participants were connected to 'uofs-secure', the secured Wi-Fi connection available at the University of Saskatchewan. Files were transferred over a secured connection to ensure the privacy of participants as required by the ethics approval. Files were transferred using a TCP socket connection. The server was set up in the DISCUS lab of the Computer Science Department and the server continuously listened for connections from the client side. Only after confirmed successful completion of file transfer, were files deleted from participants' SD card memory.

3.3.3 LogActivity

The LogActivity application was developed and installed for debugging purposes. This application stores the output of the logcat command. Logcat is used to obtain messages from different applications in Android. Log messages are generated with specific tags, usually the application names allowing identification of which application generated the message. Applications responsible for data collection during the user study generated log messages periodically along with the execution of different parts of the programs. The resulting file helped to identify the existence of problems in the software. The output of the logcat commands were saved in files using timestamp information.

These files were also uploaded to the server.

3.4 Limitations of Data Set

Appropriate measures were taken to ensure that the phones kept working properly for the period of the study. However, 2 phones had their SD cards replaced. The amount of data lost was not significant as the files were always uploaded as soon as the secured Wi-Fi connection was available.

All of the study participants were from the U. of S. Computer Science Department. Therefore, some question might arise about the generalizability of the dataset. However, participants were selected from different age groups, ethnicity and gender. Even with this biased sample heterogeneity in usage patterns were observed (see Chapters 4&5). Most of the participants used these phones along with their personal phones. Therefore, the cell phone usage behavior might not be representative. However, participants were requested to always use these phones for Internet access during the study.

The amount of data collected during the user study was not large. Most likely this is due to participants coming from the Computer Science Department, and having Internet access through personal desktop/laptop computers most of the time. A survey was conducted at the end of the study and participants were asked whether they prefer to use a smartphone for Internet access instead of a laptop/desktop and all of the participants said that they preferred computers. However, the dataset still represents usage of smartphones for Internet access.

As an undetected side effect of using the custom OS, 3G capability on the phones was disabled, and only Edge network cellular connectivity was available to the participants. This was discovered after study completion and is likely to have had an impact on the amount of cellular traffic generated by participants.

Some of the users complained about being interrupted while watching videos. They reported that youtube application exited unexpectedly before they could complete watching the videos. The amount of video traffic obtained during the user study was also low, probably due to the interruption reported by the study participants. Therefore, the amount of video traffic found in the collected dataset does not match the percentage of video traffic observed in previous studies [15, 42, 31].

3.5 Analysis Tools

Several tools have been used for analysis purposes. The server side was running a java program. Files from phones were collected over TCP connections. Files were saved in different directories using the participant's Id. Files obtained from cellular and Wi-Fi traffic were kept in separate directories and analyzed separately. Files were first unzipped using a shell script and then merged using tcpslice. Tcpslice[3]merges tcpdump files using timestamp information from the packets.

Merged files were run through a bro [35] script provided by one of the authors of [16]. Bro is an intrusion detection system which can be used to analyze tcpdump files. The bro script converted the files into human readable text files containing information required for data analysis.

Later, python and C++ were used for parsing the text files and generating necessary information such as traffic transferred at different times of the day and different days of the week, popular websites and individual usage patterns. Data generated from the Healthlogger application was saved in a SQL database. Text files obtained from the bro script were also uploaded in separate tables of an SQL database. Several SQL queries were used to obtain different information such as individual usage patterns, popular websites and content type. Graphs were plotted using Microsoft Excel and Matlab.

Some of the results presented in Section 5.2 show location information of the participants. However, participants were personally contacted and provided with the graphs in this thesis that show location information. After receiving permission from the participants, results are presented in this thesis following the protocol approved by the ethics committee.

[3]http://linux.about.com/library/cmd/blcmdl8 tcpslice.htm, access date=13/10/2011

Chapter 4

Workload Characteristics

The collected smartphone traffic is analyzed in this chapter. Initially basic traffic characteristics, such as the amount of traffic during different hours of the day and days of the week, are presented to illustrate the general characteristics of the collected smartphone traffic. Popularity analysis is provided describing which websites were the most popular among study participants. Application usage is also discussed in this chapter.

4.1 Traffic Characteristics

This section focuses on basic characteristics of the collected traffic. Table 4.1 summarizes some important aspects of the user study and expresses differences between the Wi-Fi and the cellular network use. Of the 39 participants 30 used both Wi-Fi and cellular networks and 9 used Wi-Fi networks exclusively to connect to the Internet. 42,421 transactions were performed over Wi-Fi networks and 33,035 transactions were performed over the cellular network, where the communication between client and server for a single request is considered a transaction. The total amount of data transferred over Wi-Fi networks (526.87 MB) was much higher than the amount of data transferred over the cellular network (90.25 MB). The average transfer size, maximum transfer size and average connection duration was also higher for Wi-Fi networks. A total of 128 hosts were accessed over Wi-Fi networks in contrast to 82 hosts accessed over the cellular network. All of this data suggests that participants consumed more data and performed more download-intensive activities over a Wi-Fi network than a cellular network, which is not surprising because the Wi-Fi network is a) automatically used by the phones for transferring traffic when available, b) usually faster, and c) often free. These findings are broadly consistent with trends observed by Cui *et al.* [9] which suggested presence of more traffic over Wi-Fi networks than the celluar network.

Examining traffic volumes at different hours of the day and for different days of the week illustrates basic usage patterns. Figure 4.1 shows the average amount of traffic for all participants transferred during different hours of the day. From the graph it is evident that there was almost always more Wi-Fi traffic than cellular traffic. Figure 4.1 illustrates how participants used Wi-Fi networks during different hours of the day. Most of the Wi-Fi traffic was transferred from 4 pm to

Table 4.1: Highlights of the Study

Criteria	Wi-Fi Network	Cellular Network
Number of Participants	39	30
Number of Transactions	42,421	33,035
Data Transferred	526.87 MB	90.25 MB
Avg. Transfer Size	5.31 KB	2.36 KB
Avg. Connection Duration	86.54 S	23.97 S
Hosts Accessed	128	82

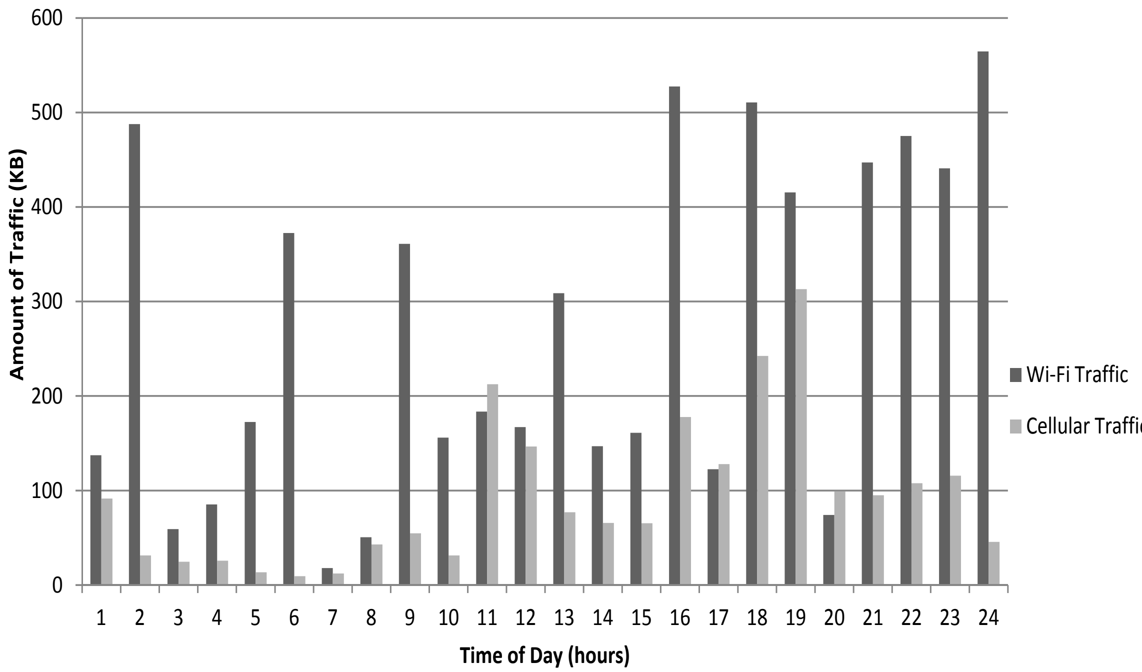


Figure 4.1: Traffic at Different Hours of the Day

12 am which is consistent with the results presented by Harvey *et al.* [19]. However, some spikes were evident from 1 am to 4 pm. Heaviest usage is found during evenings. The amount of Wi-Fi traffic decreases at 5 pm and 8 pm. A similar decrease around 8 pm has also been reported [19]. Spikes for Wi-Fi traffic at 2 am, 6 am, 9 am and 1 pm exist, but the spike at 2 am was primarily due to a single participant who browsed several websites during that time on one day of the study. Traffic at 6 am and 9 am was spread across participants and days.

A more stable pattern is found for cellular traffic throughout the day. Traffic is lower at night and picks up around 11 am in the morning. The highest peaks are seen at 7 pm. A closer look at traffic during that time revealed that participants used facebook, Android Market and maps.

Traffic volumes on different days of the week is presented in Figure 4.2. The graph clearly shows that Wi-Fi traffic is higher than cellular traffic for both weekdays and weekends. Wi-Fi

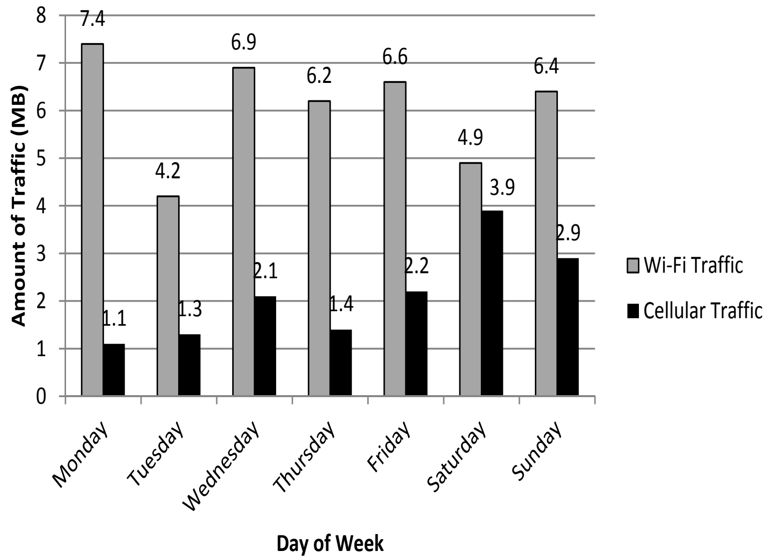


Figure 4.2: Traffic at Different Days of the Week

traffic remains more or less stable for Monday, Wednesday-Friday and on Sundays. However, Wi-Fi traffic is much lower on Tuesdays and Saturdays. The reason for having less traffic on Saturday might be the effect of weekends. Sunday is also a holiday, but a substantial amount of traffic was generated by one participant who used the phone heavily on Sundays. Cellular traffic remains stable on weekdays; however, it increases during weekends because of increased traffic from the participants who did not have access to Wi-Fi networks.

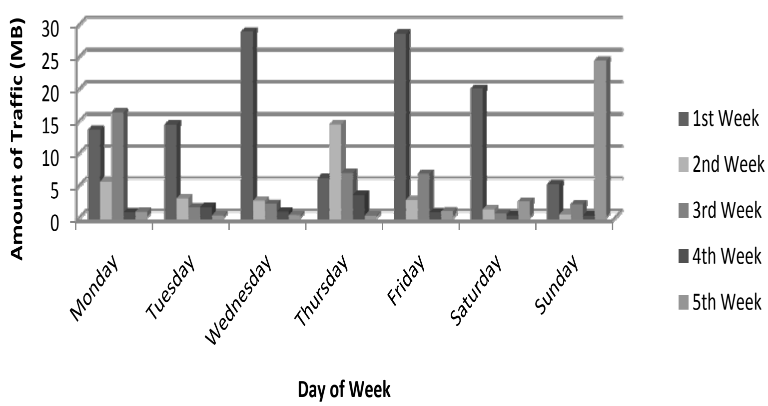


Figure 4.3: Wi-Fi Traffic in Different Weeks of the Study

Figure 4.3 represents Wi-Fi traffic over different weeks of the study. The figure shows that the first week's traffic is higher than traffic for the rest of the weeks. After getting the phones, most of the participants explored different applications and browsed different websites. After the initial exploration period only some of them used network functionality regularly which is consistent with the study performed by Shepard *et al.* [43] where participants also accessed more applications during the first week of the study. Similar results are found for cellular traffic in Figure 4.4.

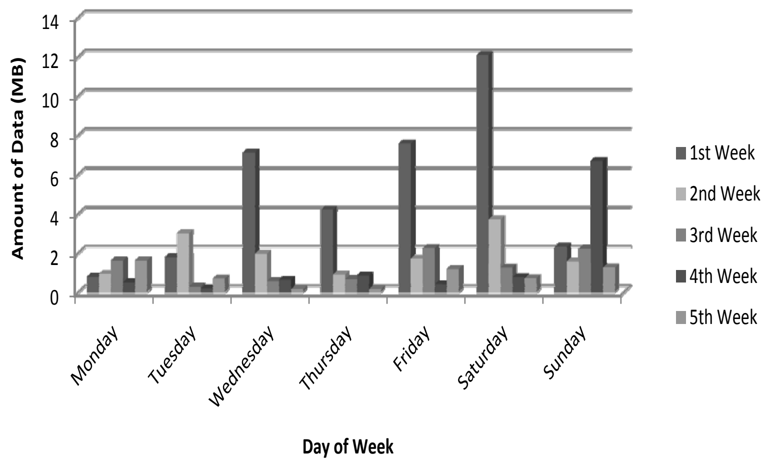


Figure 4.4: Cellular Traffic in Different Weeks of the Study

Request generation by different participants is presented in Figure 4.5. The x-axis of the figure represents a cumulative percentage of participants, when ranked in decreasing order of the number of requests generated and the y-axis represents the cumulative percentage of requests generated by participants. Figure 4.5 suggests that there exist some participants who used the network functionality more than others. The top 10% of the participants generated 61% of the Wi-Fi requests and the top 30% of the participants generated around 82% of the Wi-Fi requests. Skewed traffic distributions with respect to individual users were also observed previously [42]. Request generation is less skewed for cellular traffic. The top 10% of the participants were responsible for generating 34% of the cellular requests and the top 30% of the participants generated 63% of the cellular requests. The primary reason for the dissimilarity was the usage patterns of the top 3 users of Wi-Fi, who generated 60% of the requests. In contrast, the top 3 users of the cellular network accounted for only 35% of the requests.

Figure 4.6 shows the distribution of data transfer with respect to individual participants. The x-axis represents the cumulative percentage of participants, when ranked in decreasing order of data transfer volume, and the y-axis represents the cumulative percentage of the total data transfer volume. For this figure the number of bytes transferred is considered instead of the number of requests as in Figure 4.5. The top user alone contributed 37% of the bytes transferred for all participants over Wi-Fi networks. The top 15% of the participants (6 participants) consumed 70% of the bytes and the top 40% of the participants consumed 90% of the bytes over Wi-Fi networks. The top user for cellular traffic contributed 24% of the bytes while the top 15% of the participants consumed 62% of the bytes and the top 40% of the participants consumed 87% of the bytes. This

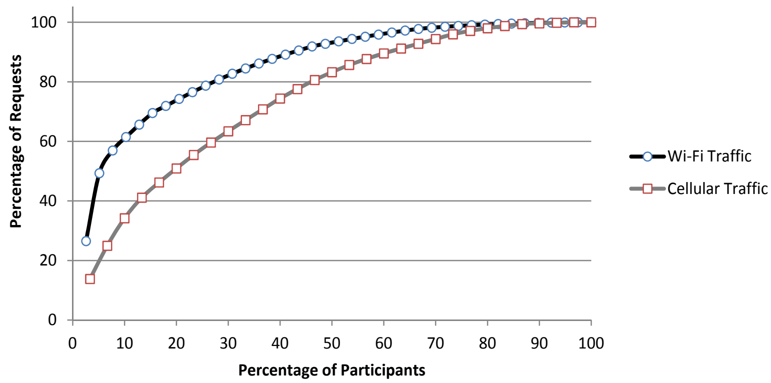


Figure 4.5: Request Generation by Participants

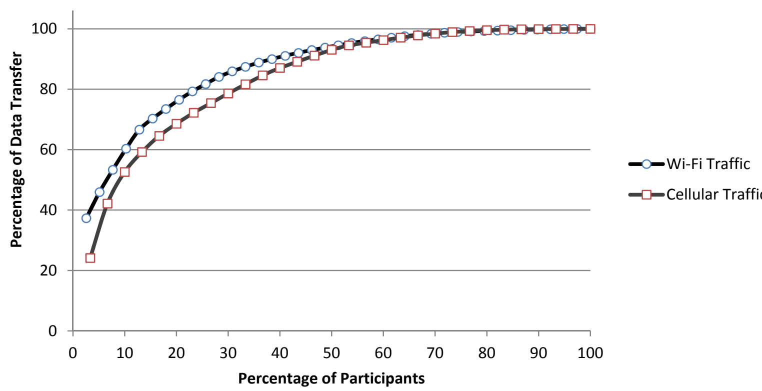


Figure 4.6: Data Transfer by Participants

pattern, characterized by a relatively small number of heavy users, was also described by Shafiq *et al.* [42] who reported that the top 5% of the users generated 90% of the traffic during their smartphone traffic collection from a backbone network.

The amount of traffic generated by individual participants over different networks is presented in Figure 4.7. From the figure it is evident that participants generated more traffic over Wi-Fi networks. The top user of the Wi-Fi networks consumed 80 MB data while the rest of the participants consumed between 137 KB and 18 MB of data. For Wi-Fi traffic the average data consumption is 5.8 MB data per participant with a median of 1.7 MB. The top users generated most of the Wi-Fi traffic leading to a higher average than median. The top users of the cellular network transferred 18 MB data over the network while the rest of the participants transferred data ranging from 10 KB to 13 MB. For cellular traffic the average was 2.5 MB per participant with a median of 1.2 MB.

The ratio of Wi-Fi to collected traffic for all participants is presented in Figure 4.8. Here participants are ranked in increasing order of this ratio, and the figure shows the percentage of participants whose ratio is at most the value given by the x-axis value. The median ratio of Wi-Fi traffic is 0.78 which indicates participants mostly used Wi-Fi networks. Twenty percent of the

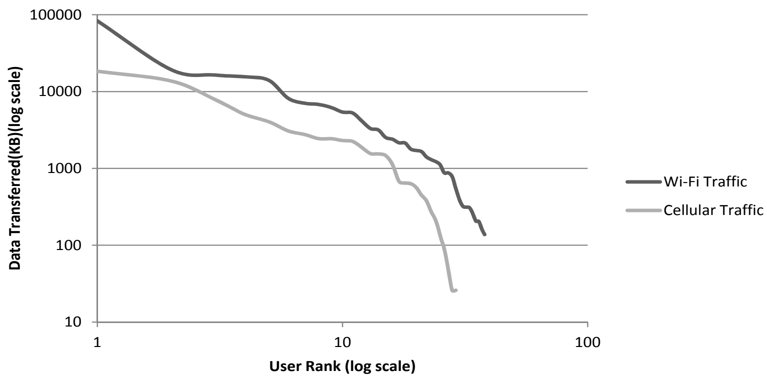


Figure 4.7: Traffic Generated by Individual Participants

participants only used Wi-Fi for connecting to the Internet. However, 25% of the participants used Wi-Fi for transferring at most 25% of their traffic. Similar results are found in the literature. Falaki *et al.* [13] reported that 20% of the participants of their study did not use Wi-Fi at all and 20% of the participants used Wi-Fi to transfer 80% of their traffic. Such results indicate that smartphone user studies based on solely Wi-Fi or cellular traffic will miss a significant portion of the traffic or miss the origin of the traffic. An on-device logger like the one used for this thesis research and discussed in Section 3.3.2 can provide more complete results.

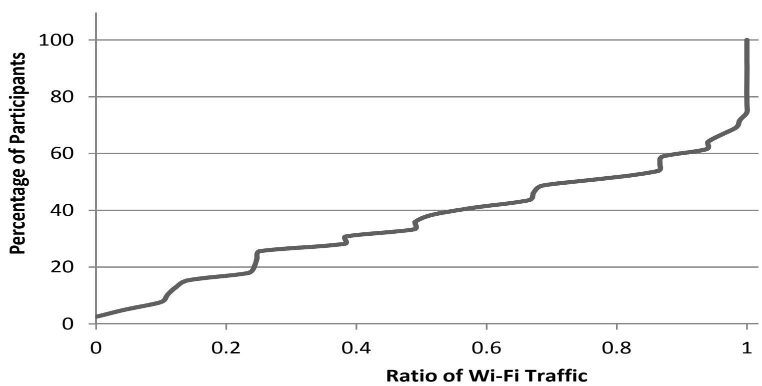


Figure 4.8: Ratio of Wi-Fi to Collected Traffic

Figure 4.9 represents the transfer size over different networks. In general, transfer sizes in a Wi-Fi network were larger than over the cellular network. There exists a straight vertical portion in both the Wi-Fi and the cellular network curves. Around 60% of the transfers over the cellular network were of almost same size (around 138 bytes). These transfers were generated from facebook, google and for providing location-related services. Participants frequently received data from location-related services and accessed google and facebook more over the cellular network while generating such transfers. The maximum transfer size over the cellular network was 3.1 MB. The average transfer size over the cellular network was 2.4 KB. Around 40% of the transfers over a Wi-Fi

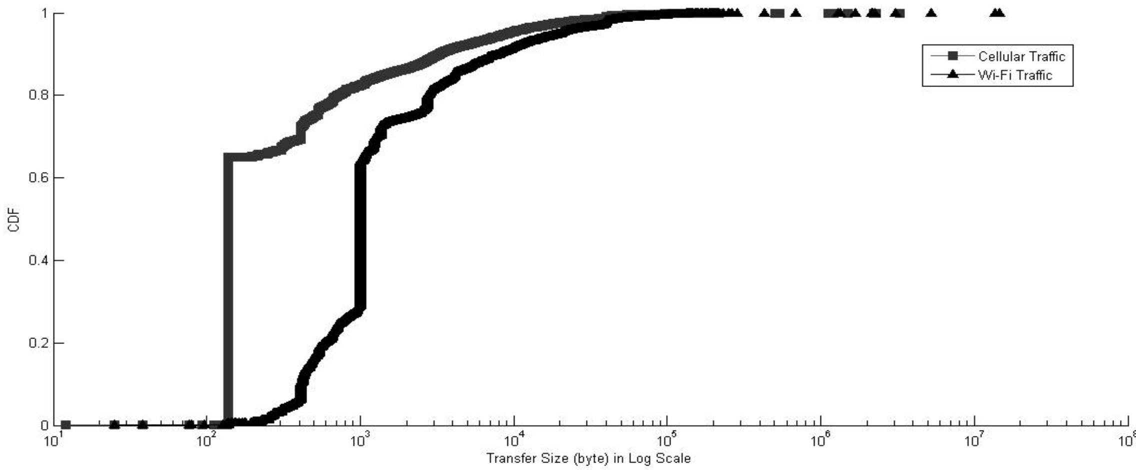


Figure 4.9: CDF of Transfer Size

network were of same size (around 1 KB). Most of these transfers were generated by the yahoo messenger application. One of the participants used yahoo messenger and was logged in for a long period of time. Similarly-sized data was transferred to keep the session alive or to send updates to the application. The maximum single transfer size over a Wi-Fi network was 14.02 MB. The average transfer size over Wi-Fi networks was 5.31 KB. One interesting thing to notice from the figure is that around 22% of the transfers over Wi-Fi networks and around 80% of the transfers over the cellular network were less than 1 KB. Such small transfer sizes were mentioned previously [13] where the authors found that 30% of the transfers were less than 1 KB.

Table 4.2: Breakdown of Traffic by Ports

Protocol/port	Wi-Fi Traffic		Cellular Traffic	
	Request(%)	Bytes(%)	Request(%)	Bytes(%)
HTTP(80)	69.54	89.18	90.50	87.30
Android-Mkt(5228)	25.05	1.68	5.96	0.80
HTTPS(443)	5.18	8.80	3.23	10.00
SOCKS(1080)	0.03	0.19	0.03	1.21
IMAP4S(993)	0.02	0.11	0.06	0.12
Other	0.18	0.04	0.22	0.57

The traffic generated from both Wi-Fi and cellular networks is categorized by ports in Table 4.2. Most of the traffic in both networks was HTTP traffic transferred over port 80. HTTP contributed 69.54% of the requests over Wi-Fi networks. The rest of the traffic was generated by the Android Market (25.05%), HTTPS (5.18%), SOCKS (0.03%), IMAP4S (0.02%) and Other (0.18%) protocols. However, if the percentage of bytes transferred for different protocols is considered, HTTP dominates even more. 89.18% of the bytes transferred were due to HTTP traffic. The breakdown of traffic for the cellular network shows a similar scenario. HTTP contributed 90.5% of the requests and 87.3% of the bytes. The rest of the traffic over the cellular network was contributed by Android Market (5.96% of the requests and 0.8% of the bytes), HTTPS (3.23% of the requests and 10% of the bytes), SOCKS (0.03% of the requests and 1.21% of the bytes), IMAP4S (0.06% of the requests and 0.12% of the bytes) and Other (0.22% of the requests and 0.57% of the bytes). The Android Market generated more requests and bytes over Wi-Fi networks than over the cellular network, which suggests participants primarily downloaded and installed applications over Wi-Fi networks.

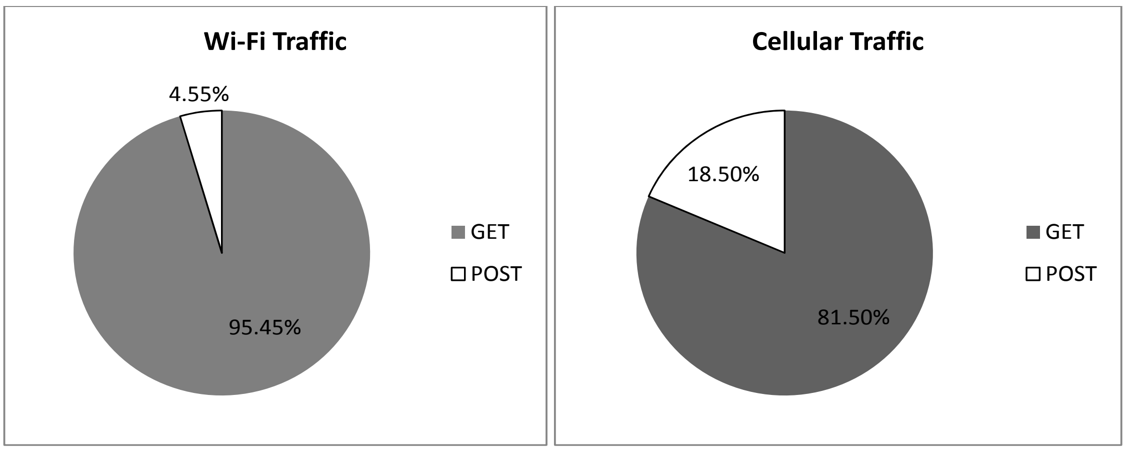


Figure 4.10: HTTP Methods for Wi-Fi and Cellular Traffic

Figure 4.10 represents the breakdown of Wi-Fi and cellular traffic by HTTP method. There are two primary HTTP methods used for data transfer, namely GET and POST. The GET method is usually used to send a request and to obtain data from a server. The POST method is generally used to upload data to a server. Figure 4.10 shows that 95.45% of the bytes were transferred by the GET method over Wi-Fi networks and only 4.55% of the bytes were transferred by the POST method. However, the breakdown of cellular traffic shows that 18.5% of the bytes were transferred by the POST method and 81.5% of the bytes were transferred by the GET method. More per capita POSTs occurred in cellular than in Wi-Fi traffic. The primary reason was the use of map applications over the cellular network which used the POST method to send location information to the server.

HTTP traffic is further categorized by content type in Figure 4.11. For all the pie charts presented in this thesis, the legend entries are arranged in decreasing percentage order. Different categories are shown in a clock-wise fashion starting from the first category mentioned in the legend. For example, this Figure 4.11 clearly shows the dominance of the application content type in both networks. Application content contributed 44.65% of the traffic over W-Fi networks and 49.77% of

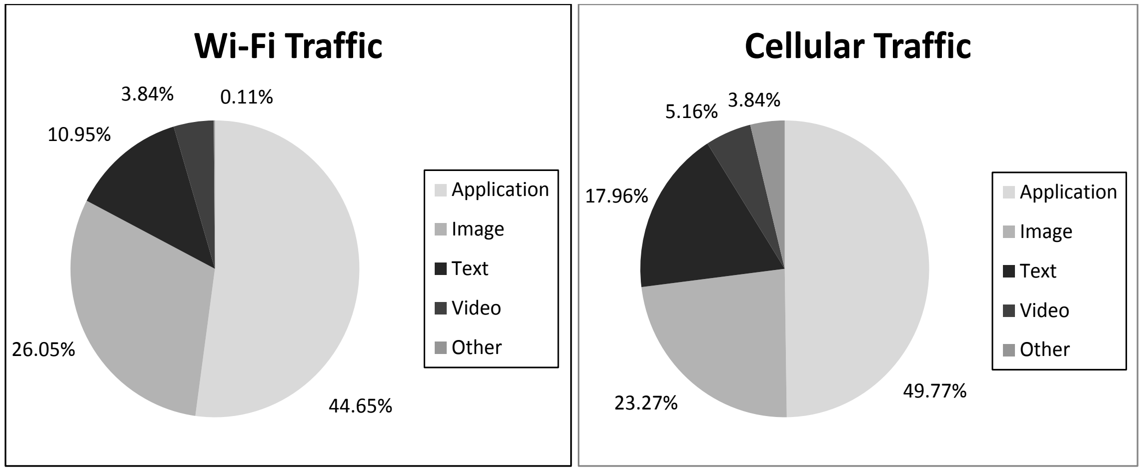


Figure 4.11: HTTP Content Type for Wi-Fi and Cellular Traffic

the traffic over the cellular network. Image content accounted for 26.05% of the traffic over Wi-Fi networks and 23.27% of the traffic over the cellular network. Text content accounted for 10.95% and 17.96% of the traffic over Wi-Fi networks and the cellular network respectively. Video type content generated only 3.84% of the traffic over Wi-Fi networks and 5.16% of the traffic over the cellular network. Some of the participants complained there were problems with the installed video viewer, leading to a lower percentage of video content than expected. Maier *et al.* [31] reported that video and audio type content contributed 35-52% of the traffic in their dataset.

Table 4.3 gives a detailed breakdown of Wi-Fi content types. Only 0.07% of requests were for Android Market content, but this small percentage of requests resulted in 18.3% of the bytes transferred. A small number of requests resulted in large download sizes given the relative size of typical apps. The presence of CSS content was also significant. It accounted for 7.79% of the bytes transferred. Only 0.13% of the requests for video content generated 3.84% of the bytes transferred. The reason behind the lower than expected percentage of video requests was discussed earlier.

A detailed breakdown of cellular traffic by content type is presented in Table 4.4. Binary application data was transferred for 23.64% of the requests resulting in 20.42% of the bytes transferred. Participants often used map applications over the cellular network, leading to a high percentage of binary application transfers because the map application transferred data using the binary content type. For the cellular network, a high percentage of data was transfered from the Android Market for a small percentage of requests similar to the Wi-Fi case. Only 0.14% of the requests generated 11.09% of the bytes transferred. Video content also generated a significant percentage of bytes transferred over the cellular network, where only 0.21% of the requests generated 5.16% of the bytes transferred.

From the preceding analysis no major differences between data type usage patterns over Wi-Fi and cellular networks is observed. However, intuitively differences between the consumption of

Table 4.3: Breakdown of Wi-Fi Traffic by Content Type

Content Type	Content Subtype	Request (%)	Bytes (%)
Application	ATOM+XML	3.44	0.74
	BINARY	14.37	9.99
	JAVASCRIPT	0.23	0.59
	JSON	8.33	2.65
	OCTET-STREAM	1.64	6.84
	VND.ANDROID.PACKAGE-ARCHIVE	0.07	18.3
	X-JAVASCRIPT	3.16	5.47
	X-SHOCKWAVE-FLASH	0.01	0.06
	HTML+XML	0.03	0.01
	XML	0.02	0.01
Image	GIF	11.55	4.41
	JPEG	11.97	18.88
	JPG	0.03	0.01
	PNG	3.29	2.02
	VND.MICROSOFT.ICON	0.01	0.01
	X-ICON	0.18	0.03
Text	CSS	2.07	7.79
	HTML	3.46	16.93
	JAVASCRIPT	31.71	1.28
	PLAIN	2.11	0.05
	X-JS	0.45	0.1
Video	3GPP	0.01	1.95
	MP4	0.12	1.89

video type content over different network might have been observed if participants could enjoy watching videos without interruption. Previous work showed no comparisons between content type usage patterns over Wi-Fi and cellular networks.

4.2 Application Use

The previous section presented the basic characteristics of network traffic generated from smartphones. This section addresses application usage to determine which applications were used by the participants which will help to identify the way people use smartphones in different circumstances in their everyday life.

Table 4.4: Breakdown of Cellular Traffic by Content Type

Content Type	Content Subtype	Request (%)	Bytes (%)
Application	ATOM+XML	5.23	1.04
	BINARY	23.64	20.42
	JAVASCRIPT	0.48	0.98
	JSON	10.7	4.33
	OCTET-STREAM	1.59	7.12
	VND.ANDROID.PACKAGE-ARCHIVE	0.14	11.09
	X-JAVASCRIPT	3.38	4.44
	XML	1.37	0.34
Image	BMP	0.02	0.05
	GIF	10.95	3.16
	JPEG	14.15	15.44
	JPG	0.02	0.01
	PNG	4.87	4.41
	ICON	0.53	0.1
Text	CSS	2.35	1.79
	HTML	13.26	11.16
	JAVASCRIPT	5.88	4.67
	PLAIN	0.96	0.34
	XML	0.03	0.01
Video	3GPP	0.21	5.16

4.2.1 Popular Applications

Figure 4.12 presents the relative popularity of applications across all participants over the entire period of the study. Lists of running processes were collected at the beginning of every 5 minute interval during the user study. There might be some applications which were used after the lists were collected and did not appear in the respective lists. Lists of applications associated with those processes are identified later for analysis. However, many system processes were found during data collection, and most of the system processes are not initiated intentionally by participants. Therefore, those processes were ignored while calculating the relative popularity of applications. In addition, processes associated with the experimental applications described in Chapter 3 are not included as they were always running during data collection. Here, applications are organized according to the number of times they have been used by the participants. The figure shows that browser and map applications are the two most popular applications among participants with

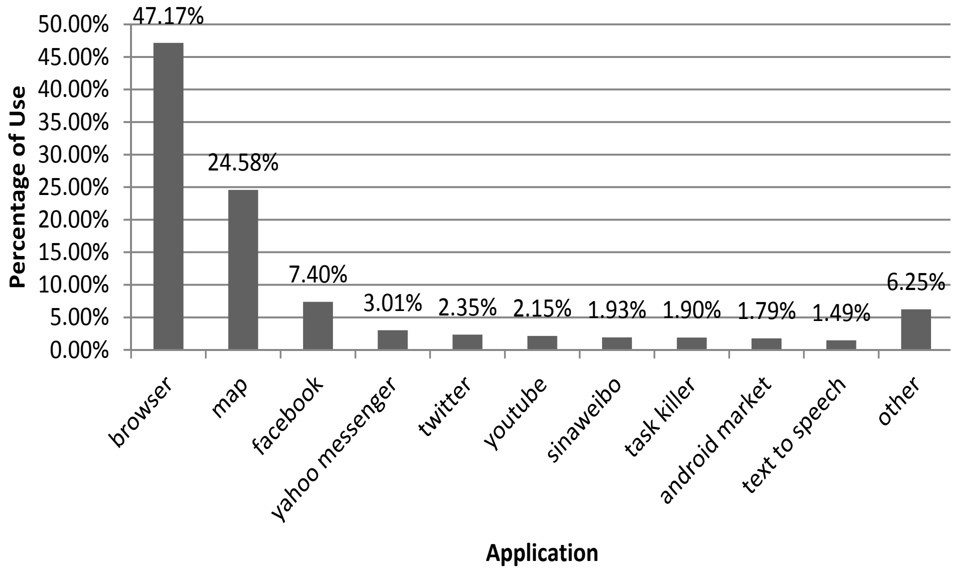


Figure 4.12: Relative Popularity of Applications

47.17% and 24.58% of use followed by facebook with 7.40% of use.

The relative popularity of different categories of application is presented in Figure 4.13. Here, productivity application refers to applications such as calculator, alarm clock or compass and communication application refers to voice calls, SMS and instant messaging. The rest of the categories are easily understandable from their names. The browsing and map categories were the most popular among participants; however, a significant percentage of productivity and communication applications were also identified. Productivity applications accounted for 6.94% of usage and communication applications accounted for 3.13% of usage. Falaki *et al.* [14] reported that the relative popularity of communication applications, browsing applications, productivity applications, media and maps are 44%, 10%, 19%, 5% and 5% respectively. The collected data suggests more browser, map, and social networking application use and less usage of communication, games and productivity applications. Most of the participants did not use study phones for voice calling and sending text messages because they kept their own phones as primary communication devices and used the supplied smartphones primarily for web functionality. Therefore, the study results showed less use of communication and productivity applications.

The above discussion provides some idea about the overall application usage in the collected dataset. Applications that generated network traffic are of particular interest because understanding the way participants used different smartphone applications to generate network traffic can lead to a better understanding of the diversity of traffic patterns.

Figures 4.14 and 4.15 show the HTTP traffic generated by various applications. The user-agent field was used to identify applications that generated traffic. For some cases, different applications generated similar user-agent field values which were distinguished by using host-name along with

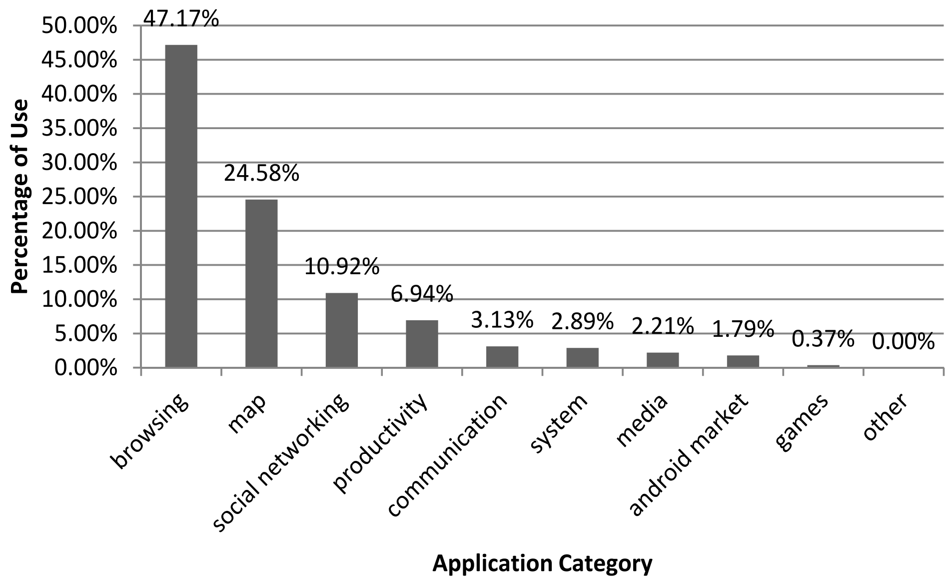


Figure 4.13: Relative Popularity of Application Categories

user-agent. Figure 4.14 shows the number of requests generated for different applications over different networks. The amount of data transferred for different applications over Wi-Fi and cellular networks are presented in Figure 4.15.

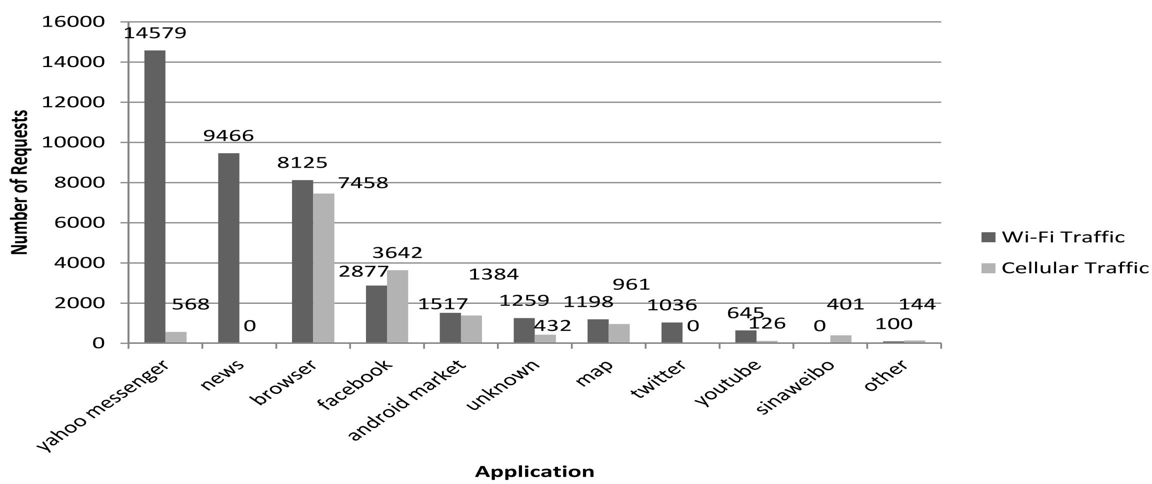


Figure 4.14: Applications Used Over Different Networks (Requests)

The above figures show that the news application is mostly used over Wi-Fi networks. However, the news application was only used by 2 participants during the study. Heavier usage of the youtube app was observed over Wi-Fi networks than the cellular network. Although, the number of requests generated by the Android Market over Wi-Fi networks is slightly higher than that of the cellular network, Figure 4.15 suggests more downloads were performed over a Wi-Fi network. The twitter application was only used over Wi-Fi networks and the sinaweibo application was only used over

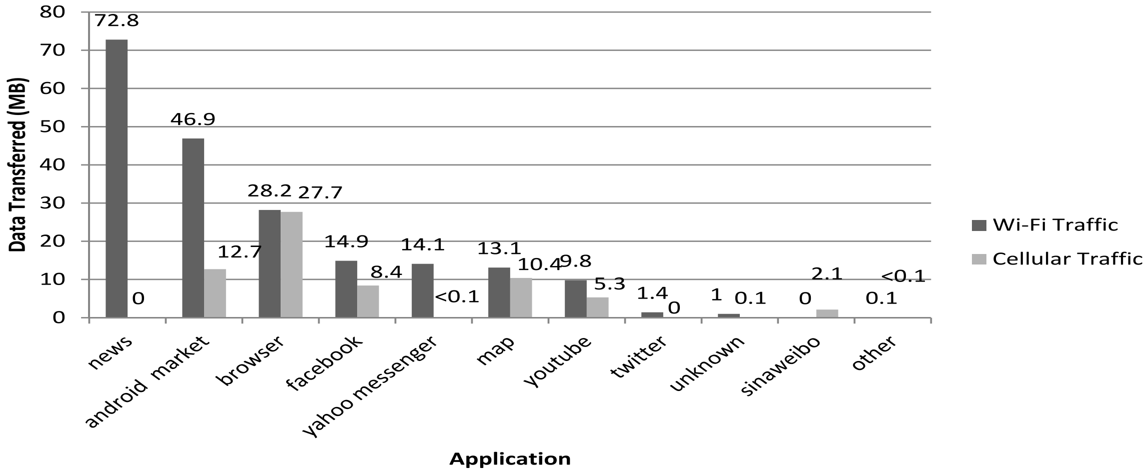


Figure 4.15: Applications Used Over Different Networks (Bytes)

the cellular network. However, the sinaweibo application was only used by a single participant. The browser application was essentially equally employed over both networks. The collected data suggests participants tend to use some applications more over Wi-Fi networks than over the cellular network.

4.2.2 Top Users

This section analyzes applications used by the top users. This will clarify application usage at the individual participant level and help in understanding what differences typically exist among individual participants. Participants are ranked according to their application use. The participant who generated greatest number of requests for different applications is considered as the first rank application user. The rest of the top users are determined in a similar fashion. Table 4.5 shows the participant id (assigned to participants at the beginning of the data collection) and rank of the top users. The participant id will be used to identify the participant for the rest of this thesis: P21 will always refer to the participant whose id is P21.

Table 4.5: Top Application Users

Participant ID	Application Usage Rank
P31	1
P13	2
P24	3
P30	4
P29	5

Figure 4.16 presents the application usage of the participant P31 (application usage rank=1).

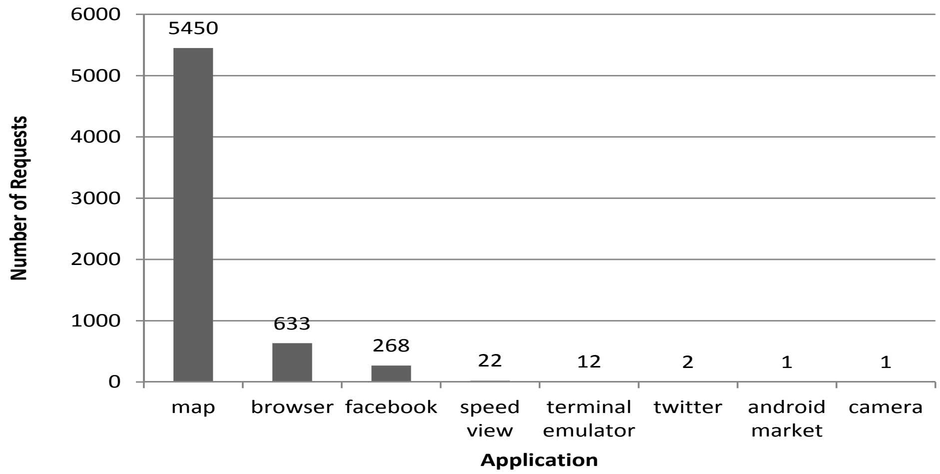


Figure 4.16: Application Usage of Participant P31 (Application Usage Rank=1)

The participant predominantly used map applications. Facebook and the browser are other two applications employed by participant P31. The participant used the terminal emulator more than the Android Market or twitter which illustrates a bias in the collected data set, particularly that participants were selected from the Computer Science Department, as we do not expect to see widespread console use in the larger population.

The application usage of participant P13 (application usage rank=2) is presented in Figure 4.17 which shows the dominance of the browser with 4723 requests. The participant also launched the youtube application a significant number of times. However, the rest of the applications were hardly used by the participant.

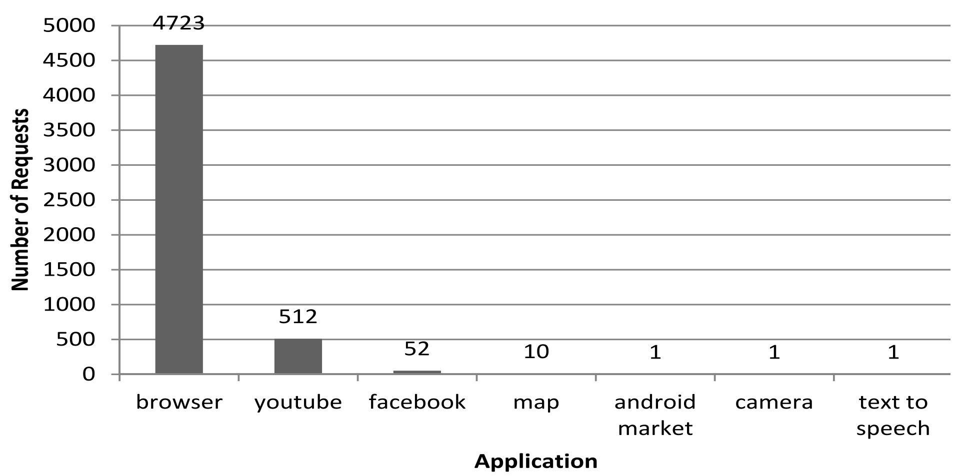


Figure 4.17: Application Usage of Participant P13 (Application Usage Rank=2)

The application usage of participant P24 (Application Usage Rank=3) was dominated by yahoo messenger and is presented in Figure 4.18. The task killer application accounted for 1266 requests which suggest that either the participant continuously ran task killer for long periods of time or this application was frequently launched for killing processes.

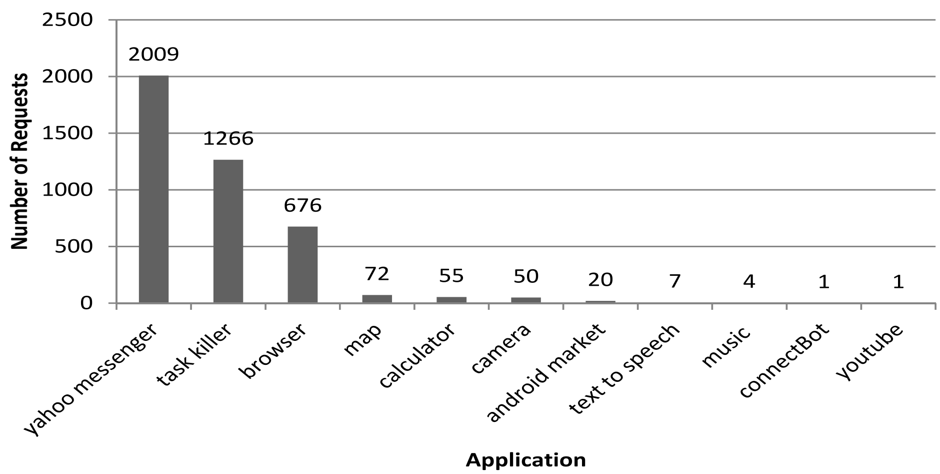


Figure 4.18: Application Usage of Participant P24 (application usage rank=3)

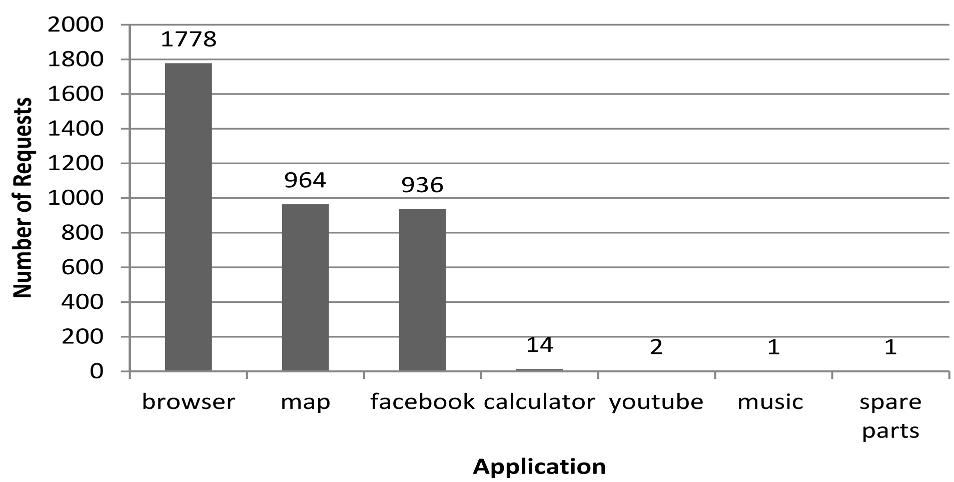


Figure 4.19: Application Usage of Participant P30 (application usage rank=4)

Figure 4.19 presents the application usage of participant P30 (Application Usage Rank=4). Top applications for the participant include the browser, map and facebook with 1778, 964 and 936 requests respectively. Therefore, it can be concluded that, unlike other top users, the participant used the popular applications more evenly.

Clear dominance of the browser application for participant P29 (Application Usage Rank=5) is observed in Figure 4.20. The map application accounted for 112 requests. However, surprisingly, the participant did not use any other applications.

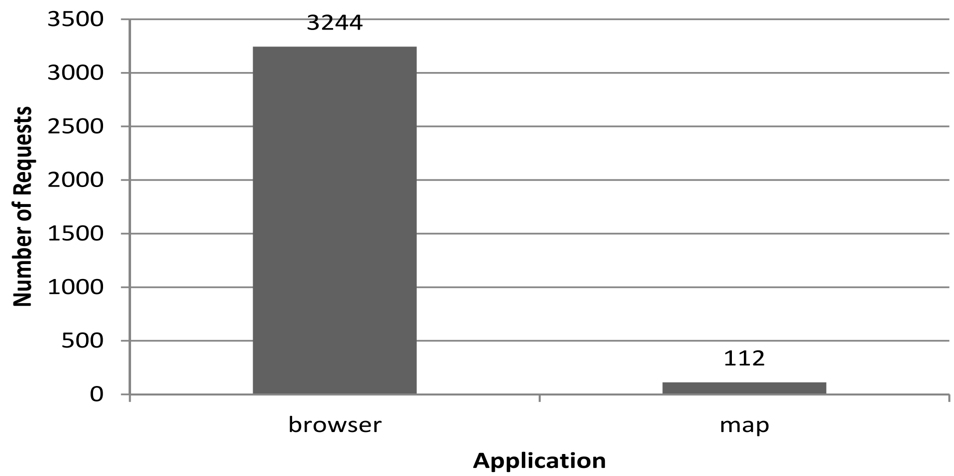


Figure 4.20: Application Usage of Participant P29 (application usage rank=5)

4.3 Popularity Analysis

This section presents a popularity analysis of the collected data over different networks. Popularity analysis refers to investigating the most popular websites visited by participants and the website usage patterns of top users. This will certainly provide insight about what kinds of websites participants like to visit over different networks and how frequently they used these websites.

4.3.1 Popular Websites

A common task in analysis is to determine if the popularity distribution of web page requests follows a Zipf-like model. Halvey *et al.* [20] reported that mobile web browsing follows a Zipf-like model. However, Adya *et al.* [1] suggested that, for a particular website, document popularity does not follow a Zipf-like model for cell phone browsing.

The popularity distribution of web pages is plotted in Figure 4.21 where the horizontal axis represents the page rank (by number of requests) on a log scale. The vertical axis refers to access frequency on a log scale which is actually the number of requests generated for each web page. From the figure we find that the popularity distribution follows a Zipf-like model for both Wi-Fi and cellular networks with a heavy tail. The primary reason behind such a distribution is the popularity of the top websites which accounted for a large portion of the traffic and some of the websites were only accessed a few times resulting in a lower byte transfer. Such heavy tails are also observed for previous dataset [20].

The popularity analysis of websites over different networks is presented in later sections. An analysis of website popularity for smartphone traffic is not presented elsewhere in the literature. However, a report published by Opera Software suggests that google, facebook and youtube are

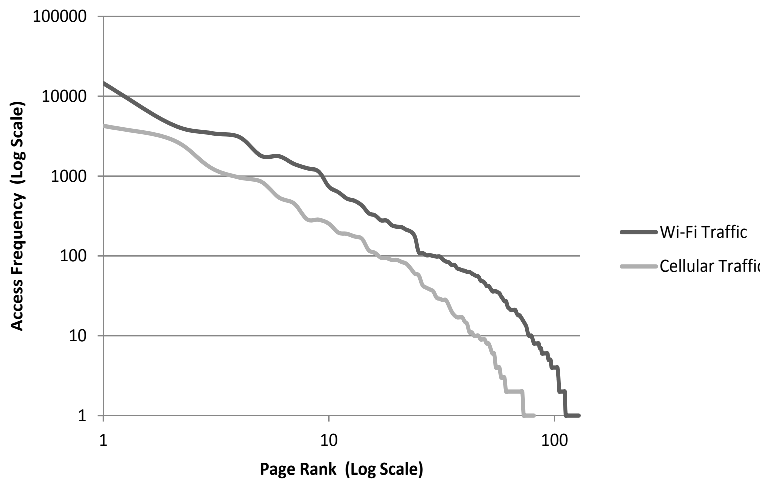


Figure 4.21: Popularity Distribution of Web Pages

the most popular sites among the users worldwide.[1]

Wi-Fi Network

Figure 4.22 shows the top websites over Wi-Fi networks by number of requests. Throughout the remainder of this thesis websites are named by striping off the scheme name (i.e. http and ftp) and the www part from the URL of the website's home page. Yahoo messenger is also included separately in the list of websites even though it is a messenger application. The top-level domain name portion of each website is also stripped off, if the top-level domain is dot-com. For example, the facebook website(http://www.facebook.com) is named 'facebook' in this thesis and the cbc website (http://www.cbc.ca) is named 'cbc.ca'. Figure 4.22 suggests that most of the requests were made to yahoo messenger, followed by facebook, google and vancouversun. However, yahoo messenger was used by a single participant and it constantly generated small automatic updates generating a high number of requests.

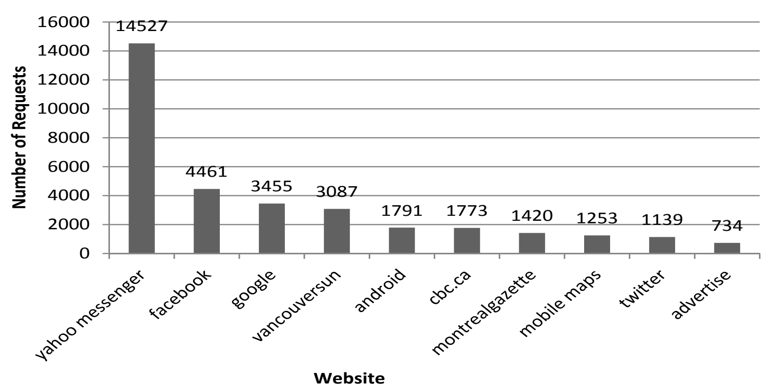


Figure 4.22: Top 10 Websites by Hits Accessed Using Wi-Fi Networks

[1] http://media.opera.com/media/smw/2011/pdf/smw052011.pdf, access date=18/8/2011

Figure 4.23 shows the top 10 websites for Wi-Fi networks by percentage of request bytes. Google and mobile maps accounted for 77.68% and 17.30% of the request bytes respectively. A closer look at the data revealed that most of the request bytes were generated for search on google and transferring location related information to google and mobile maps.

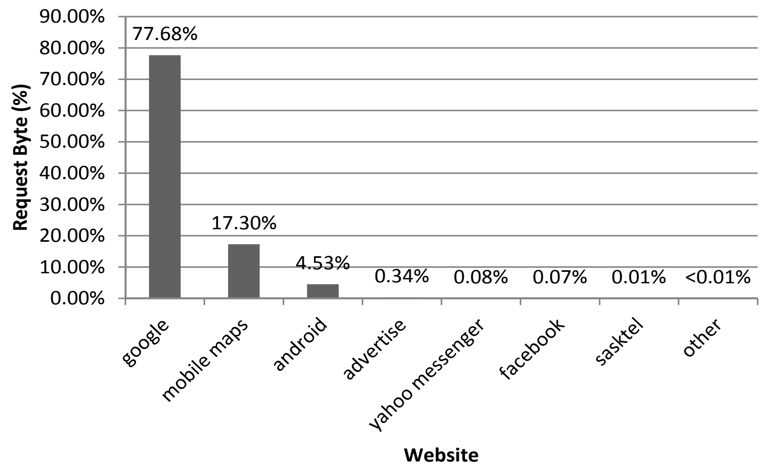


Figure 4.23: Top 10 Websites by Request Byte(%) Accessed Using Wi-Fi Networks

Figure 4.24 shows the top 10 websites contacted over Wi-Fi networks by percentage of response bytes. Android, used for downloading applications and packages for Android phones, was responsible for 22.90% of the bytes. Vancouversun, a news website, was responsible for generating 8.91% of the response bytes. Facebook, yahoo messenger and mobile maps accounted for 7.74%, 6.94% and 6.49% of the bytes respectively. Googlevideo generated 4.24% of the bytes from only a small number of requests.

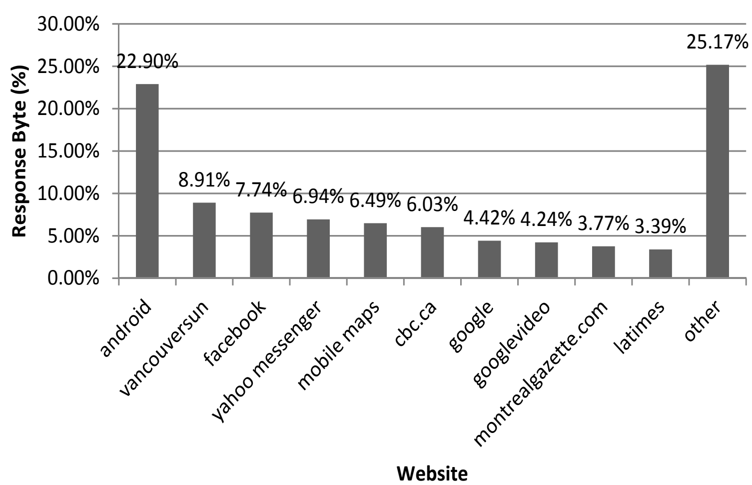


Figure 4.24: Top 10 Websites by Response Byte(%) Accessed Using Networks

The number of requests to particular categories of website are presented in Figure 4.25. Yahoo messenger and Skype were considered in the instant messenger category; news related websites such as vancouversun and montrealgazette in the news category; facebook and twitter as social networking websites; android in the apps category, and yahoo as a web portal. The instant messenger category accounted for the greatest number of requests followed by news, social networking and web browsing.

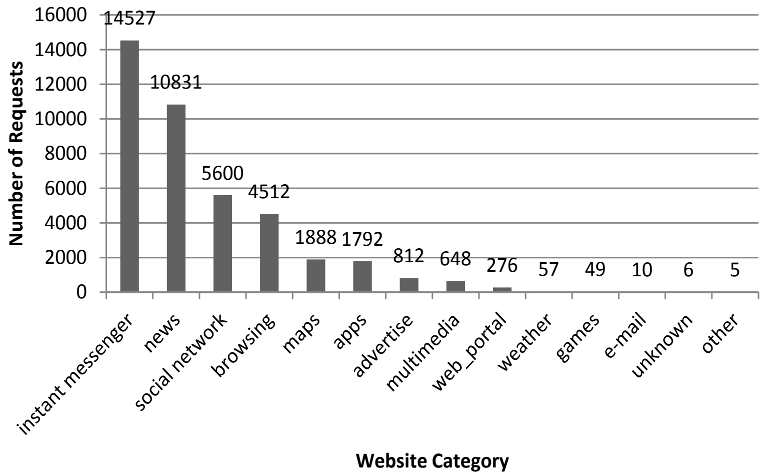


Figure 4.25: Categories of Websites by Requests Accessed Using Wi-Fi Networks

The percentage of request bytes generated by different website categories is presented in Figure 4.26. It shows search engine and location related services were responsible for 77.68% of the request bytes. Map related websites generated 17.3% of the bytes followed by the android website which accounted for 4.53% of the bytes.

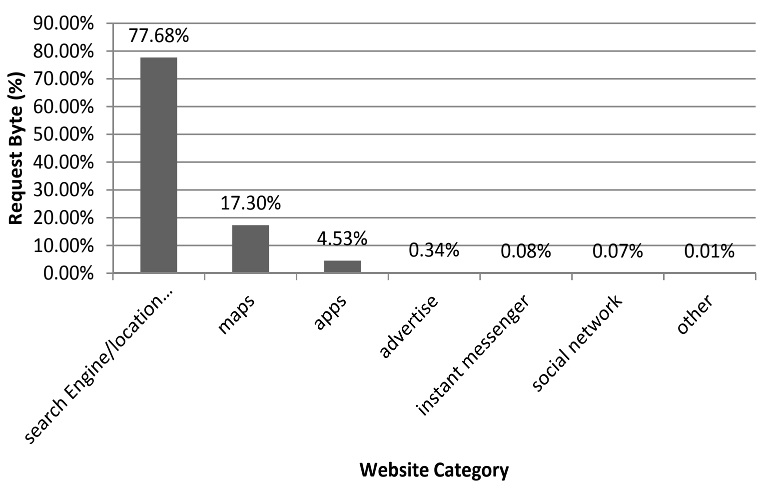


Figure 4.26: Categories of Websites by Request Byte (%) Accessed Using Wi-Fi Networks

Figure 4.27 shows the percentage of response bytes generated for different website categories. News related websites accounted for generating 38.58% of the response bytes. However, more detailed analysis revealed that one particular participant was generating most of the traffic by accessing many news websites. The participant used a news app to find news headlines and then followed articles. Websites responsible for downloading different applications and packages accounted for 22.91% of the response bytes followed by social networking (8.61% byte), browsing (8.42% byte) and instant messenger (6.94% byte).

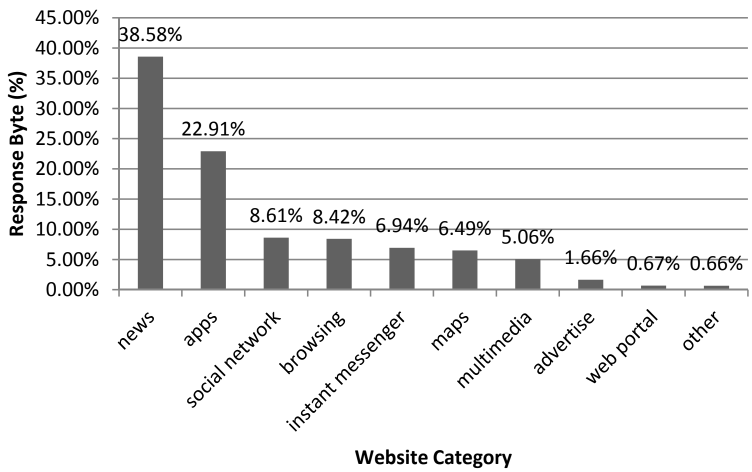


Figure 4.27: Categories of Websites by Response Byte (%) Accessed Using Wi-Fi Networks

Cellular Network

Figure 4.28 shows the top 10 websites accessed over the cellular network sorted by number of requests obtained for each website. Facebook accounted for 4251 requests and was considered as the top website, request-wise. Google, android and mobile maps accounted for 2886, 1265 and 960 requests respectively. While a Wi-Fi network was preferred by the participants for downloading applications, the collected data suggests many requests were also made to android over the cellular network.

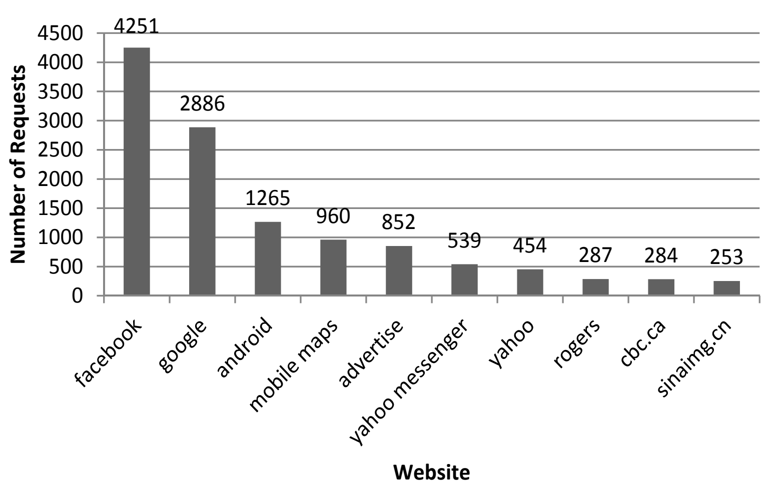


Figure 4.28: Top 10 Websites by Requests Accessed Using the Cellular Network

The top 10 websites for the cellular network, sorted by percentage of request bytes is presented in Figure 4.29. Google alone accounted for 89.38% of the bytes. A closer look at the data confirmed that most of these request bytes were generated for search queries and location related services. Mobile maps accounted for 7.47% of the request bytes.

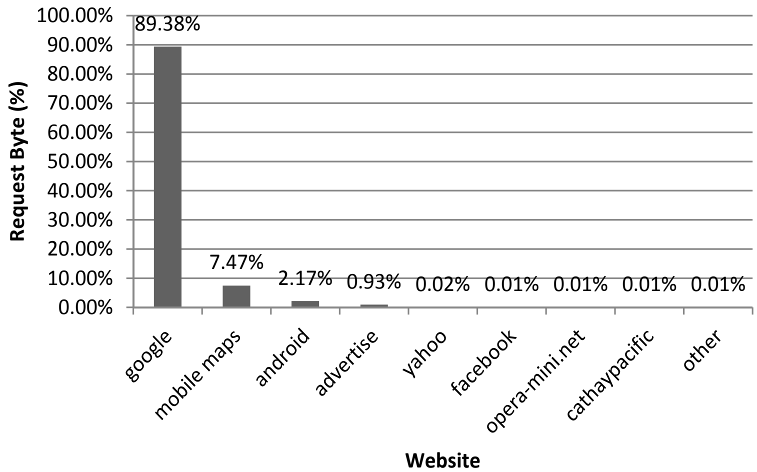


Figure 4.29: Top 10 Websites by Request byte(%) Accessed Using the Cellular Network

Figure 4.30 shows the top websites accessed over the cellular network ranked by percentage of response bytes. The figure clearly shows the dominance of android, mobile maps and facebook. Android accounted for 17.64% of the response bytes, implying a large number of application downloads were completed over the cellular network. Mobile maps accounted for 15.03% of the response bytes.

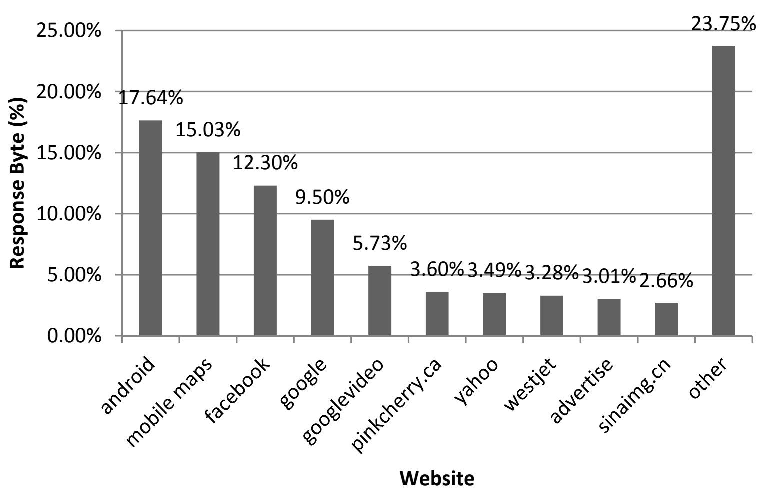


Figure 4.30: Top 10 Websites by Response Byte (%) Accessed Using the Cellular Network

The websites accessed over the cellular network are categorized using the categories as for the Wi-Fi network. Figure 4.31 gives the categories of websites and the number of requests generated for each category. Browsing and social networking websites accounted for 4865 and 4291 requests respectively.

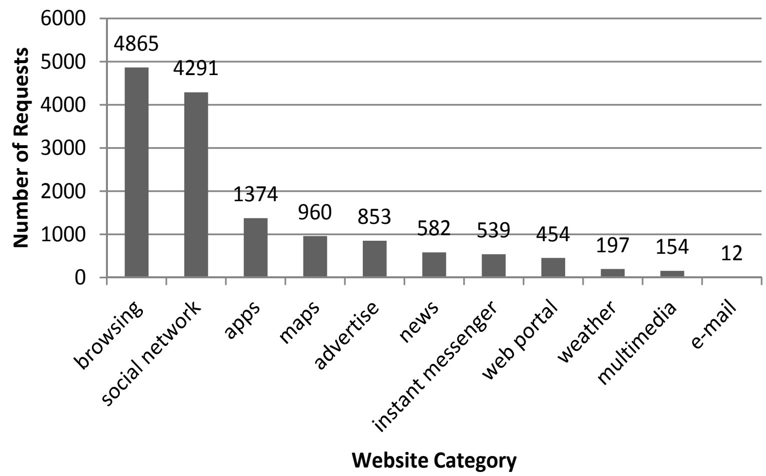


Figure 4.31: Categories of Websites by Requests Accessed Using the Cellular Network

Figure 4.32 shows the percentage of request bytes for different website categories. Search engines and location-dependent services were responsible for generating 89.39% of the request bytes and map applications accounted for 7.47% of the request bytes.

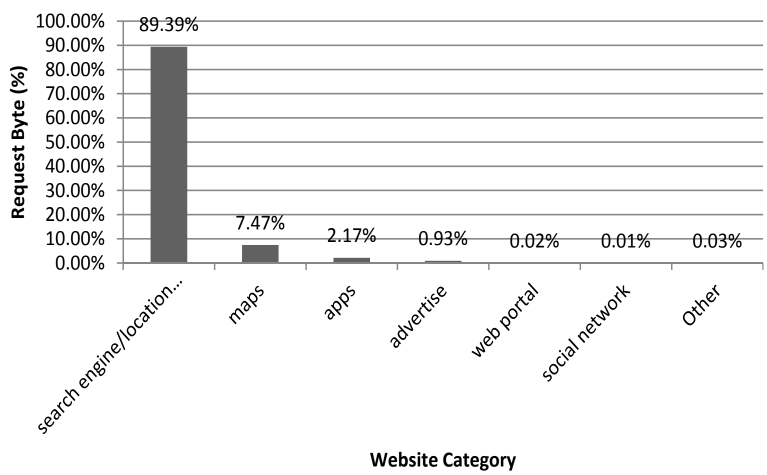


Figure 4.32: Categories of Websites by Request Byte (%) Accessed Using the Cellular Network

Figures for the percentage of the request bytes over both networks clearly show the dominance of search engines, location-based services and maps. Essentially, most of the bytes were generated for search queries or generated automatically for providing location information to the server.

Figure 4.33 gives the categories of websites and the corresponding percentage of response bytes for each category. The dominance of browsing, application download, map and social networking applications is observed over the cellular network. Browsing accounted for 33.12% of the response bytes, followed by application download, maps and social networking, generating 18.67%, 15.03% and 12.66% of the response bytes respectively.

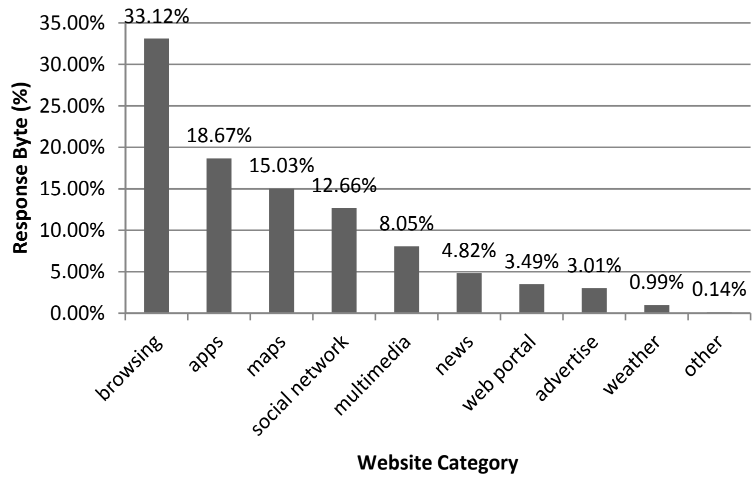


Figure 4.33: Categories of Websites by Response Byte (%) Accessed Using the Cellular Network

From the above discussion, some similarities and dissimilarities between usage of popular websites over Wi-Fi and cellular networks are observed. Intuitively, more applications were downloaded over Wi-Fi networks and greater map application usage occurred over the cellular network. Substantial use of an instant messager application occurred only on Wi-Fi networks, not on the cellular network. However, it is worth mentioning that only one participant used the messenger application. Therefore, the apparent correlation between network type and instant messenger usage could be due to the individual's habits and not a generalizable effect. Social networking applications were almost equally used over both networks. News related websites were also mostly visited over a Wi-Fi network which is again primarily due to a particular participant. Therefore, the only consistent difference between networks was the heavy usage of map application over the cellular network and application download over Wi-Fi networks.

4.3.2 Top Users

The websites accessed by the top users over different networks were examined to understand how different participants use network functionality and if there exist any differences among the usage patterns of the top users. The top 5 users contributed around 65% of the bytes over both networks and 65% of the requests over Wi-Fi networks and 46% of the requests over cellular network. Analyzing traffic generated by top users will help understanding basic characteristics of smartphone usage on an individual participant level. Such an analysis of websites accessed by top users over different networks has not been presented in the literature. Therefore, looking at the usage patterns of the top users will certainly provide a better insight to the dataset. One of the earlier studies also suggested that top 5% of the users are responsible for generating 90% of the traffic [42].

The top users were ranked by the number of requests they made for different networks. Therefore, the top ranked user of the Wi-Fi network refers to the participant who generated most requests over Wi-Fi networks. Similarly, the top ranked participant of the cellular network refers to the participant who generated the highest number of requests over the cellular network. Table 4.6 shows participant id and rank of the top 5 users of both networks. The table shows that the participant with id P7 had rank 3 for the cellular network and rank 15 with respect to Wi-Fi network usage. Interestingly, only one participant made both top 5 lists (P30). This suggests that the top users had different browsing habits, and that participants did not consume content equally across different networks.

Table 4.6: Top Ranked Users

Participant ID	Wi-Fi Network Rank	Cellular Network Rank
P7	15	3
P12	5	28
P16	12	5
P23	2	25
P24	1	7
P25	9	2
P30	4	1
P31	3	N/A
P32	10	4

Wi-Fi Network

The websites accessed by the top users of Wi-Fi networks are presented in this section. Figure 4.34 shows the websites used by participant P24 (Wi-Fi rank=1). The left side of the figure gives the number of requests generated and the right side of the figure gives the percentage of bytes transferred from the corresponding websites. From the figure it is evident that the participant primarily used yahoo messenger over Wi-Fi networks. Yahoo messenger accounted for 14527 requests and 92.49% of the bytes transferred. Android accounted for 413 requests and 4.20% of the bytes.

The websites accessed by participant P23 (Wi-Fi rank=2) are presented in Figure 4.35, which shows that the participant was primarily interested in reading news over Wi-Fi networks. A closer look at the traffic generated by the participant confirmed that a news application was used to obtain different headlines and then particular news articles were selected taking the participant to various news sites.

Figure 4.36 shows the websites accessed by participant P31 (Wi-Fi rank=3) over Wi-Fi networks. The participant primarily visited facebook, mobile maps and google. Participant P31 made 898

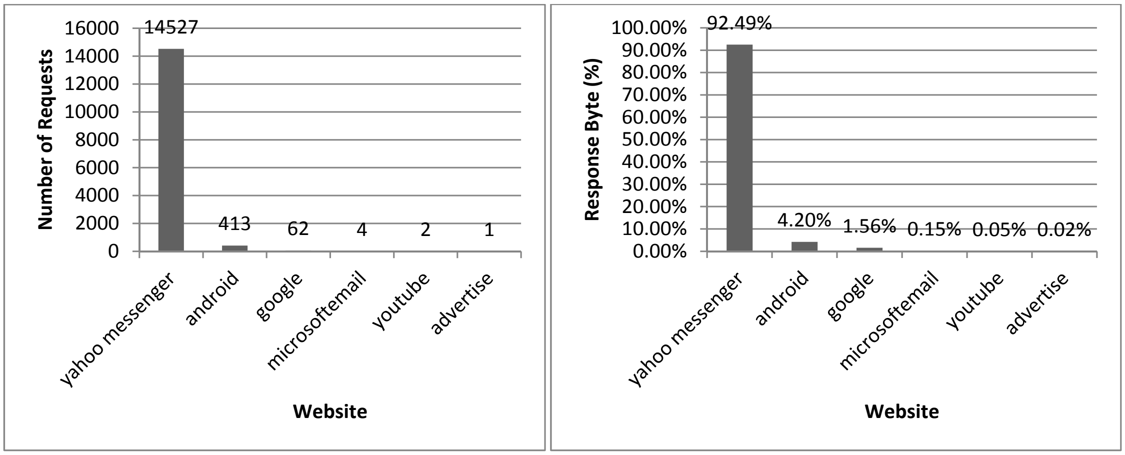


Figure 4.34: Websites Accessed by Participant P24 Over Wi-Fi Networks(Wi-Fi Network Rank=1)

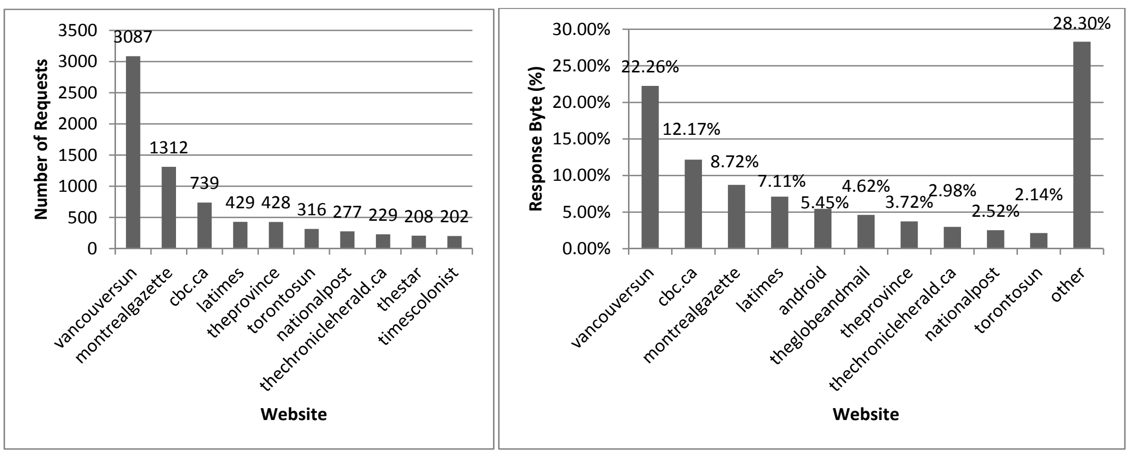


Figure 4.35: Websites Accessed by P23 Over Wi-Fi Networks (Wi-Fi Network Rank=2)

requests to mobile maps. The participant also made 619 requests to facebook and 466 requests to google. The percentage of response bytes generated by websites accessed over Wi-Fi networks are presented in the right hand side of Figure 4.36. Googlevideo accounted for 31.32% of the response bytes. However, only 7 requests made to googlevideo generated this percentage of byte transfer. Meanwhile android, mobile maps and facebook accounted for 25.08%, 19.05% and 13.54% of the response bytes respectively.

The websites accessed over Wi-Fi networks by participant P30 (Wi-Fi rank=4) are presented in Figure 4.37. The dominance of facebook over Wi-Fi networks is obvious in the graph. Facebook accounted for 819 requests and 63.86% of the bytes transferred. Google and yahoo constitute the majority of the remaining requests and bytes for P31.

Figure 4.38 gives the websites accessed by participant P12 (Wi-Fi rank=5) over Wi-Fi networks.

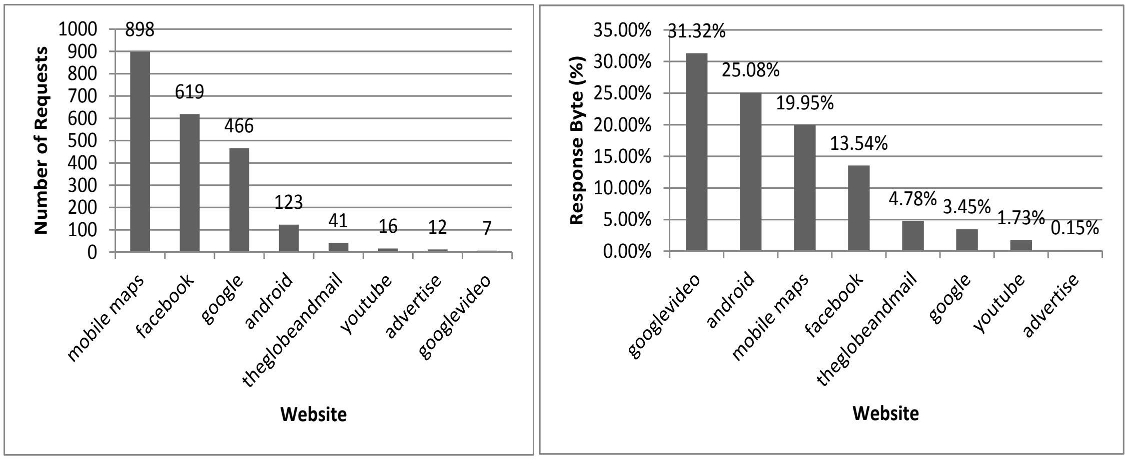


Figure 4.36: Websites Accessed by Participant P31 Over Wi-Fi Networks (Wi-Fi Network Rank=3)

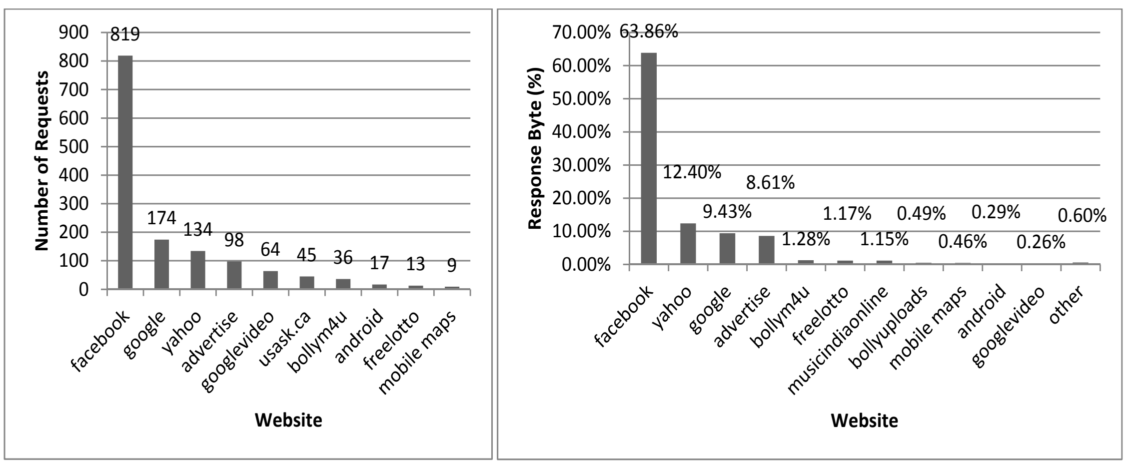


Figure 4.37: Websites Accessed by Participant P30 Over Wi-Fi Networks (Wi-Fi Network Rank=4)

The participant primarily used twitter over Wi-Fi networks. Twitter accounted for 1131 requests and 95.16% of the bytes transferred. Android only accounted for 21 requests and 3.36% of the bytes transferred.

From the above figures and discussion it is evident that substantial heterogeneity exists among participants. Different participants like to do different things over Wi-Fi networks. The collected data shows that the top 5 participants were interested in different applications and websites. Facebook, google and android were found as the mostly accessed websites across all participants. However, participant profiles are different if we consider all the websites they accessed. The top user used yahoo messenger, the second ranked primarily accessed news websites, the third ranked was more interested in mobile maps and facebook, the fourth ranked used facebook and yahoo, and, the fifth ranked participant used twitter. Such diversity among users for session related information

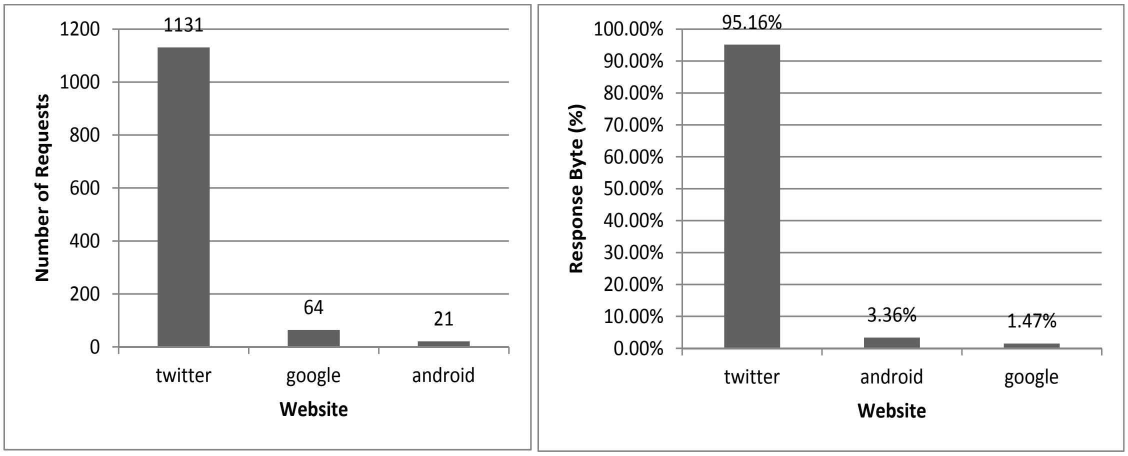


Figure 4.38: Websites Accessed by Participant P12 Over Wi-Fi Networks (Wi-Fi Network Rank=5)

such as session length or interaction time and application use is also mentioned by Falaki *et al.* [14]. However, the authors were primarily concerned about the overall application use of smartphones and did not concentrated on web access or network traffic for individual users.

The above results suggest that a global caching technique might not provide the best response time. Techniques which consider profile-specific information and make more user-centric decisions may provide better performance to the user.

Cellular Network

The websites accessed by the participants over the cellular network are presented in this section. Figure 4.39 presents websites accessed by participant P30 (cellular rank=1) over the cellular network. This participant used facebook over the cellular network which accounted for 1352 requests and 18.59% of the bytes transferred. Map applications and yahoo were also largely used by participant P30 which accounted for 15.71% and 12.40% of the bytes transferred respectively. Weather information was also accessed by the participant over the cellular network accounting for 196 requests and 3.91% of the bytes transferred. A significant percentage of requests were made to websites related to different airlines. Most likely the participant was looking for details of specific flights on different airline websites.

The websites accessed by participant P25 (cellular rank=2) over the cellular network are presented in Figure 4.40. Participant P25 used android and google over the cellular network. Android and google accounted for 48.96% and 14.15% of the bytes transferred respectively. However, a substantial use of cricket-related websites was also observed. The participant accessed espncricinfo and cricbizz both of which provide live scores and news related to the cricket world. These websites were mostly accessed during cricket world cup'2011 and shows how external events can effect web access patterns which is well known for wired networks [2].

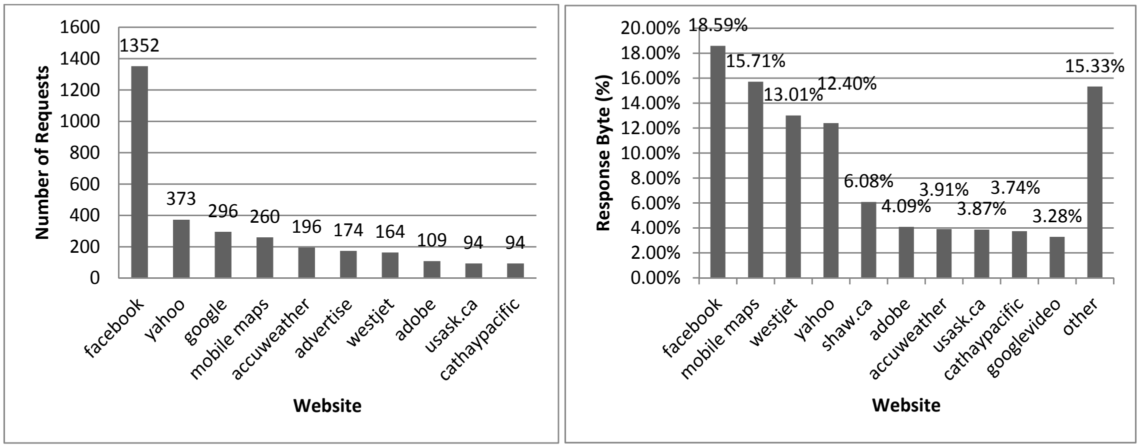


Figure 4.39: Websites Accessed by Participant P30 Over the Cellular Network (Cellular Network Rank=1)

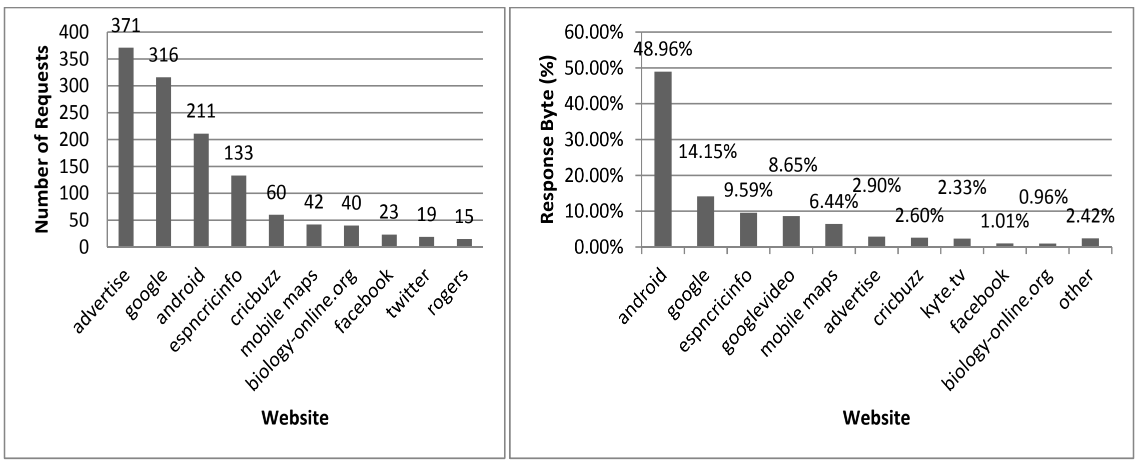


Figure 4.40: Websites Accessed by P25 Over the Cellular Network (Cellular Network Rank=2)

Figure 4.41 shows the websites accessed by participant P7 (cellular rank=3) over the cellular network. Facebook and mobile maps accounted for 33.87% and 29.43% of the bytes respectively. The participant also downloaded a significant number of applications over the cellular network accounting for 12.76% of the bytes. Only a handful of request to youtube generated 4.07% of the bytes for participant P7.

The websites accessed by participant P32 (cellular rank=4) over the cellular network are presented in Figure 4.42. The dominance of facebook and fsdn is also visible here. Facebook and fsdn accounted for 921 and 10 requests and 45.04% and 22.96% of the bytes respectively. Fsdn is a website for open source application developers.

Figure 4.43 presents the websites accessed by participant P16 (cellular rank=5) over the cellular network. Facebook and pinkcherry.ca visibly dominate. Facebook accounted for 639 requests and

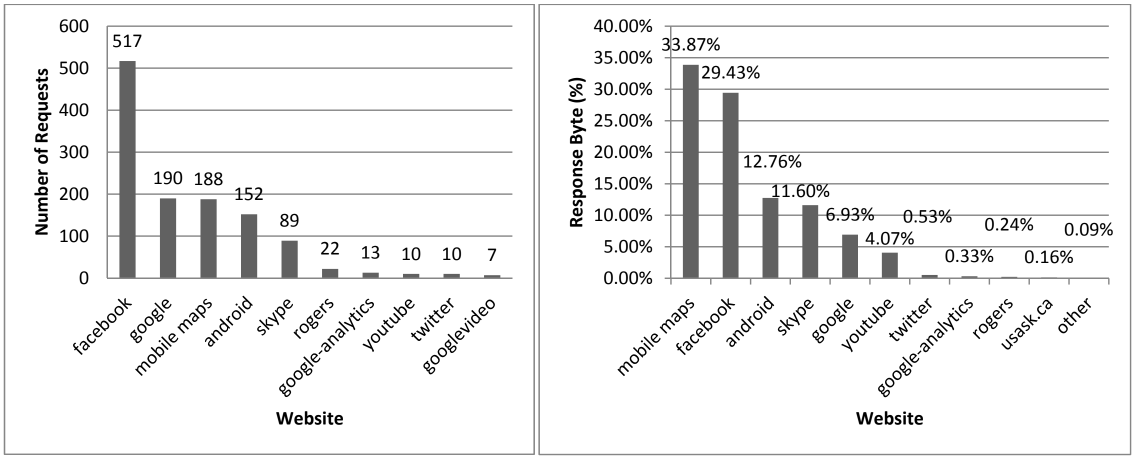


Figure 4.41: Websites Accessed by Participant P7 Over the Cellular Network (Cellular Network Rank=3)

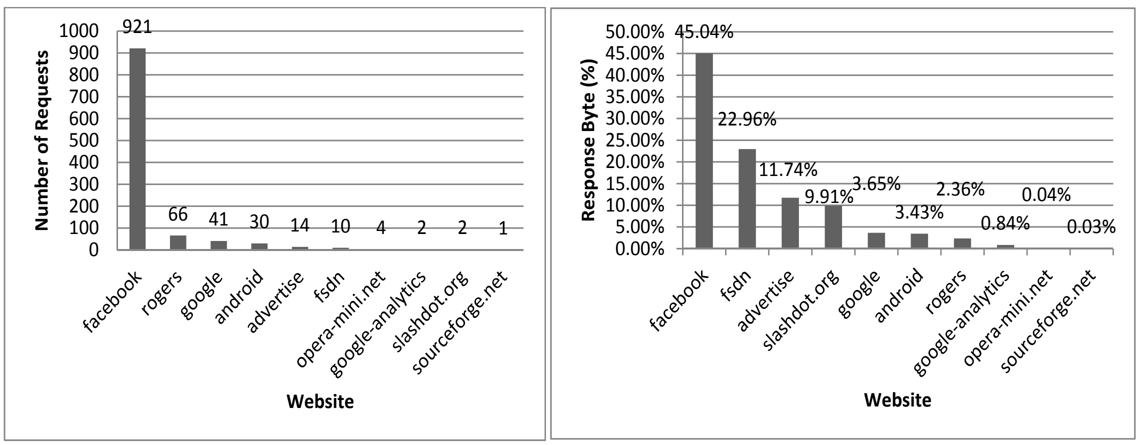
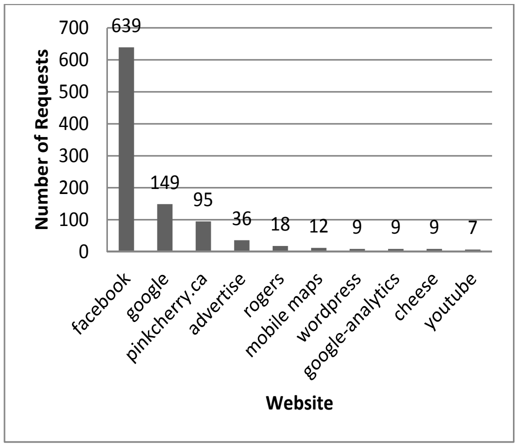


Figure 4.42: Websites Accessed by Participant P32 Over the Cellular Network (Cellular Network Rank=4)

18.55% of the bytes transferred. Pinkcherry.ca accounted for 95 requests and 48.61% of the bytes transferred. However, a closer look at the data revealed that pinkcherry.ca generated that amount of traffic in a single session. Substantial use of mobile maps and google is also observed from the figure.

From the above discussions and figures it might be said that there exists less diversity between participants with respect to their usage of the cellular network, compared to that for Wi-Fi networks. Participants predominantly used facebook and mobile maps over the cellular network with some exceptions.



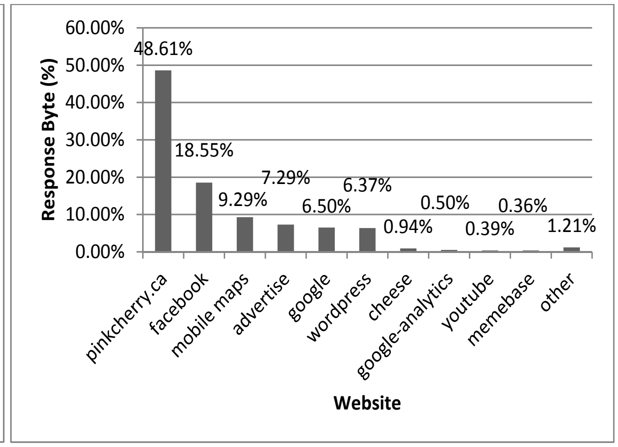


Figure 4.43: Websites Accessed by Participant P16 Over the Cellular Network (Cellular Network Rank=5)

The analysis presented in this section suggests that heterogeneity exists among the top users of Wi-Fi and cellular networks. However, the diversity of websites is lower for the cellular network than Wi-Fi networks. More similarities among the top users of the cellular network are observed.

Chapter 5

Context Dependencies

This chapter focuses on relationships between smartphone use and participant contexts as detected using the Healthlogger software. Of interest is whether participants browse particular websites and use specific applications under specific circumstances. The main objective here is to identify whether there exist dependencies between smartphone usage and different participant contexts. The type of network providing Internet connectivity, location, time-of-day, movement and proximity to other people are considered as different contexts.

5.1 Network Dependency

This section explores the relationship between network type and smartphone use. It is investigated whether particular participants tend to browse specific sets of websites and use particular applications while connected through Wi-Fi networks and some other sets of websites and applications while connected through the cellular network. Some similarities and dissimilarities between usage of Wi-Fi and cellular networks have been observed in the previous chapter. Individual participant level analyses are presented in this section to further study the impact of network type on smartphone usage. This type of analysis of network dependencies at the individual level is not found in the literature.

Figure 5.1 shows traffic volumes for the top Wi-Fi network hosts over both Wi-Fi and cellular networks. The figure suggests that news related websites were predominantly accessed over Wi-Fi networks. Android also contributed more traffic over Wi-Fi networks.

Figure 5.2 shows traffic volumes for the top cellular network hosts over both Wi-Fi and cellular networks. This provides some idea about how the top cellular network hosts were accessed over Wi-Fi networks. The figure suggests relatively more usage of mobile maps over the cellular network.

Figure 5.1 and 5.2 suggest that participants preferred to access some websites over a particular network. Individual usage patterns are presented in this section to determine network dependencies for an individual's smartphone usage. Initially the usage patterns of the top users of Wi-Fi networks will be observed to determine how they accessed both Wi-Fi and cellular networks. Later, usage patterns of the top users of the cellular network will be analyzed to understand the way they utilized

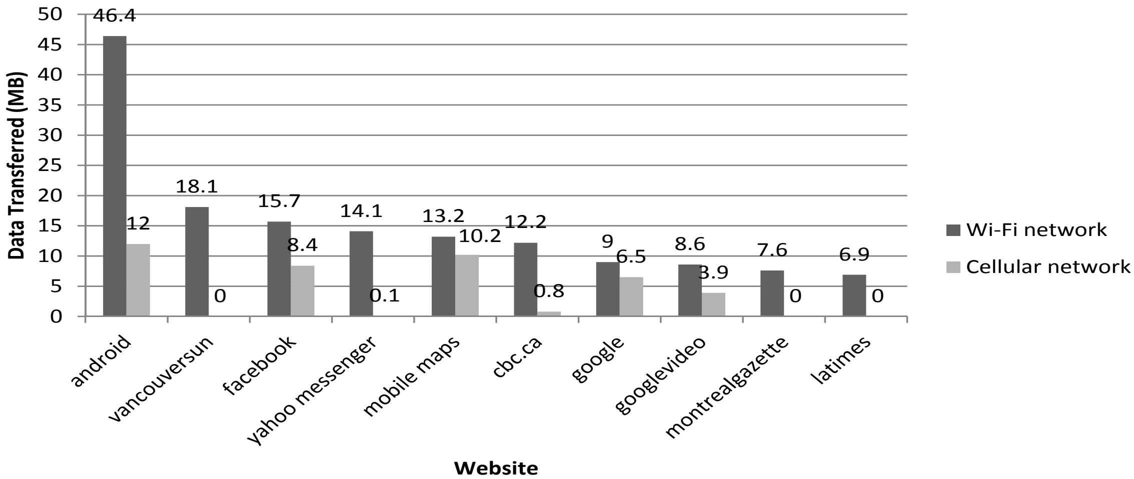


Figure 5.1: Data Transfer for Top Hosts Accessed Using Wi-Fi Networks

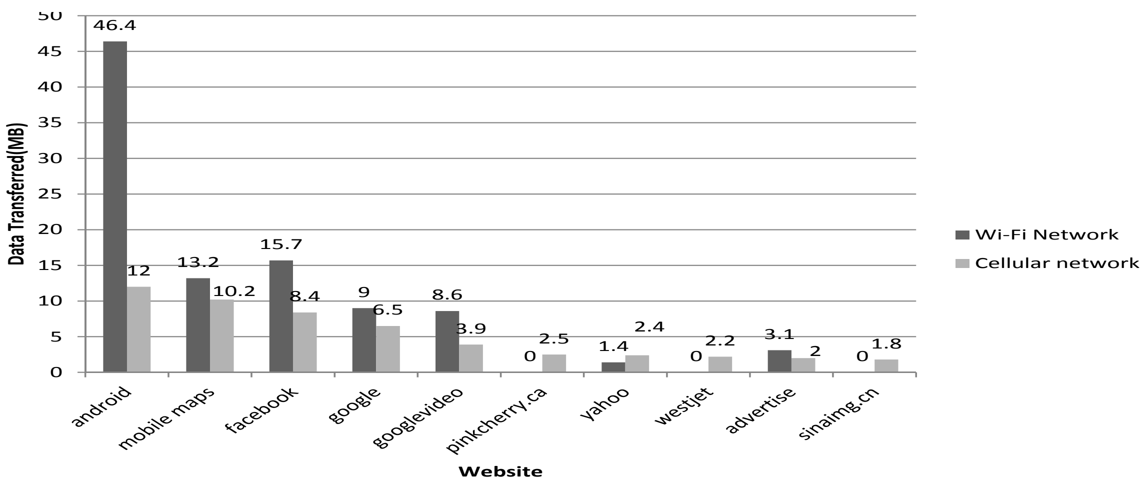


Figure 5.2: Data Transfer for Top Hosts Accessed Using the Cellular Network

both Wi-Fi and cellular networks.

The traffic volumes for different hosts accessed by participant P24 (Wi-Fi rank=1, cellular rank=7) is presented in Figure 5.3. Yahoo messenger was commonly accessed over Wi-Fi networks. Around 14 MB of data was transferred for yahoo messenger over Wi-Fi networks in contrast to only around 100 KB over the cellular network. Map websites were exclusively accessed over the cellular network. Heavier usage over Wi-Fi networks for the rest of the websites is also observed.

Figure 5.4 shows the websites accessed by participant P23 (Wi-Fi Rank=2, Cellular Rank=25) over different networks. This participant accessed news related websites only over Wi-Fi networks. Application downloads were primarily performed over a Wi-Fi network by the participant. However, map websites were only accessed over the cellular network.

The websites accessed by participant P30 (Wi-Fi Rank=4, Cellular Rank=1) over different networks are shown in Figure 5.5. Participant P30 was a heavy user of the cellular network and

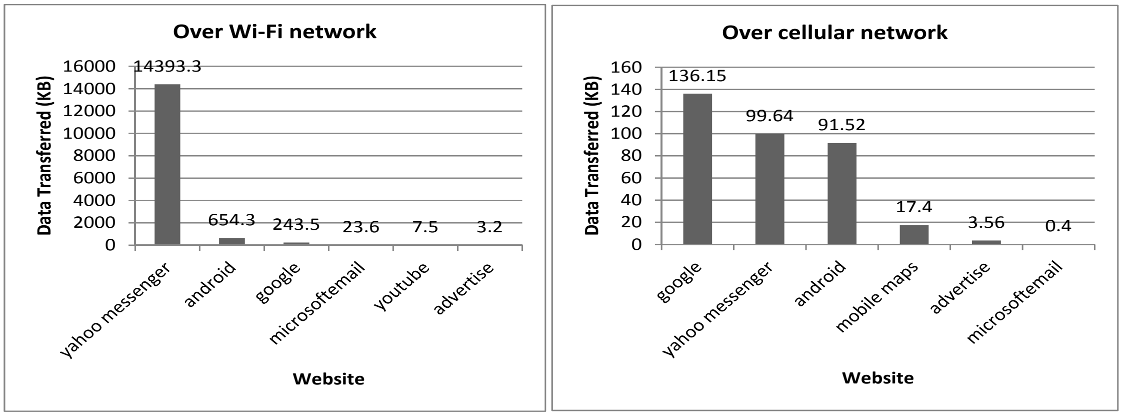


Figure 5.3: Data Transfer for Top Hosts for Participant P24 (Wi-Fi Rank=1, Cellular Rank=7)

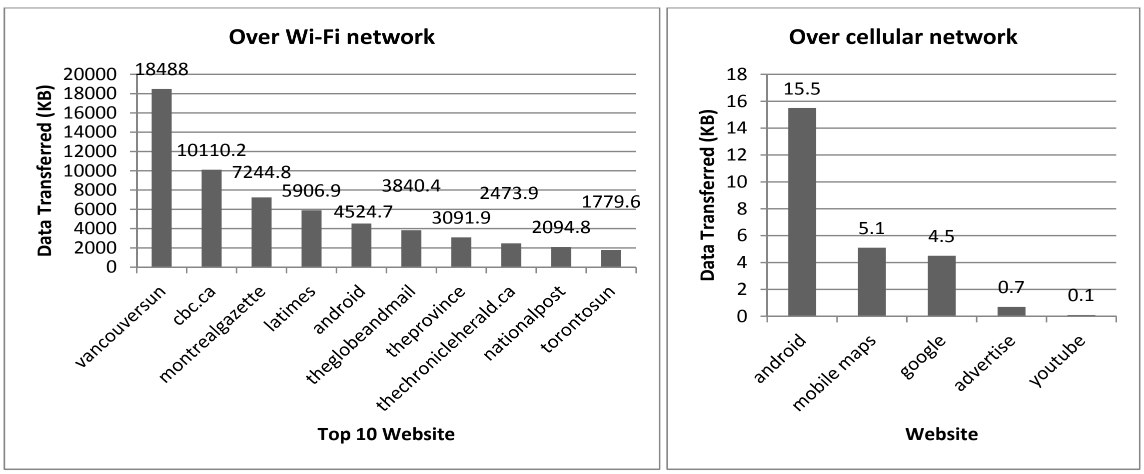


Figure 5.4: Data Transfer for Top Hosts for Participant P23 (Wi-Fi Rank=2, Cellular Rank=25)

mainly transferred data over the cellular network. However, a substantial amount of data was also transferred over Wi-Fi networks. The participant's Wi-Fi network usage was dominated by facebook, google and yahoo. However, data transfer volumes for those sites over the cellular network were much higher. Map, weather and airline related websites were mainly accessed through the cellular network. The university website and googlevideo website was also accessed mostly over the cellular network.

Figure 5.6 presents the usage pattern of participant P12 (Wi-Fi Rank=5, Cellular Rank=28) over different networks. This participant used twitter and android only over Wi-Fi networks and the map website only over the cellular network which suggests a strong network dependency for the usage patterns of the participant.

The usage patterns of the top users of the cellular network is now analyzed. The top ranked participant of the cellular network and the fourth ranked participant of Wi-Fi networks are the

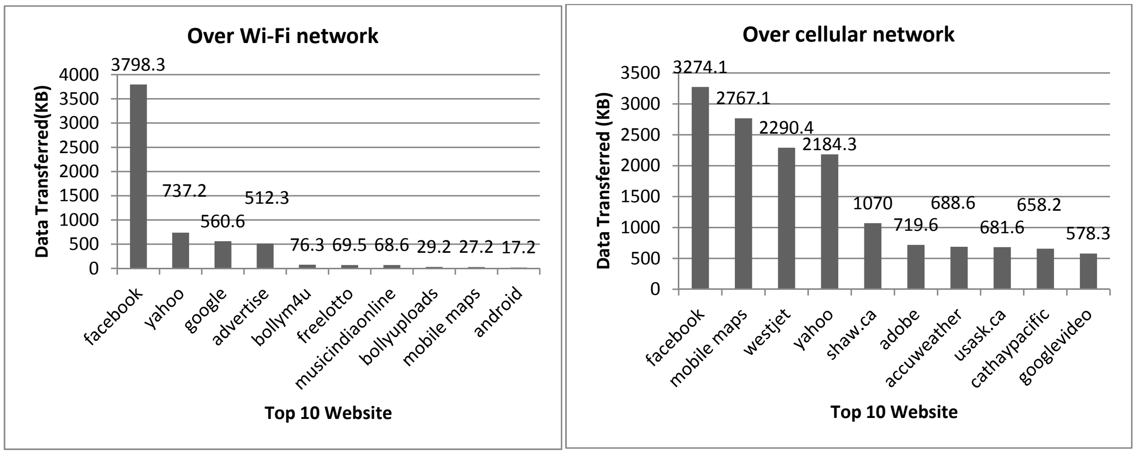


Figure 5.5: Data Transfer for Top Hosts for Participant P30 (Wi-Fi Rank=4, Cellular Rank=1)

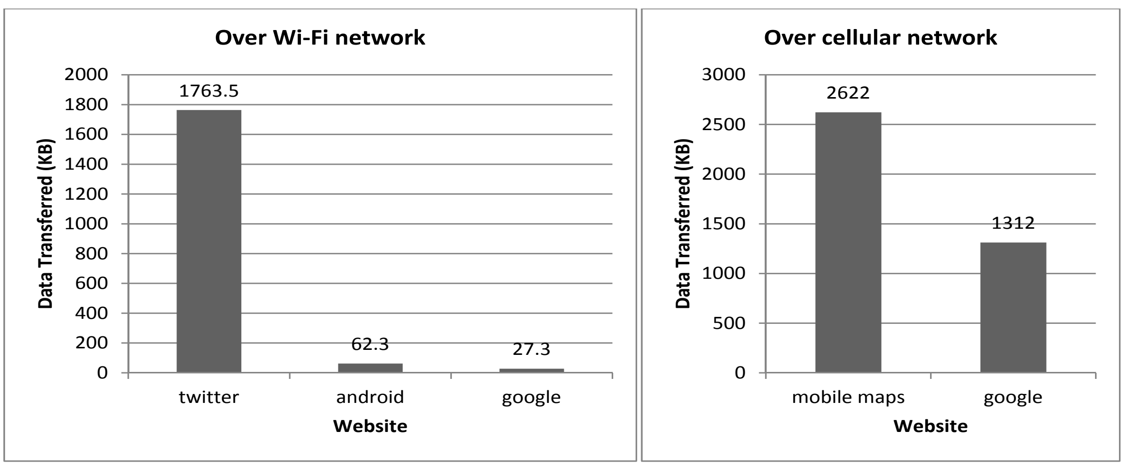


Figure 5.6: Data Transfer for Top Hosts for Participant P12 (Wi-Fi Rank=5, Cellular Rank=28)

same participant. Therefore, that usage pattern is not presented again.

The usage pattern of participant P25 (Wi-Fi Rank=9, Cellular Rank=2) is presented in Figure 5.7. Downloads from android, and facebook access occurred preferentially over Wi-Fi networks rather than the cellular network. However, cricket-related websites such as espncricinfo and cricbuzz were mainly accessed over the cellular network which suggests that the participant might have followed some matches and checked live scores. The participant also accessed google and googlevideo more over the cellular network than Wi-Fi networks.

Figure 5.8 shows the websites accessed by participant P7 (Wi-Fi Rank=15, Cellular Rank=3) over different networks. Map websites and facebook were primarily accessed over the cellular network by the participant. The participant downloaded more applications over the cellular network than over Wi-Fi networks. The participant only accessed websites such as wikipedia and wikimedia

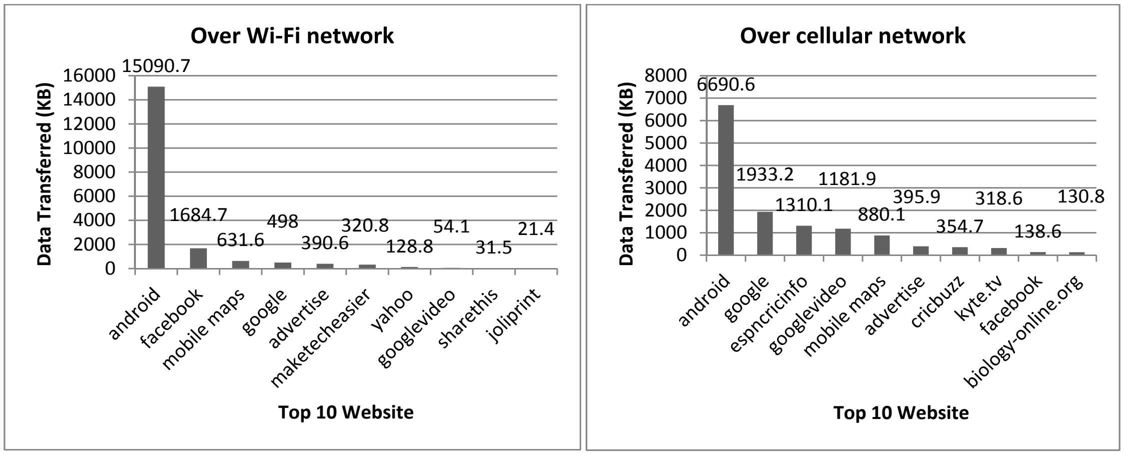


Figure 5.7: Data Transfer for Top Hosts for Participant P25 (Wi-Fi Rank=9, Cellular Rank=2)

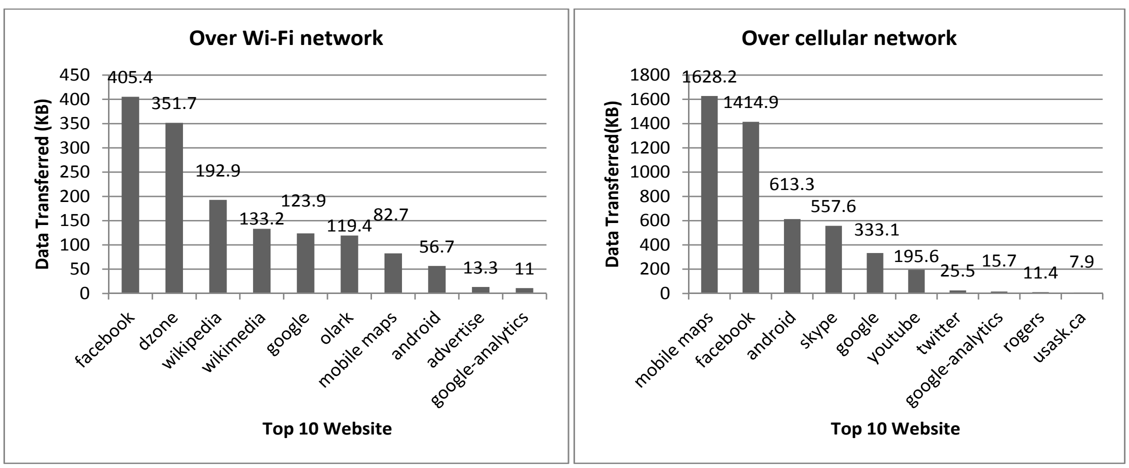


Figure 5.8: Data Transfer for Top Hosts for Participant P7 (Wi-Fi Rank=15, Cellular Rank=3)

over Wi-Fi networks.

The websites accessed by participant P32 (Wi-Fi Rank=10, Cellular Rank=4) over Wi-Fi and cellular networks are given in Figure 5.9. Heavy facebook use over both Wi-Fi and cellular networks is observed. However, unlike the other participants, participant P32 used map websites only over Wi-Fi networks. Websites such as fsdn and slashdot.org are only accessed over the cellular network by the participant.

The traffic volumes generated by participant P16 (Wi-Fi Rank=12, Cellular Rank=5) are shown in Figure 5.10. The participant transferred more facebook data over Wi-Fi than over the cellular network. Map application use over the cellular network is also observed from the figure. Websites such as pinkcherry.ca and chese were only accessed over the cellular network.

The above analysis shows substantial differences in usage patterns for different participants

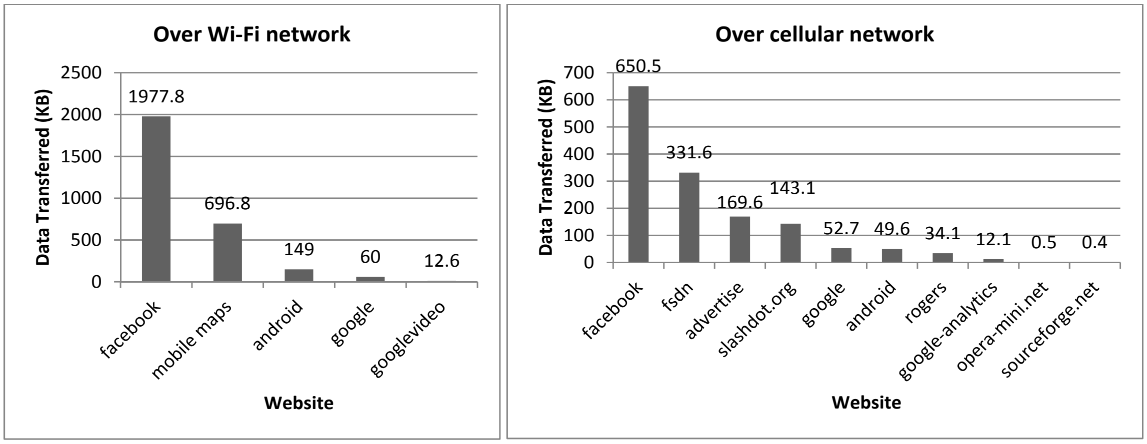


Figure 5.9: Data Transfer for Top Hosts for Participant P32 (Wi-Fi Rank=10, Cellular Rank=4)

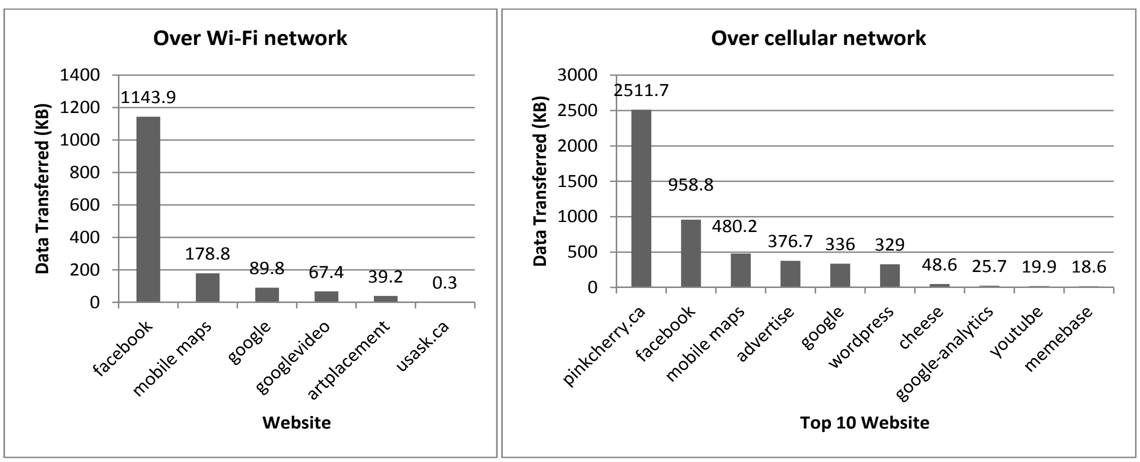


Figure 5.10: Data Transfer for Top Hosts for Participant P16 (Wi-Fi Rank=12, Cellular Rank=5)

over different networks. Most of the heavy users accessed different sets of websites while connected through different networks. Participants tended to do specific tasks over one network. However, these network dependencies might be observed because of location dependencies, which are investigated in the next section.

5.2 Location Dependency

This section explores relationships between a participant's location and his smartphone usage. Of interest is whether there are relationships between being at home versus at work/school, and the websites participants visit. The set of websites visited while at work or school (the University of Saskatchewan for these participants) is compared with the set of websites participants visit off-campus. Participants predominantly used Wi-Fi networks to connect to Internet while they were at home or work/school. Therefore, Wi-Fi traffic is analyzed in this section to identify potential relationships between location and smartphone usage. Wi-Fi access point and GPS information is used to separate Wi-Fi traffic between traffic generated at the University of Saskatchewan (U. of S.) and traffic generated outside university campus. 'U. of S.' and 'Off-Campus' are used to denote the location of traffic in all of the following figures.

Table 5.1 presents the percentage of time participants spent at different locations and how they accessed the Internet through Wi-Fi networks at each locations. The table shows that 30.10% of the time participants were at the U. of S. and 69.90% of time they were off-campus. 55.75% of the requests and 39.21% of the bytes were transferred from the U. of S. through Wi-Fi networks. Off-campus locations accounted for 44.26% of the requests and 60.79% of the bytes. These records suggest that participants completed larger transfers from off-campus locations.

Table 5.1: Participants at Different Locations

Location	**Time Spent (%)**	**Request Generated (%)**	**Byte Transferred (%)**
U. of S.	30.10	55.74	39.21
Off-campus	69.90	44.26	60.79

Data transfer at different times of day at the U. of S. and off-campus is presented in Figure 5.11. Most of the traffic at the university campus was between 9 am and 5 pm, as well as around 7 pm and 10 pm to midnight. The high night time traffic volumes suggest some participants (most likely the graduate students) stayed till midnight at the university campus and used the phones for connecting to the Internet. The figure also shows more usage outside the university during evenings, nights and mornings.

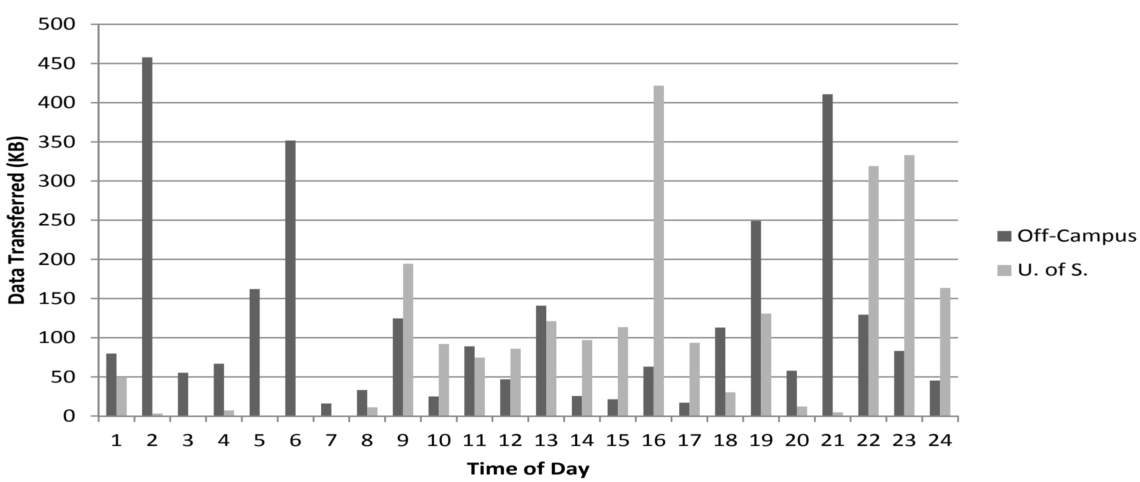


Figure 5.11: Data Transfer at Different Times of the Day

Traffic volumes on different days of the week for U. of S. and Off-Campus usage are presented in Figure 5.12. Traffic on the university campus during weekdays was substantial although the traffic volume on Fridays was lower. Little traffic at the U. of S. accounted on Saturdays. Surprisingly, the highest amount of U. of S. traffic occurred on Sundays.

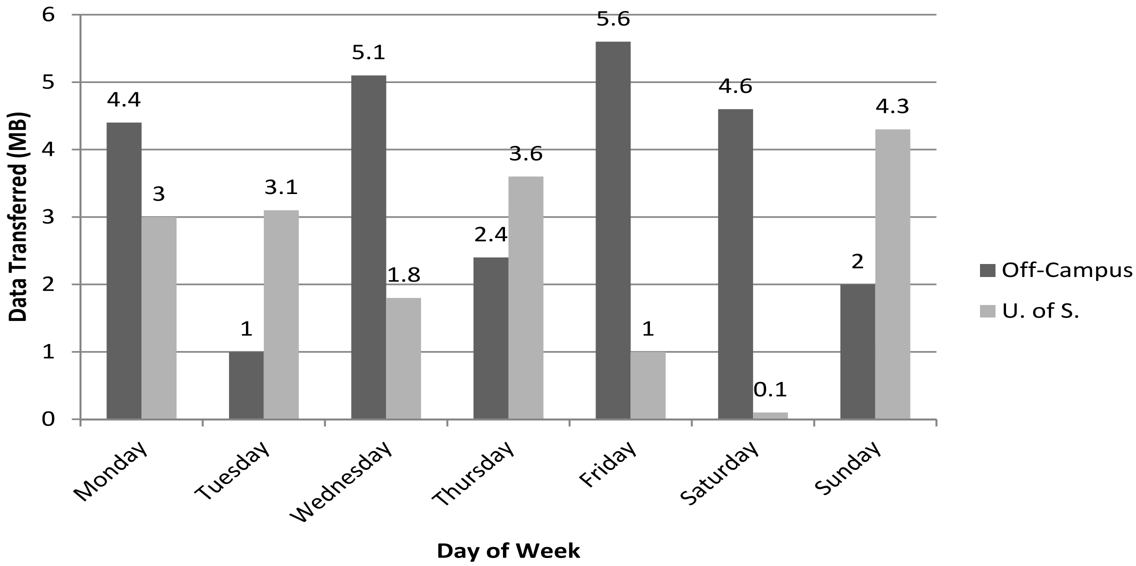


Figure 5.12: Data Transfer at Different Days of the Week

Figure 5.13 breaks down the Wi-Fi traffic by content type and shows the percentage of traffic of each type at the U. of S. and off-campus. One important thing to observe is the consumption of video content. Video contributed 8.07% of the traffic while participants were off-campus. In contrast, only 0.1% of the traffic was generated when participants were at the university campus. This implies that the participants did not normally watch videos on their phone at their workplace or school.

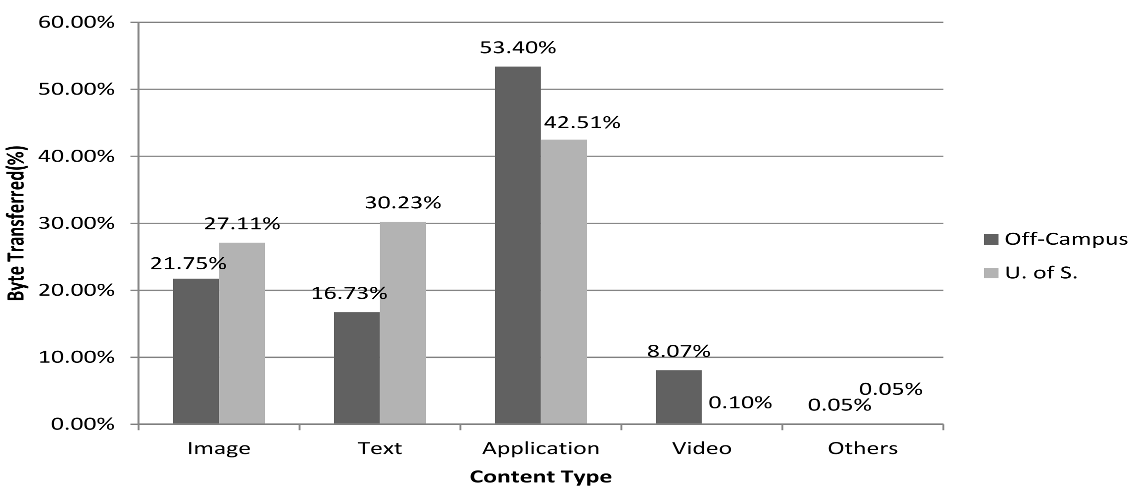


Figure 5.13: Content Type for Wi-Fi Traffic at Different Locations

The top hosts accessed on or off the U. of S. campus are presented in Figure 5.14. Android was used more off-campus than on-campus. In contrast, yahoo messenger was primarily used at the university campus. Heavy usage of googlevideo is only present off-campus.

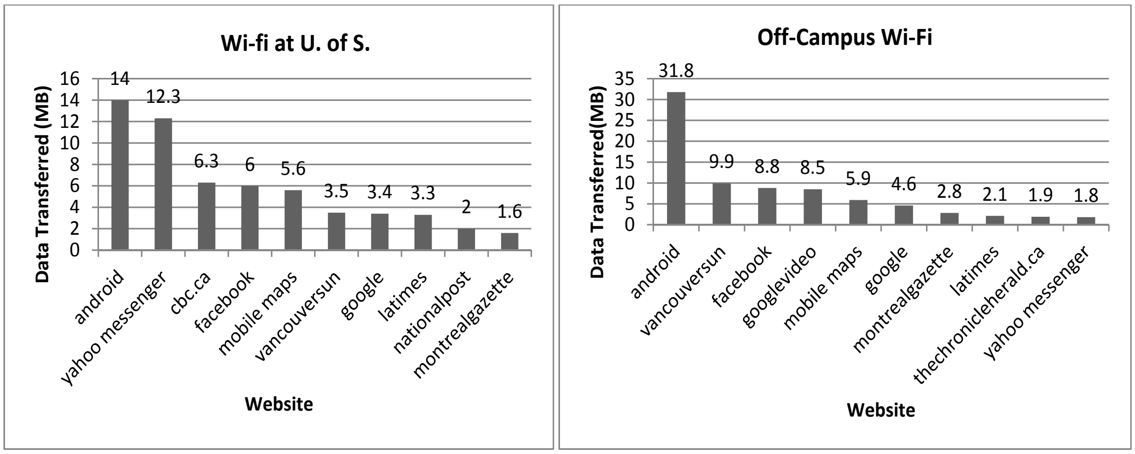


Figure 5.14: Top Websites for Wi-Fi Traffic at Different Locations

The categories of websites accessed at different locations are presented in Figure 5.15. Application download contributed more data off-campus. Multimedia, web portal and weather related websites were also primarily accessed from off-campus. Instant messenger was primarily used at the university campus. However, almost no difference between browser and map applications at different locations is observed.

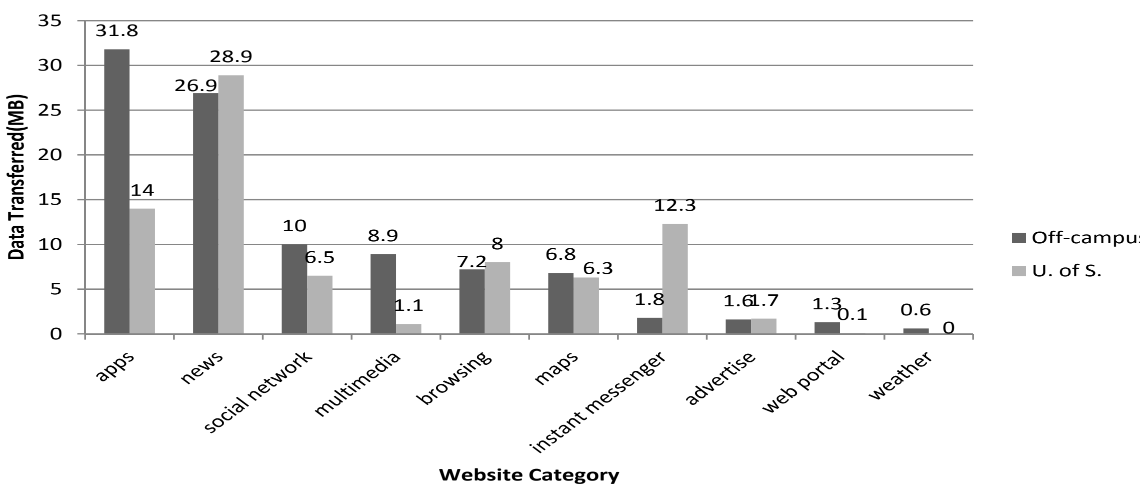


Figure 5.15: Website Categories for Wi-Fi traffic at Different Locations

The results suggest that there exists some differences between smartphone usage of participants when they were on the university campus and when they were outside the campus. The observed lower consumption of multimedia content suggests participants usually did not prefer to enjoy videos at their workplace or school.

Individual usage patterns of the top users will be investigated now to determine if location dependencies exist on an individual level. The usage patterns of the top users of Wi-Fi networks are used for analysis.

The usage pattern of participant P24 (Wi-Fi rank=1, cellular rank=7) is presented in Figure 5.16. The participant made usage of yahoo messenger on the university campus, but not in other locations. The participant also downloaded most applications while on-campus.

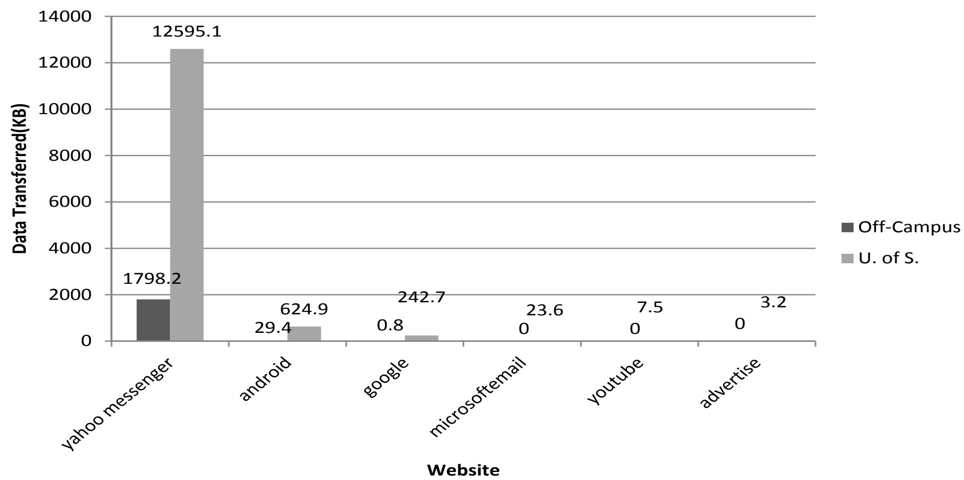


Figure 5.16: Data Transfer for Top Hosts for Participant P24 (Wi-Fi Rank=1)

The websites accessed by participant P23 (Wi-Fi rank=2, cellular rank=25) at different locations is presented in Figure 5.17. Most of this participants' traffic was from visits to news websites. No dependencies have been observed between location and news website access for participant P23 because the participant used a news aggregator application from both locations. However, some differences for other websites are observed. The participant primarily downloaded applications while on-campus, while map websites were primarily accessed off-campus.

Figure 5.18 shows the locations from which participant P23 accessed the Internet via Wi-Fi networks as indicated by the recorded GPS and Wi-Fi AP data, and the websites visited at each location. The colors of the location markers indicate the visited websites as shown at the bottom of the figure. The figure shows that the participant accessed all these websites from a single location outside the campus, most likely the participant's home. Therefore, the participant's usage can be considered as usage at workplace/school and home which suggests that the participant primarily used map applications from home, preferred downloading applications from U. of S. and accessed news websites equally from both locations. These information might be used for designing a caching policy suitable for users like participant P23.

The amount of traffic generated from visits to various websites by participant P31 (Wi-Fi rank=3) at different locations is presented in Figure 5.19. Participant P31 primarily used network

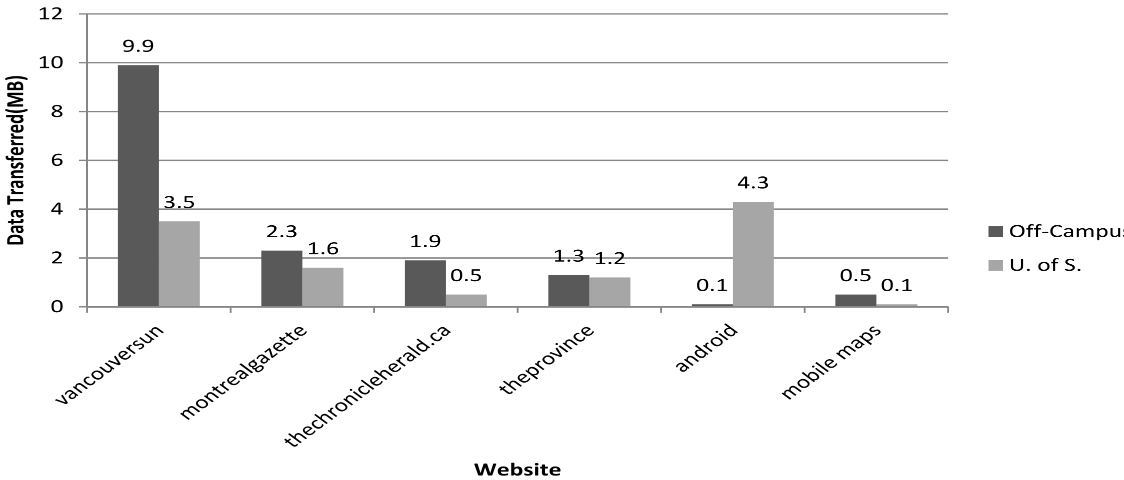


Figure 5.17: Data Transfer for Top Hosts for Participant P23 (Wi-Fi Rank=2)

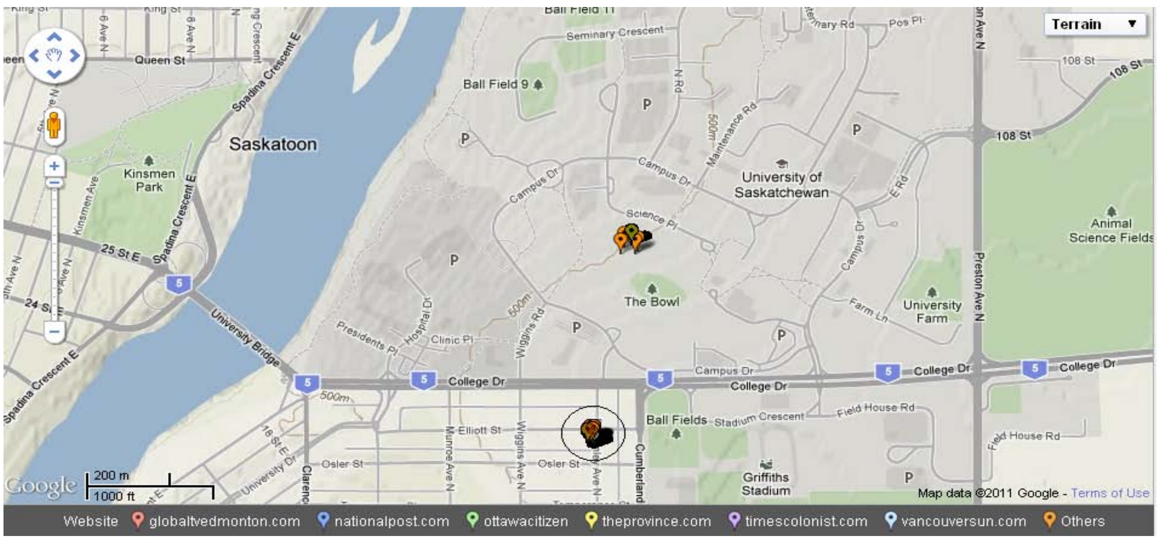


Figure 5.18: Websites Accessed on Wi-Fi Networks by Participant P23 at Different Locations

functionality when off-campus, particularly googlevideo and android. Mobile maps and facebook were also accessed off-campus.

Figure 5.20 presents the recorded GPS locations of participant P31, from which the participant accessed the Internet through Wi-Fi networks, and the websites accessed at these locations. Unlike the other top ranked Wi-Fi users, P31 accessed Wi-Fi from locations other than the university campus and home.

The usage pattern of participant P30 (Wi-Fi rank=4, cellular rank=1) is presented in Figure 5.21. The figure shows that the participant mostly accessed Wi-Fi networks off-campus. The participant only accessed facebook, google and the university website on-campus. Most websites were accessed from off-campus and generated a significant amount of traffic.

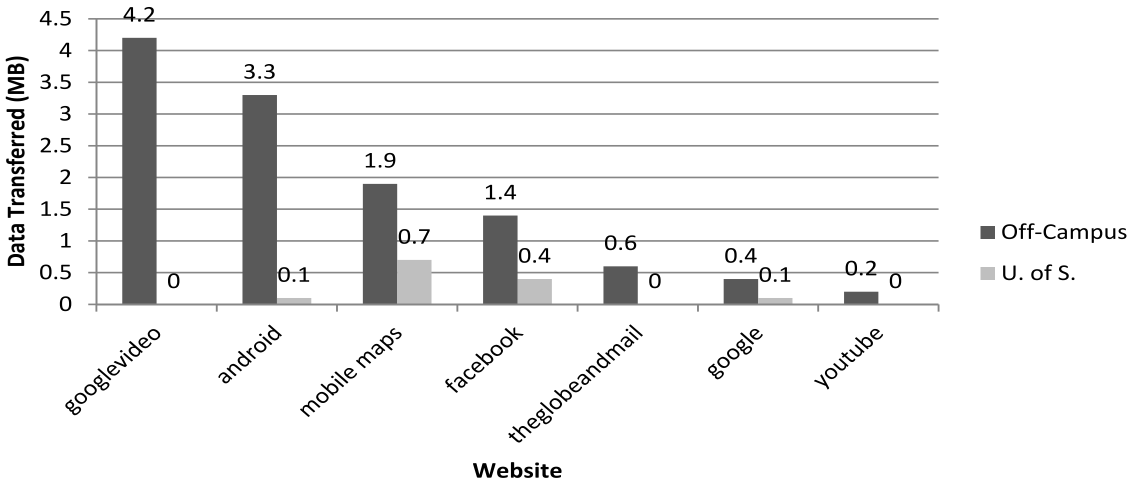


Figure 5.19: Data Transfer for Top Hosts for Participant P31 (Wi-Fi Rank=3)

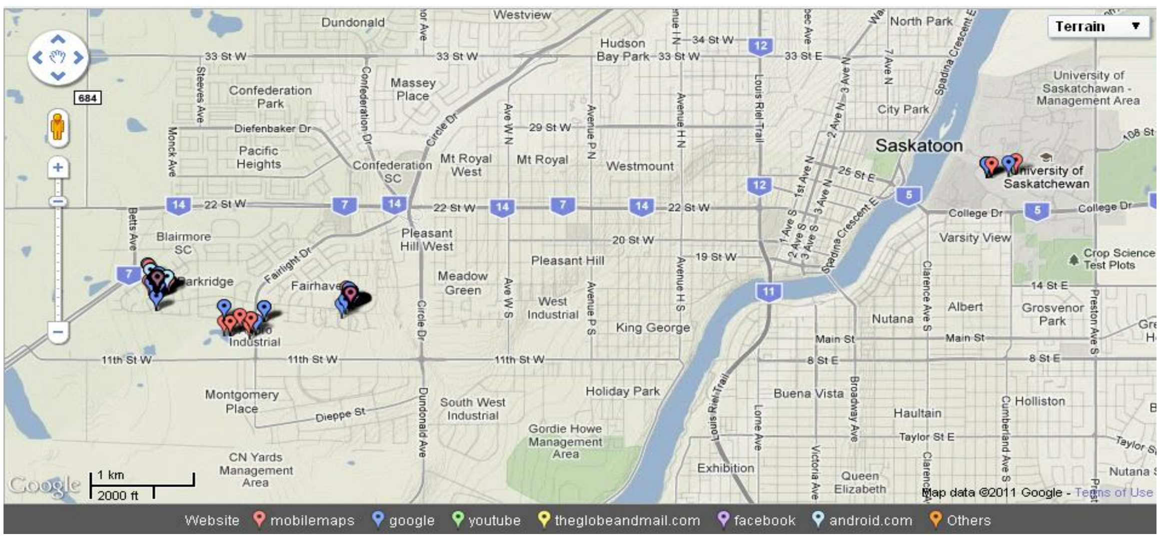


Figure 5.20: Websites Accessed On Wi-Fi Networks by Participant P31 at Different Locations

The recorded GPS locations of participant P30 while accessing different websites via Wi-Fi networks are presented in Figure 5.22. From the figure it is evident that the participant performed off-campus network activities from one single location, most likely the participant's home. Therefore, the participant's usage can be separated into usage at home and usage at the workplace/school. The participant's usage patterns indicate that most usage occurred at home.

Figure 5.23 presents websites accessed by participant P12 (Wi-Fi rank=5, cellular rank=28) on-campus and off-campus. Participant P12 also consumed more data when outside the university campus. The participant used twitter and google more when off-campus. Applications were only downloaded off-campus.

Figure 5.24 presents the recorded GPS locations of participant P12 while the participant was

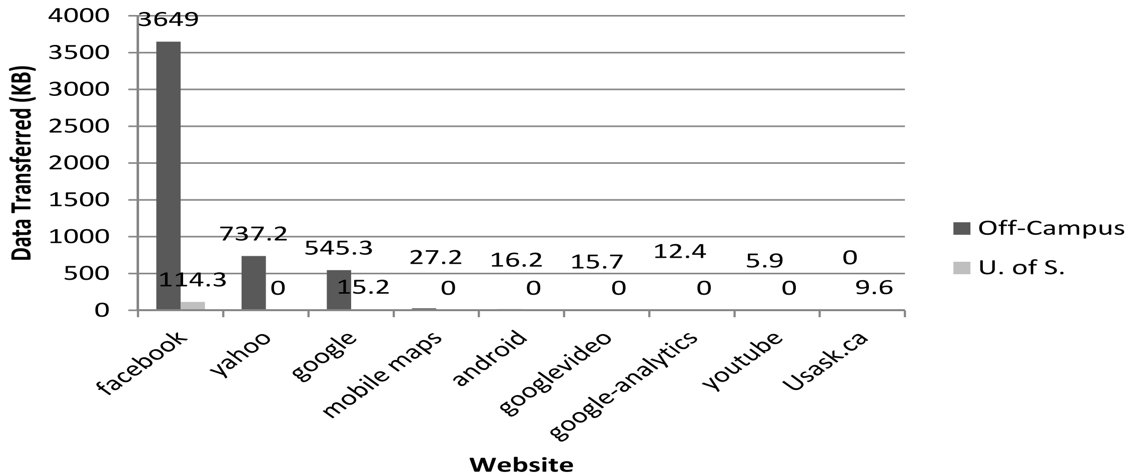


Figure 5.21: Data Transfer for Top Hosts for Participant P30 (Wi-Fi Rank=4)

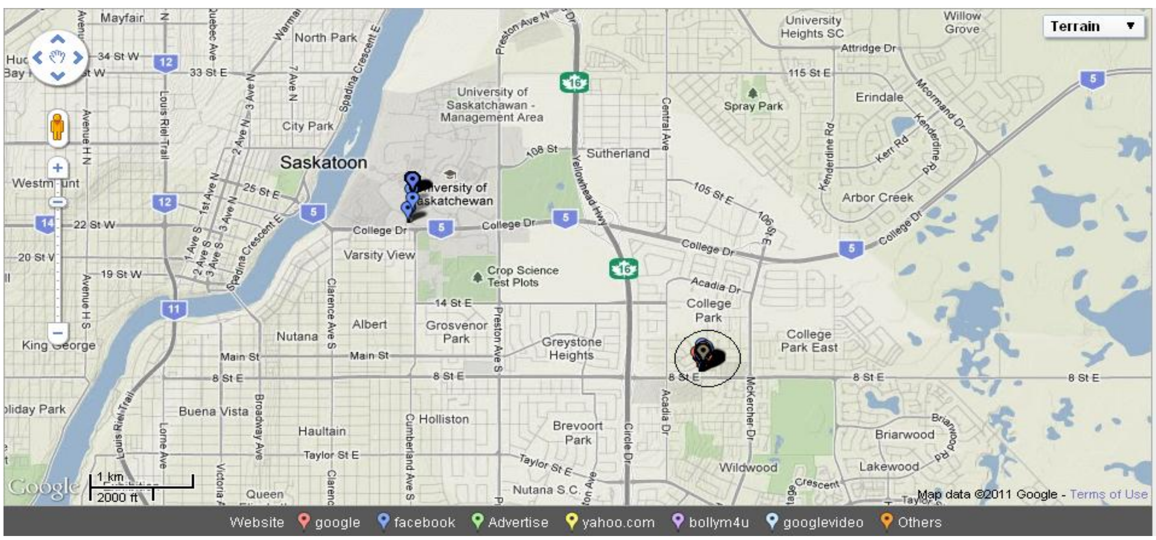


Figure 5.22: Websites Accessed On Wi-Fi Networks by Participant P30 at Different Locations

using Wi-Fi networks to access websites. The participant accessed a Wi-Fi network from one single location other than the university campus, again likely the participant's home.

The results suggest some relationship exists between participant's location and the types of websites they visit via Wi-Fi. Participants tended to consume less multimedia when they were at work or school, and more at home. Location-dependent browsing patterns were also found at the individual level. Individual participants preferred to access particular sets of websites or applications at home while others did so at school or the workplace. Some location dependencies have been investigated in the literature [44, 43]. Trestian *et al.* [44] reported location-dependent application usage; however, they collected traffic from only the cellular network and hence could not capture the usage over Wi-Fi networks. Shepard *et al.* [43] also reported location-dependent

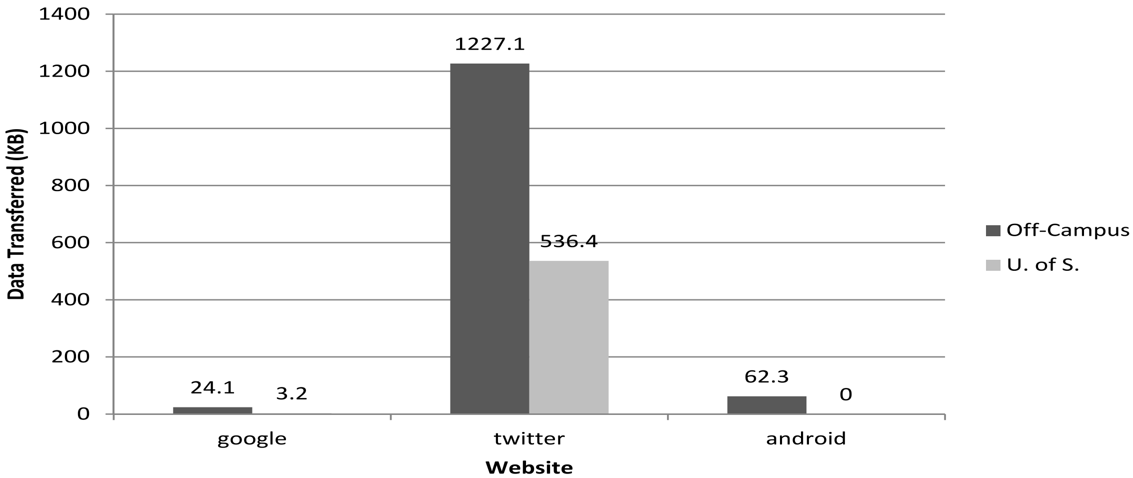


Figure 5.23: Data Transfer for Top Hosts for Participant P12 (Wi-Fi Rank=5)

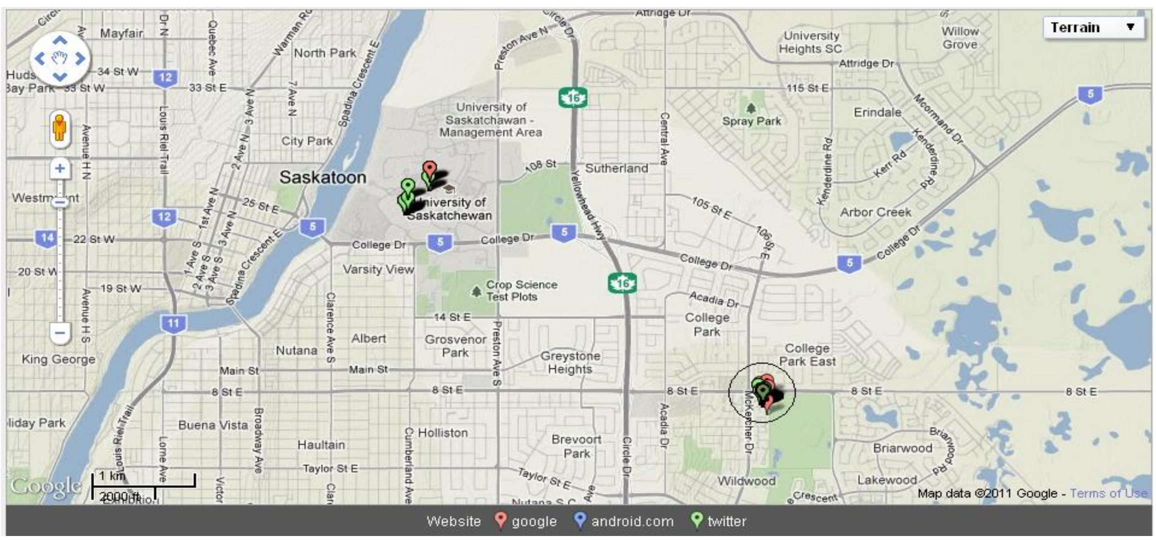


Figure 5.24: Websites Accessed On Wi-Fi Networks by Participant P12 at Different Locations

web access. They identified 10 different location areas using Wi-Fi access point information and presented probability of accessing top websites in those areas. However, this thesis considers web access from two primary locations: workplace/school and home and presents detailed analysis at the aggregate and individual user level, which is not found in the literature. Trestian *et al.* [44] only presented analysis at aggregate level and Shepard *et al.* [43] only considered individual user level dependencies. Analysis based on both individual and aggregate level data provide a better understanding suggesting the use of particular types of applications such as multimedia applications might drop substantially for all participants at specific places, and at an individual level participants prefer doing different things at different places.

5.3 Time-of-Day Dependency

This section explores the relationships between smartphone usage and times of the day. Of interest is whether there exists any patterns in which particular websites are accessed at particular times of day.

For determining time-of-day dependency the usage of facebook and news related websites was analyzed. Facebook is currently one of the most popular social networking websites. The traffic generated by the participant who accessed facebook on the most days (participant P30) is analyzed. News related applications were primarily accessed by one participant (P23), so analysis of news browsing behavior is constrained to that participant.

Figure 5.25 presents the facebook usage of participant P30. On the x-axis of the figure are dates. On the y-axis are times of the day and on the z-axis facebook traffic volume. The figure shows some time periods where a substantial amount of data was transferred, as well as consecutive periods of non-zero but low traffic. The latter are likely due to facebook not automatically logging the user out when the application is closed. The user must logout from facebook by pressing the menu button on the phone; there is no logout option in the facebook application itself. While logged in, facebook periodically sends data to the client application to keep the session alive. The figure shows that this participant tends to use facebook at particular times of day. The participant accessed facebook late at night. Substantial usage around 9 am is also observed. However, most of the facebook accesses occurred between 5 pm to 9 pm. Interestingly, no substantial use is observed from 10 am to 4 pm.

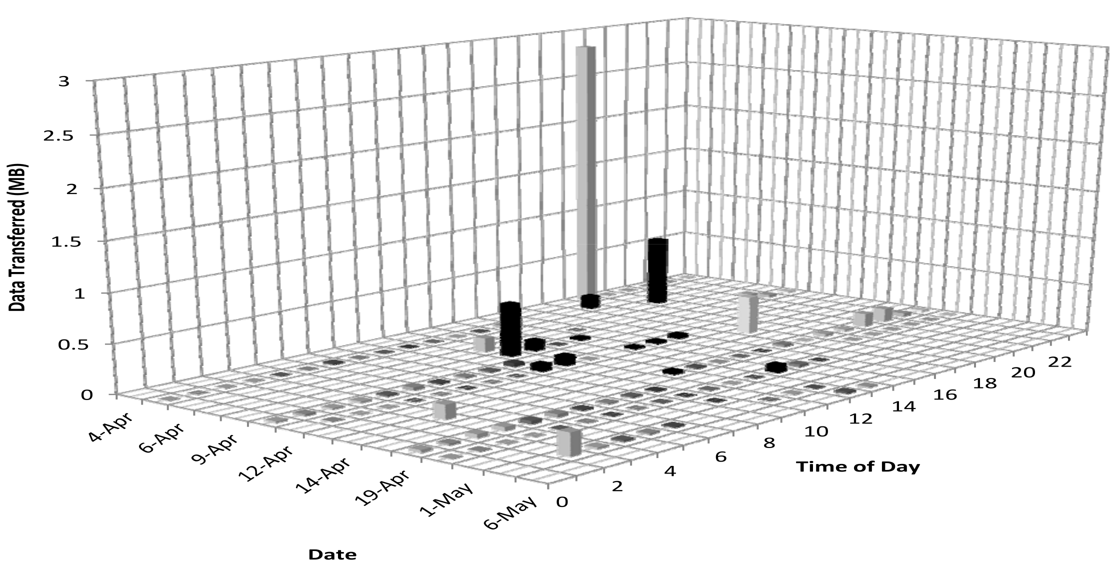


Figure 5.25: Facebook Access Pattern of Participant P30 (participant accessing facebook on the most days)

To more precisely understand the bytes transferred by participant P30 for facebook on each active session, sizes of less than 30 KB are excluded, each hour is divided into 10 slots 6 minutes each to represent active facebook. The results are presented in Figure 5.26.

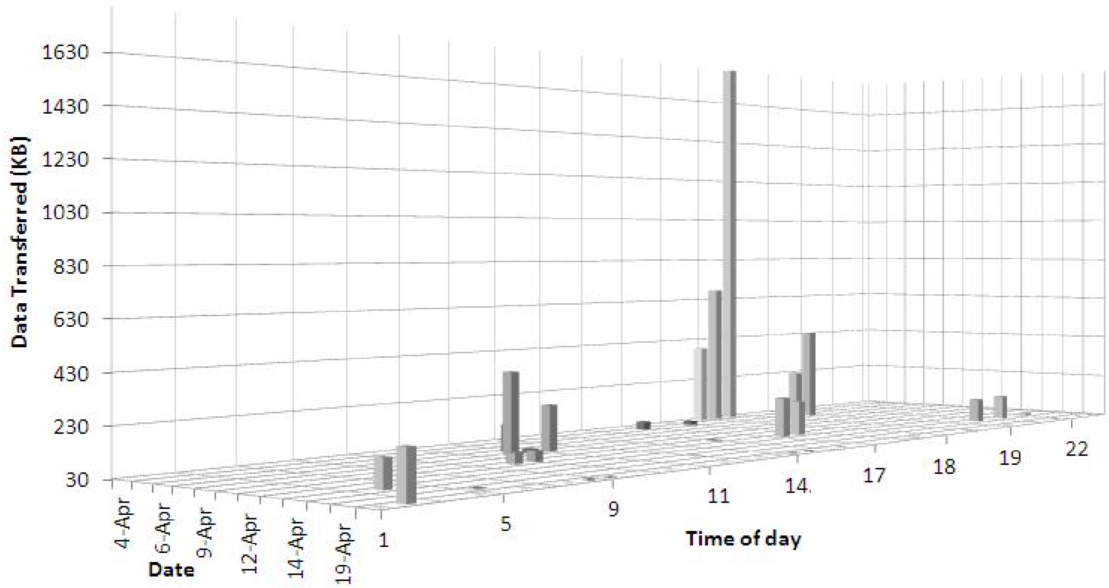


Figure 5.26: A Closer Look at Facebook Access Pattern of Participant P30

A news aggregator application was exclusively used by participant P23. The participant used the news aggregator to find news headlines and then followed stories from there. The participant's news reading patterns are analyzed to determine any relationships with the time-of-day.

Figure 5.27 presents the news reading patterns of participant P23. The x-axis of the figure represents different dates and the y-axis of the figure represents different times of the day. The black markers represent news site accesses on the university campus and the others represents news site accesses outside of the university campus. No dependency between location and news reading is observed. However, news websites were only accessed at particular times of day. On one particular date the participant read news early in the morning around 5 am. However, no access occurred on any date during the daytime until 3 pm. Substantial usage from 3 pm to 6 pm and from 9 pm to midnight is observed. This graph suggests that the participant read news during late afternoon and at night.

Figure 5.28 presents the traffic volume associated with visits to news websites for participant P23 at different times of the day. On the x-axis are different dates, the amount of data transferred is on the z-axis and the y-axis gives the times of the day. This graph suggests that when this participant accessed news websites, access tended to be extensive.

The above discussions suggest that there exist some dependencies between time-of-day and

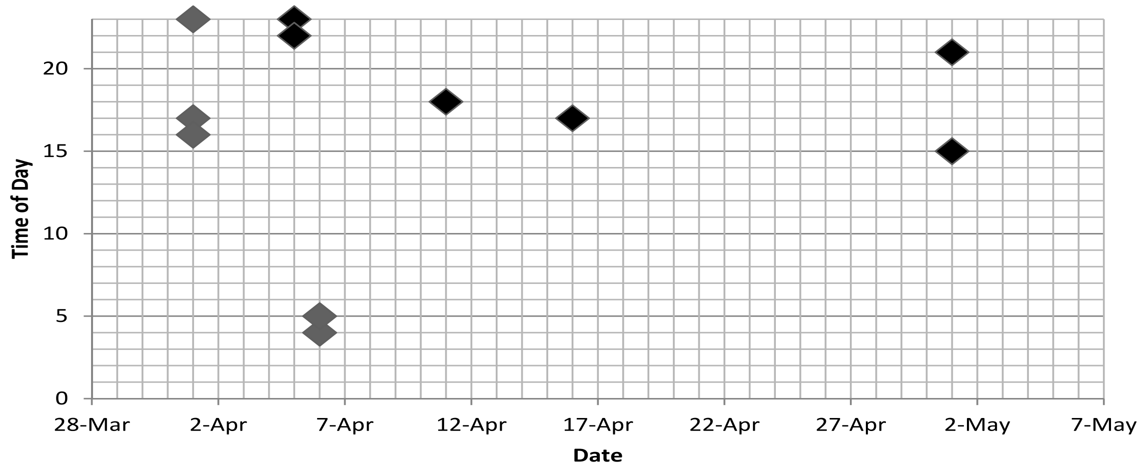


Figure 5.27: News Reading Pattern of Participant P23 (top visitor to news websites)

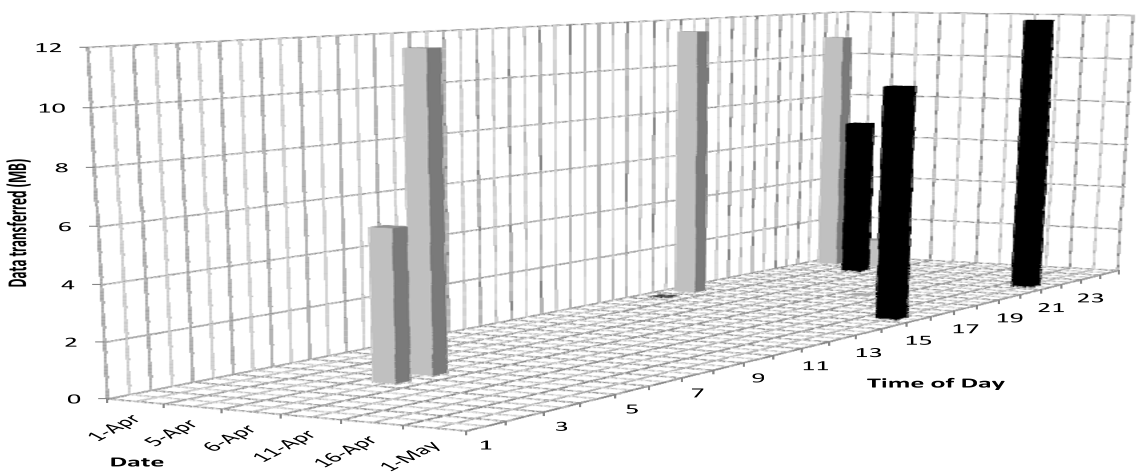


Figure 5.28: A Closer Look at News Reading Pattern of Participant P23

browsing behavior for individual heavy network users. Participants tended to access particular websites at particular times of day in general for the two participants whose behavior was investigated in detail.

5.4 Movement Dependency

This section investigates the relationship between the movement of participants and their Internet usage, attempting to determine if a particular usage pattern exists when moving and how it might differ from browsing patterns when stationary. Only cellular traffic was analyzed to find out whether participants prefer accessing some sets of websites while they are moving, because of interest in broad movement patterns such as commuting, which are dominated by cellular connectivity. Applications used by moving and stationary participants is investigated in [44, 47]. However, both of the studies used base station information to identify movements which fails to provide precise movement information. Individual user behavior is also not investigated in [44, 47]. This study used more reliable GPS information to identify participants' movement and perform individual user level analysis. GPS data contained speed information of the participants which was used to identify whether participants were moving or stationary. During data collection, 10 GPS records were collected in each duty cycle if a satellite fix could be obtained. The average speed is calculated for each duty cycle and whenever average speed was more than 5 m/s (18 km/h) for a participant, that participant was considered moving.

Table 5.2 presents the percentage of time participants were in a different movement state and how they accessed the cellular network to generate traffic in these different states. 2.79% of the records suggested participants were moving and in that state they made 5.82% of the requests and transferred 9.55% of the bytes. Most of the time participants were stationary (97.21%) and generated more traffic.

Table 5.2: Participants at Different State

State	**Time Spent (%)**	**Request Generated (%)**	**Byte Transferred (%)**
Moving	2.79	5.82	9.55
Stationary	97.21	94.18	90.45

Figure 5.29 breaks down cellular traffic by content type for moving and stationary usage. The figure shows that application content contributed greatest amount of traffic for both cases. Image content contributed more traffic for moving usage.

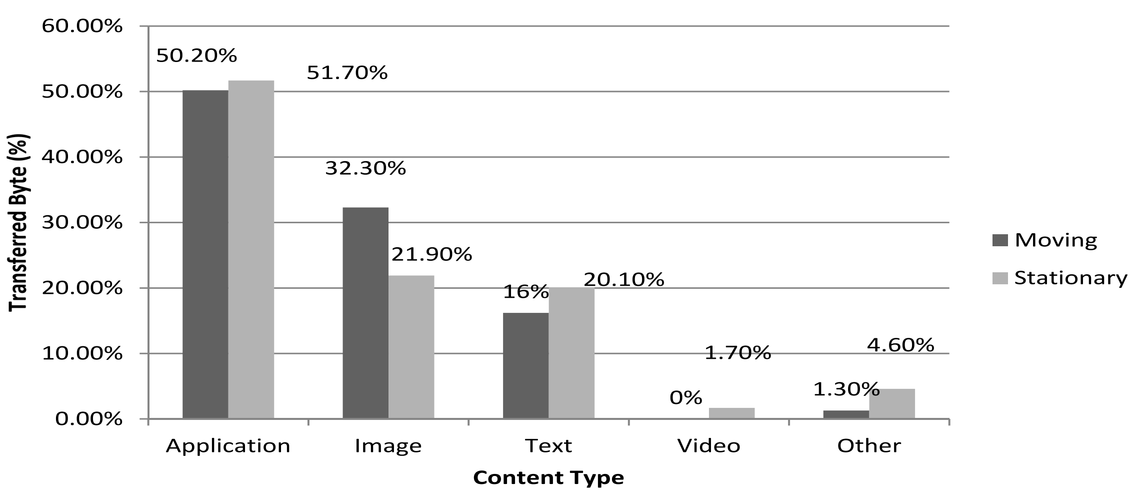


Figure 5.29: Content Type for Cellular Traffic

The top websites accessed by moving and stationary participants over the cellular network is presented in Figure 5.30. The figure reflects the dominance of stationary usage. Most of the bytes for popular hosts were generated when the participants were stationary. However, substantial usage of mobile maps, espncricinfo, facebook and google is observed for moving participants.

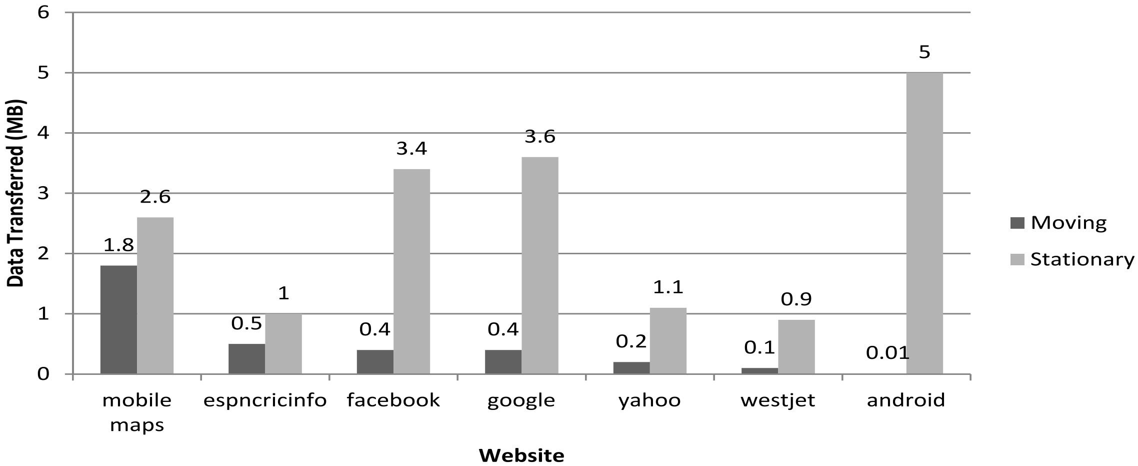


Figure 5.30: Top Website for Cellular Traffic

The website categories accessed by moving and stationary participants is presented in Figure 5.31. As with Figure 5.30, this figure reflects the dominance of stationary usage. More data was transferred when participants were stationary for all categories. However, substantial usage of browsing, map and social networking websites for moving participants is observed. One interesting observation is that the usage of map applications by moving participants almost matches the usage by stationary participants. Therefore, caching decisions based on movement might employ different policies for different applications.

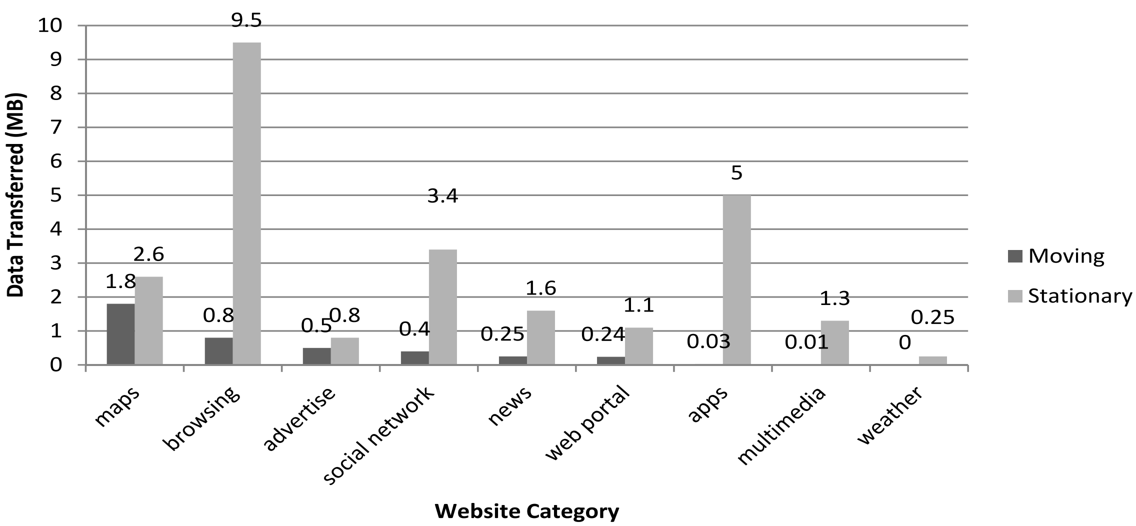


Figure 5.31: Website Categories for Cellular Traffic

The relative popularities of categories of websites accessed by moving and stationary participants are presented in Figure 5.32. The figure suggests that maps were the most popular categories among moving participants, accounting for 44.60% of the bytes transferred. In contrast, a relatively lower popularity of map applications (10.17%) is found for stationary participants. Apps were predominantly downloaded by stationary participants. App downloads only accounted for 1% of the bytes transferred for moving participants.

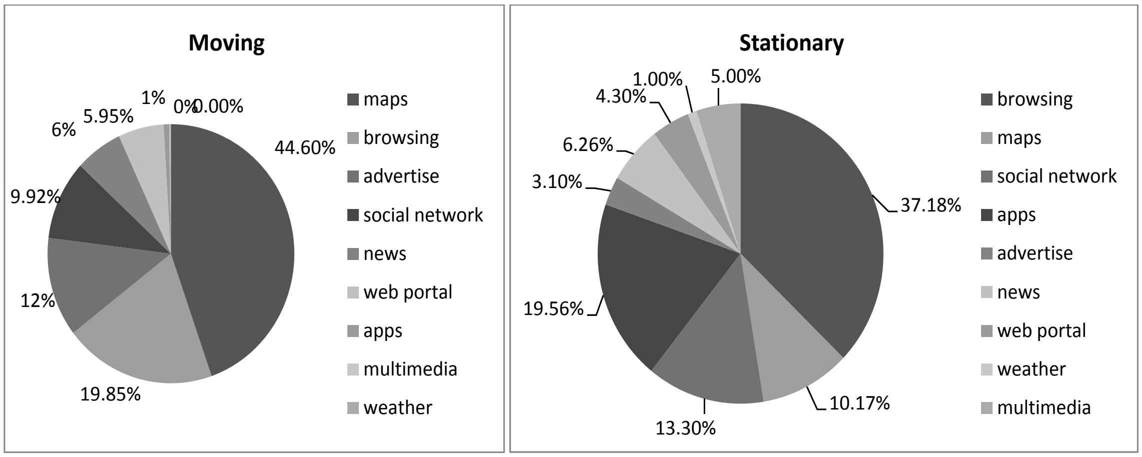


Figure 5.32: Relative Popularity (response byte %) of Website Categories for Cellular Traffic

The usage patterns of the top users of cellular network will be investigated to determine how their usage patterns changed while they were moving or stationary.

Figure 5.33 shows the websites accessed by participant P30 (Cellular Rank=1) on the cellular network. As expected, for most of the websites more data was transferred when the participant was stationary. However, greater usage of mobile maps occurred while the participant was moving.

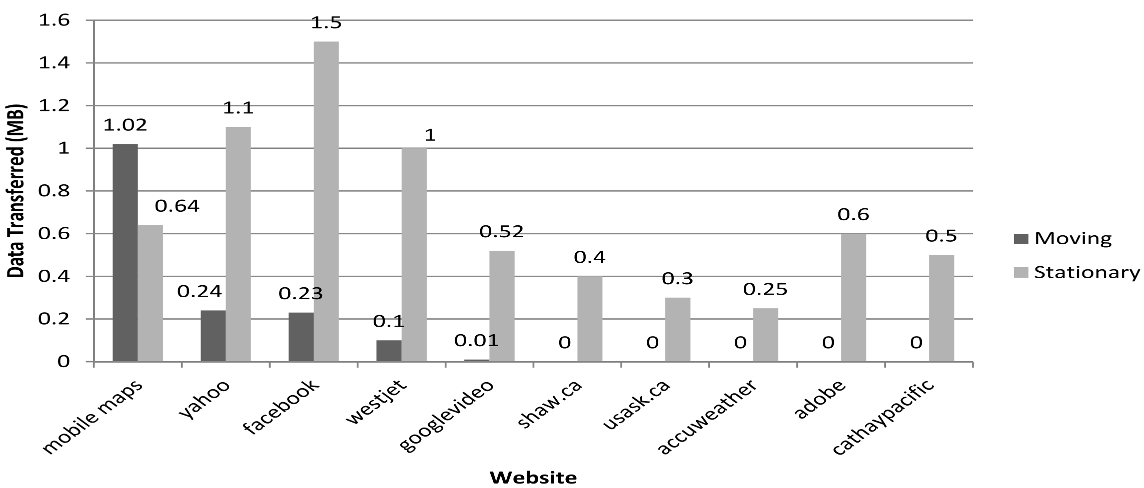


Figure 5.33: Websites Accessed by Participant P30 on the Cellular Network (Cellular Rank=1)

Figure 5.34 shows the relative popularities of websites accessed by participant P30. The participant predominantly accessed mobile maps (62.71% of response bytes) while moving. A much lower proportion of the traffic was due to usage of mobile maps when stationary.

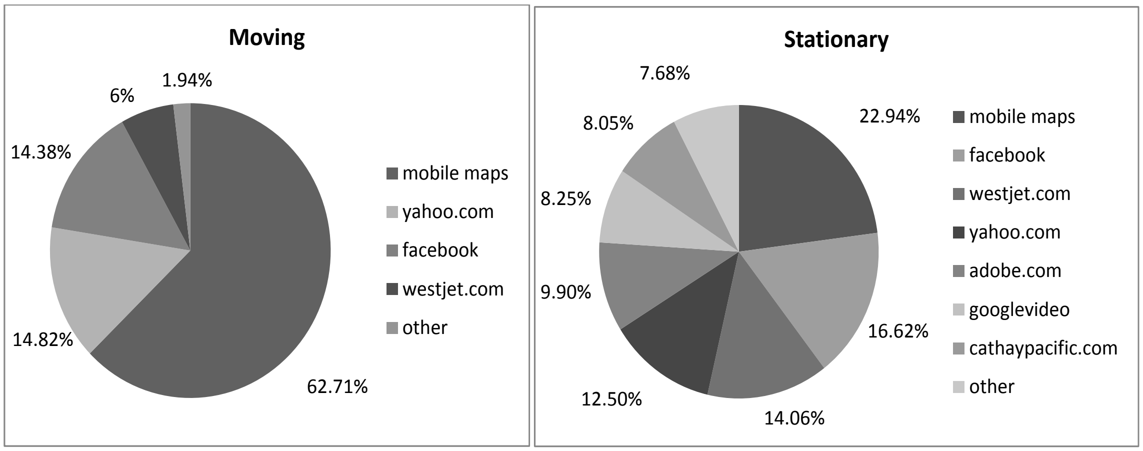


Figure 5.34: Relative Popularity of Websites (response byte %) for Participant P30 on the Cellular Network

The websites accessed by participant P25 (Cellular Rank=2) on the cellular network are presented in Figure 5.35. Again, stationary usage is dominant. However, usage of espncricinfo and google occurred while the participant was moving. Earlier analysis performed in Section 4.3.2 suggests that the participant might have followed some cricket matches and checked live score during 2011 cricket world cup.

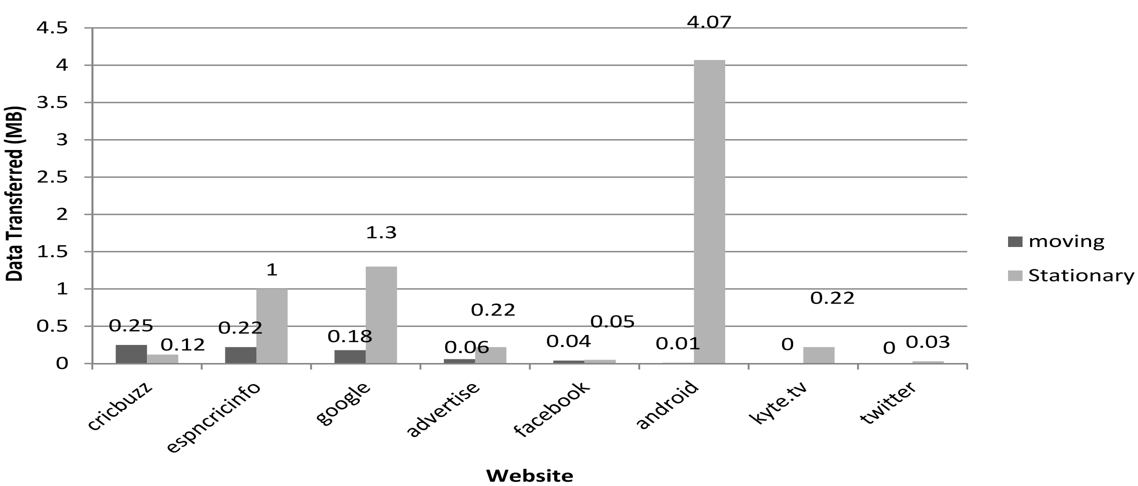


Figure 5.35: Websites Accessed by Participant P25 on the Cellular Network (Cellular Rank=2)

Figure 5.36 presents the relative popularities of the websites accessed by participant P25. The figure shows substantial differences between moving and stationary behavior. Stationary use is dominated by android.

The traffic generated by participant P7 (Cellular Rank=3) over cellular network is presented in Figure 5.37. As before, stationary usage dominates. However, significant usage of map applications occurred while the participant was moving.

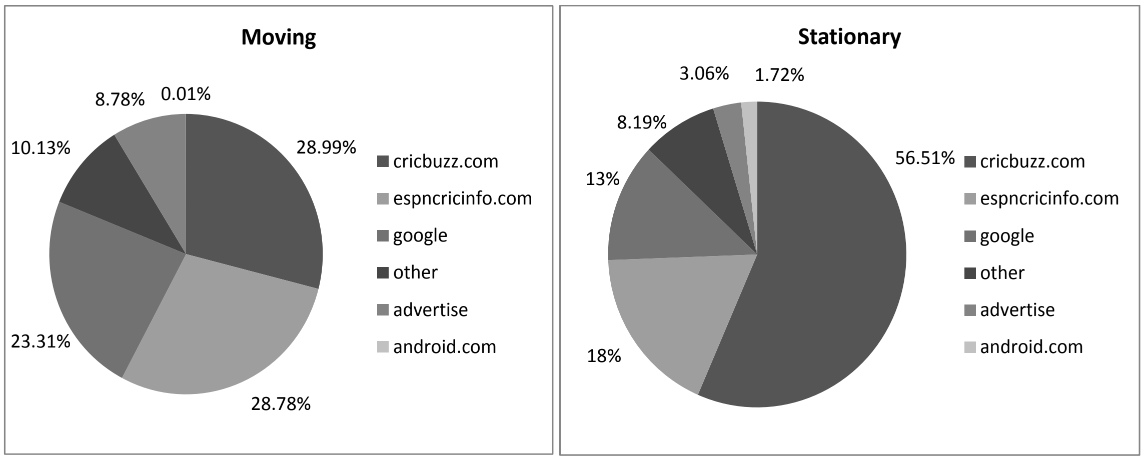


Figure 5.36: Relative Popularity of Websites (response byte %) for Participant P25 on the Cellular Network

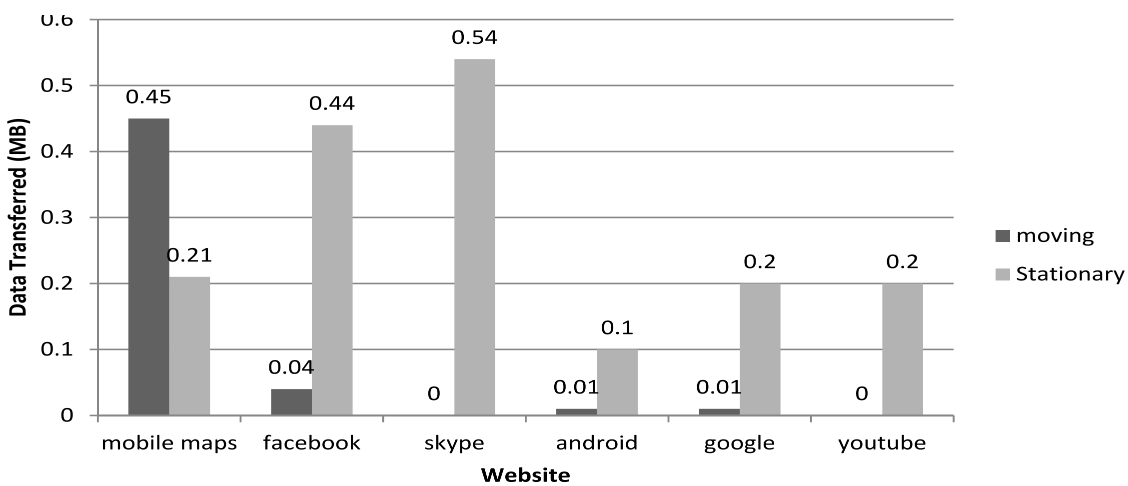


Figure 5.37: Websites Accessed by Participant P7 on the Cellular Network (Cellular Rank=3)

The relative popularities of the websites accessed by participant P7 (Cellular Rank=3) is presented in Figure 5.38. This figure shows that 87.39% of the response bytes were contributed by mobile maps while the participant was moving.

The remaining top ranked users of the cellular network generated most of their bytes while they were stationary. Very little usage occurred while the participants were moving. Therefore, their usage is not analyzed in detail here.

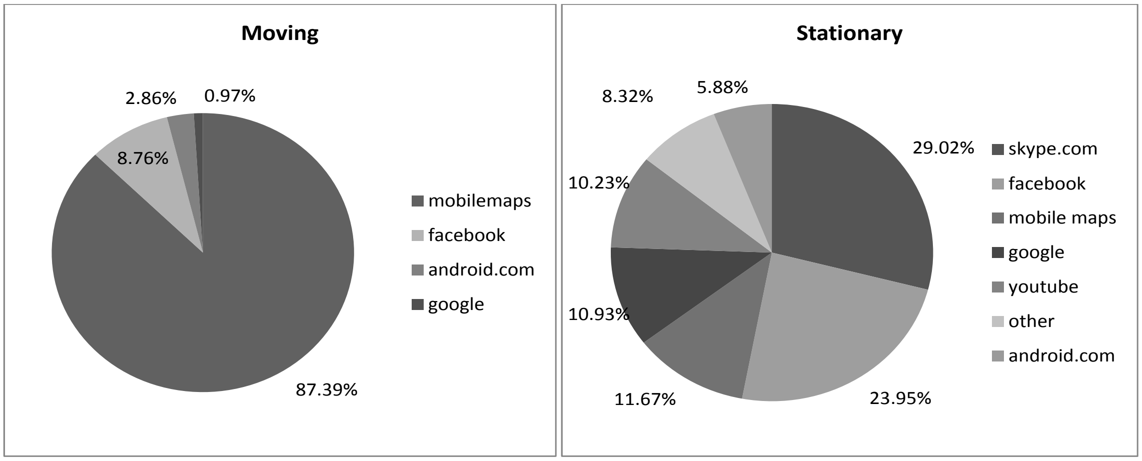


Figure 5.38: Relative Popularity of Websites (response byte %) for Participant P7 on the Cellular Network

The collected data unsurprisingly shows that most of the traffic was generated by participants while they were stationary and that participants uniformly used mobile maps more while moving. The greater traffic volumes generated while stationary reflect greater time durations spent stationary rather than moving. More use of map applications over the cellular network suggests that caching policies for map applications might help providing better services to moving users.

5.5 Proximity Dependency

This section explores relationships between smartphone usage and proximity to other participants. Of interest is whether smartphone use changes when participants are close to other people, compared to when they are alone. Similar analyses have not been performed in previous work.

Bluetooth data is used to identify proximity to another person. During data collection, data obtained from any Bluetooth device was recorded. However, for this section only Bluetooth discoverable cell phones including both smartphones and non-smartphones are considered. If a participant's smartphone detected other cell phones (not necessarily study phones) with Bluetooth discoverable and an RSSI value of 45 dBm or more, then that participant is considered not alone. RSSI value of 45 dBm refers to couple of meters (generally 5-6 meters) distance. Of course, it is still possible that the participant was with people without a Bluetooth phone or with such a phone but without their Bluetooth discoverable. 'One ph avail.' is used to denote that the participant's smartphone detected a single cell phone with an RSSI value of 45 dBm or more, 'two or more ph avail.' denotes that the participant's smartphone detected two or more cell phone with an RSSI value of 45 dBm or more, and 'alone' indicates that no such device was detected.

Table 5.3 shows the percentage of time participants in different states and their Wi-Fi network utilization. Most of the time participants were alone and generated more bytes. However, substantial usage is also observed in other scenarios.

Table 5.3: Participants at Different Scenario

State	Time Spent (%)	Request Generated (%)	Byte Transferred (%)
Alone	76.21	75.95	67.53
One ph. available	14.78	10.98	15.34
Two or more ph. available	9.01	13.07	17.13

Figure 5.39 presents the content type for Wi-Fi traffic, with usage broken down into 'alone', 'one ph Avail' and 'two or more ph Avail' categories. Substantial differences between the categories are not observed, although video type traffic was mostly generated when participants were alone or when a single other Bluetooth phone was discoverable.

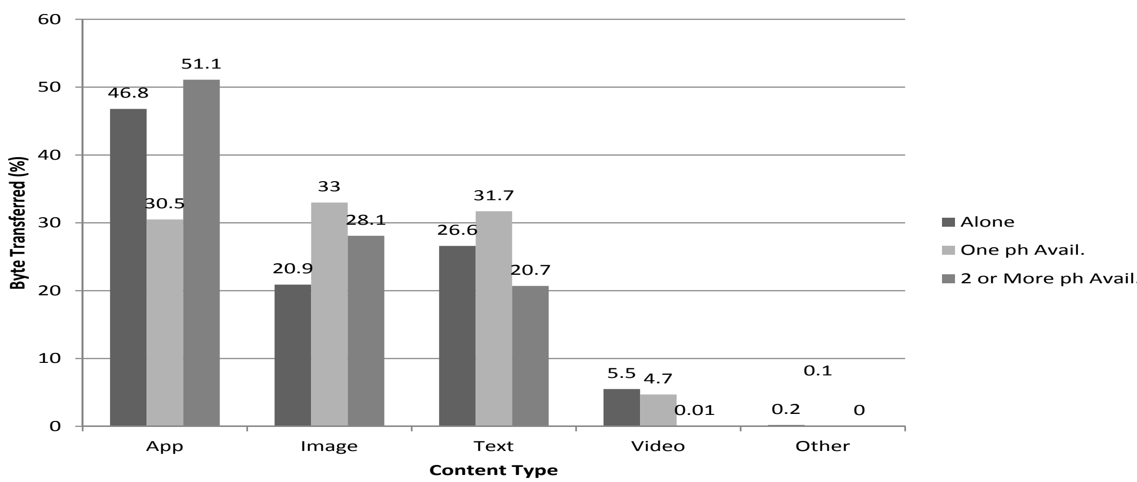


Figure 5.39: Content Type for Wi-Fi Users

Figure 5.40 presents the top websites for Wi-Fi traffic categorized by availability of Bluetooth discoverable devices in proximity. The figure shows participants consumed more data when they were alone. However, significant usage is still visible when one and two or more phone was discoverable via Bluetooth. The use of yahoo messenger was predominantly observed when the participant was alone. A substantial amount of traffic was transferred for android when two or more phone was discoverable via Bluetooth. However, a closer look at the data suggested that one of the participants generated that traffic over a single session. One interesting thing to notice is the use of facebook, mobile maps and google. Although participants generated 75.95% of their requests while alone, substantial amount of traffic was transferred for those three websites.

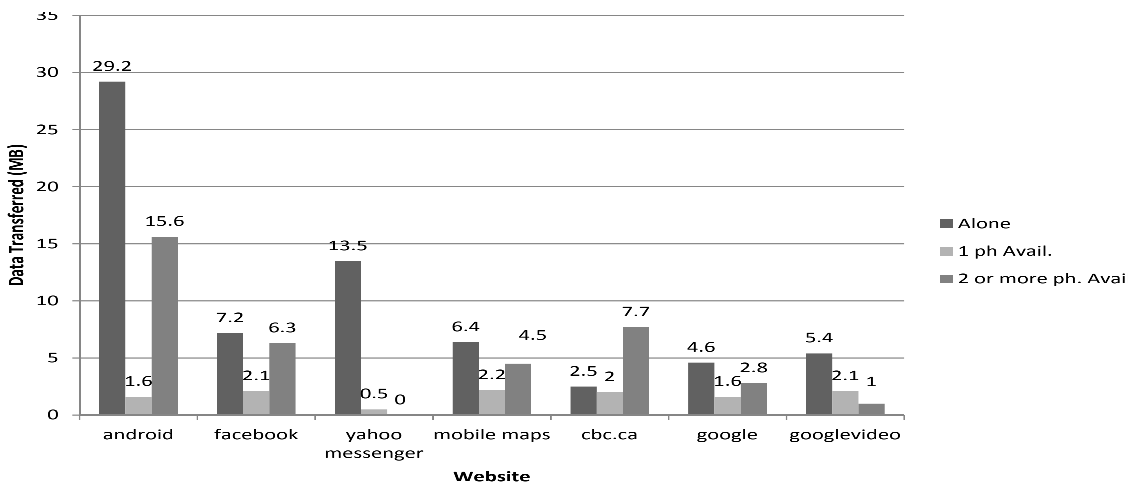


Figure 5.40: Top Websites for Wi-Fi Users, Usage Categorized by Proximity

Website categories for Wi-Fi traffic categorized by proximity to other people are presented in Figure 5.41. Higher usage for all website categories occurred when participants were alone. Substantial usage also occurred when one or more phone was available. However, participants preferred to access messenger application, multimedia type content and weather-related websites when they were alone. Although participants spent 76.21% of their time alone, browsing contributed almost equally for all three cases. Substantial use of social networking and map applications were also observed.

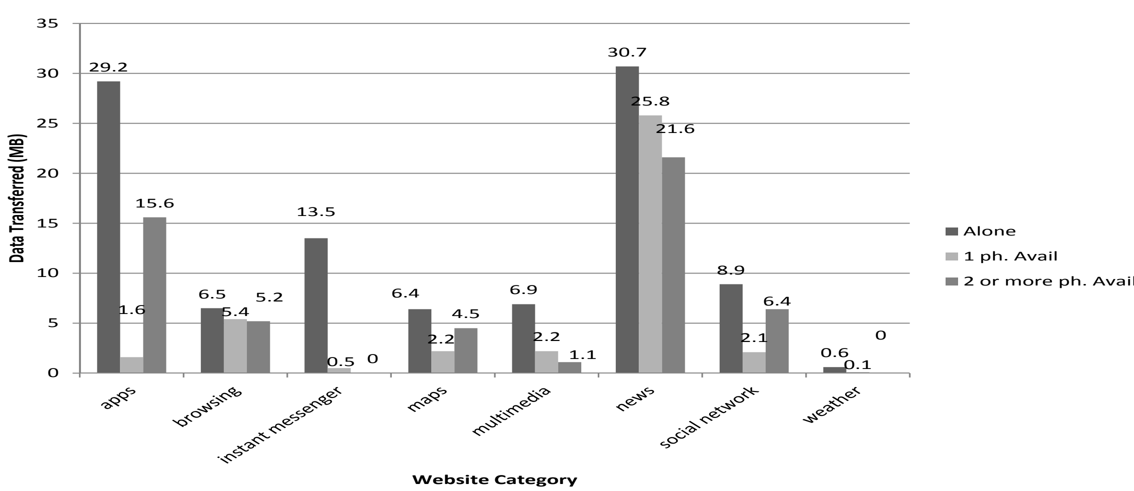


Figure 5.41: Website Categories for Wi-Fi Users, Usage Categorized by Proximity

From the above discussion it is evident that participants generated more traffic when no phone was close enough with Bluetooth discoverable. Participants also preferred accessing multimedia websites and using messenger applications when they were alone. Individual usage patterns of the top users of Wi-Fi networks are analyzed next to determine if usage patterns differ at the individual user level, when participants are in proximity to other people.

Figure 5.42 shows the websites accessed by participant P24, the top Wi-Fi user, when alone, close to one Bluetooth discoverable phone and two or more Bluetooth discoverable phones. From the figure it is evident that the participant primarily used messenger applications when alone and no phone was available with Bluetooth discoverable.

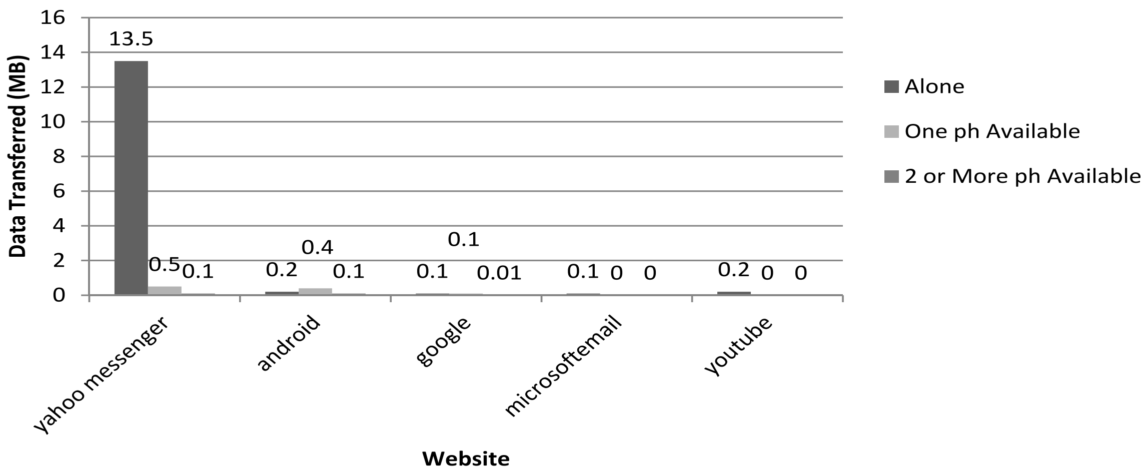


Figure 5.42: Websites Accessed by Participant P24 (Wi-Fi Rank=1), Usage Categorized by Proximity

Websites accessed by participant P23 (Wi-Fi Rank=2) in different situations are presented in Figure 5.43. The usage pattern for the participant when alone is similar to that when in proximity to other people. The Participant accessed many news websites in all cases. However, the participant only downloaded applications when alone.

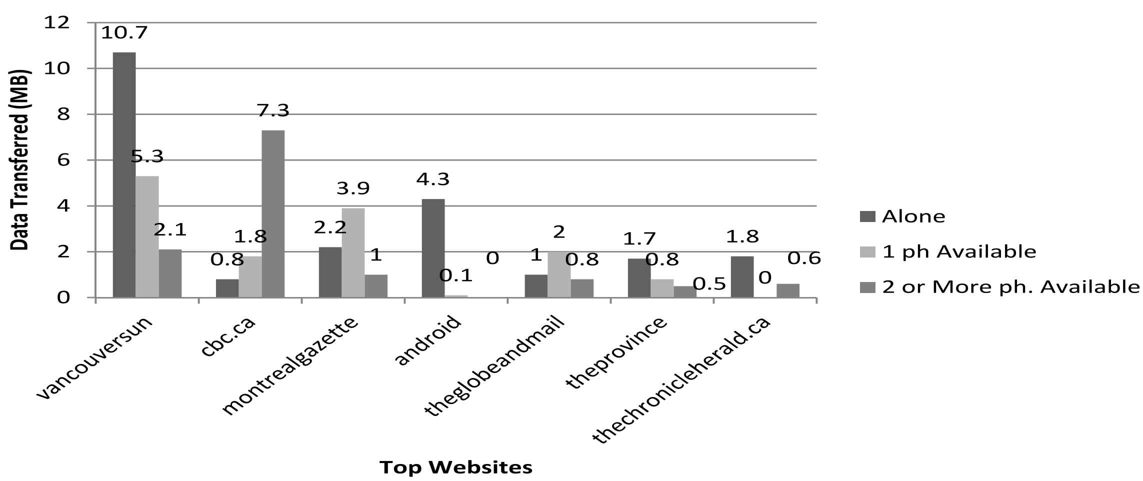


Figure 5.43: Websites Accessed by Participant P23 (Wi-Fi Rank=2), Usage Categorized by Proximity

Figure 5.44 shows the usage pattern of participants P31 (Wi-Fi Rank=3). The participant generated more traffic when alone. Most of the browsing activities were performed when the participant was not within the proximity of other Bluetooth enabled phones.

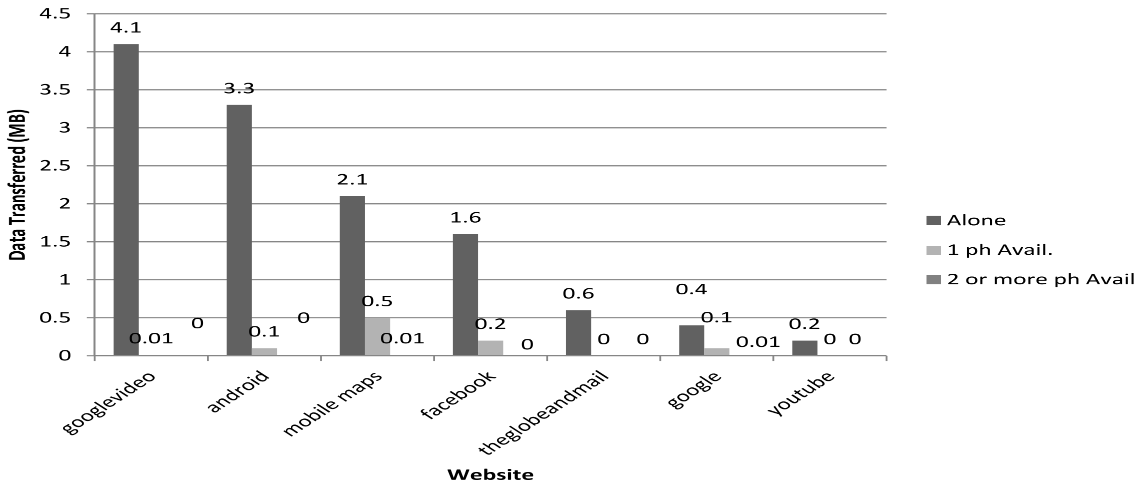


Figure 5.44: Websites Accessed by Participant P31 (Wi-Fi Rank=3), Usage Categorized by Proximity

The remaining top Wi-Fi users have similar usage patterns. They did not generate substantial amounts of traffic when they were in proximity to other people. Rather they accessed websites and generated more traffic when they were alone. Analyses presented in Section 5.2 shows that participant P30 (Wi-Fi rank=4) and participant P12 (Wi-Fi rank=5) transferred most of their bytes from home and no Bluetooth discoverable devices were available at that time which resulted in such usage patterns for those two participants.

The above discussions suggest that participants generated more traffic while alone. However, substantial use of facebook, mobile maps and google is observed for all three cases. Most of the top users did not generate much traffic while they were with other people. However, the usage patterns of participant P23 did not change when the participant was with other people. Therefore, it is observed that smartphone usage changed significantly for some participants when they were in proximity to others and for some participants it did not change much.

This chapter focused on relationships between smartphone usage and different participant contexts. Network and location dependencies at the individual level have been observed. The collected data suggests presence of time-of-day dependent access patterns for particular website access. Relationships between movement and smartphone usage is also investigated in this chapter. The collected data suggests that participants generated more traffic while stationary. Greater use of map applications is also observed for moving participants. The collected data also suggests that participants generated more traffic while alone. Some participants changed their usage patterns in the presence of other people while others did not.

Chapter 6

Summary and Conclusion

A considerable amount of Internet traffic is generated by smartphone users [31]. A few recent studies have tried to provide insight into basic characteristics of smartphone traffic. However, there are still many unanswered questions in this area. This study explores the basic characteristics of smartphone traffic and analyzes traffic dependencies at the individual user level to obtain a broader understanding of smartphone usage patterns. The relationship between different user context and smartphone usage on both individual and aggregate levels was explored.

6.1 Thesis Summary

Little is known about characteristics of smartphone traffic. Recent studies have explored this area to obtain a preliminary understanding of smartphone traffic. However, applications suitable for capturing smartphone traffic for research work are not publicly available. An application capable of capturing Internet traffic generated over Wi-Fi and cellular networks on the Android platform has been developed for this study. A user study was conducted with 39 participants for 37 days to capture smartphone traffic. Analyses were performed to understand the characteristics of this traffic.

The amount of traffic generated at different times of the day and different days of the week, along with protocols used and content types generated has been explored. Application usage was also analyzed to gather insight about the applications people use on their smartphones. Popularity analysis has been performed to obtain a better understanding of popular websites and how they are accessed over different networks. The traffic generated by the top users over different networks has been explored to understand the way participants use those networks. Strong heterogeneity has been found even in this small group of participants. Participants had different usage patterns. Some participants preferred to use messenger applications while some others preferred news or social networking applications, and many were only light users of these smartphones.

Internet connectivity is available through Wi-Fi and cellular networks in smartphones. Utilization of both Wi-Fi and cellular networks and characteristics of the traffic generated over different networks have been explored in this work. The collected data demonstrated that download-intensive applications are mostly used over Wi-Fi networks. Individual usage patterns also change when participants used different networks. Participants tend to access different sets of websites when they are connected through different networks.

Extensive analyses have been conducted to capture the dependency of usage patterns on different user contexts. Analyses have been performed to understand the dependency on location of individual and aggregated smartphone use. The collected data suggests participants tend to access multimedia content more when they are at home and less when they are at work or school. Location-dependent individual application usage patterns have also been identified.

Time-of-day dependency on access patterns was studied to understand whether participants have temporally stable access patterns for specific applications and websites. Data collected from the user study confirmed the existence of time-of-day dependent access patterns for some participants. The top facebook user seemed to prefer accessing facebook at particular times of day. Similar time-of-day dependent patterns were found for the participant who accessed the news websites.

The relationships between smartphone usage and movement is studied. Participants generated more traffic when they were stationary. Significant usage of the mobile maps has also been found for moving participants.

The relationship between smartphone use and proximity to other people has also been explored in this study. Heterogeneity has been found with respect to the impact on smartphone usage of proximity to other people. Some participants seemed to change their browsing activities around people and some did not.

6.2 Thesis Contribution

The main contributions of this thesis are as follows.

- One of the major contributions of this thesis is the methodology used for data collection and analysis. An application for capturing network traffic on Android phones has been developed. Using this platform it is possible to capture network traffic from both Wi-Fi and cellular networks and upload data to a server over a TCP connection. The user study was conducted providing a limited amount of data plans to participants, unlike other studies where unlimited voice, SMS and data plans were provided. Such techniques increases the possibility of reflecting real life usage patterns because unlimited plans might motivate the users to use communication applications and perform download-intensive activities more over the cellular network than usual time. This study also used data, collected from different sensors by another project to determine user context. Such methodologies can be employed in some other settings to obtain a better understanding of the impact of user context on smartphone use.

- Basic characteristics of smartphone traffic, and application and website popularity analysis over different networks have been studied to understand the way people use smartphones. Heterogeneity has been found among the participants in different circumstances. Participants with different usage patterns have been identified. Differences between the usage of Wi-Fi and cellular networks have been explored at both an individual user and aggregate level. A preference for performing download-intensive activities over Wi-Fi networks has been confirmed, and other preferences for performing specific tasks over certain networks have been discovered.

- Dependencies of browsing activities on location and time-of-day have been identified. Caching and prefetching decisions could potentially use such results to enhance performance and reduce delays. Relationships between the movement of participants, proximity to other users and smartphone usage patterns have been studied. Analyses presented could facilitate research related to sharing data among collocated people in ad-hoc networks, for example, one of the top participants who followed live scores of a cricket match could share that data among collocated people.

- Analyses have been performed to help understand the use of apps, maps and multimedia applications in different contexts. Apps were primarily downloaded over Wi-Fi networks. Maps were predominantly used by moving participants over the cellular network. However, substantial use of map applications over Wi-Fi networks was also identified. Multimedia applications were primarily accessed from home. All of these suggests that participants used different applications under different circumstances and most of them had a pattern which could be utilized for designing caching techniques.

6.3 Discussion

Network traffic can be collected from different locations using different methods. Researchers have collected data from user surveys, network backbones and from devices to analyze traffic characteristics. Data collection from backbone networks enables studying larger populations and provides a view from the aggregate level. However, on-device data collection facilitates the analysis of individual user characteristics.

The generalizability of the data set is an issue of concern when performing user studies. It is always difficult to determine if a user population would reflect significant characteristics of larger populations. In this research, participants for the user study had age, race and gender differences. However, the user study was conducted with a small group of participants and most of them were graduate students in the U. of S. Computer Science Department who stayed till midnight at the university and generated smartphone traffic.

Experimental settings sometimes lead to unrealistic behaviors. Smartphone-related studies have been performed in different settings and have provided different incentives such as unlimited voice, SMS and data plans. There is the possibility that participants in these studies used more resources than they would have done in real life scenarios and therefore generated measurement data which does not reflect usage in the real world. In this study, participants were only given limited resources and possibly provided more realistic data. However, technical problems that occurred during the user study (see Section 3.4) might have had some impact on the collected dataset.

Confidentiality and security are critical issues for study participants. Participants may have avoided accessing websites containing sensitive information such as banks or even checking their e-mail. Therefore, the data collected from the user study might not have included a portion of the traffic which participants would have generated otherwise.

The characteristics of the application platform can have substantial impacts on usage patterns and on the resulting smartphone traffic. Different operating systems and handsets of different qualities are available in the market. Usage patterns might differ for users of different platforms, for example iPhone users might use their phones differently than Symbian users depending on the applications available for those platforms. The quality of the handsets might also drive changes in user behavior. This study used the Android platform which is the second most popular OS as discussed in Section 1.2 and the open source Android OS provides opportunities for modifications. Therefore, the dataset used for this study is representative of a considerable portion of smartphone use.

6.4 Future Work

While capturing network traffic, the whole packet was captured, instead of just headers. Study participants did not generate a huge amount of traffic and data upload was frequent. Therefore, no smartphone storage capacity related problems occurred. However, in different settings, data collection might suffer from such problems. It is possible to capture only the headers during data collection and implementation of such flexibility will permit data collection in different settings.

Investigations could be performed with different participant groups to determine how their usage differs from that of the participant group and for this thesis. Data collection should also be performed on different platforms to determine if platform dependencies exist.

Desktop and laptop traffic could be collected to compare desktop, laptop and smartphone traffic to understand the way people use different devices to connect to the Internet. Comparisons between data collected in different years and, in different places will possibly help better understand the trends and changes in traffic characteristics.

The collected data shows the availability of both locked and unlocked Wi-Fi access points while participants were using the cellular network for Internet access. Studies could be performed to determine the possibility of using those access points for Internet connectivity to reduce load on the cellular network.

Traces collected during the experiment can be used for studying caching and prefetching policies. Design and development of an intelligent policy could potentially help to reduce response times and network traffic. Such policies could be investigated and then performance could be compared using the collected data. For example, one of the participant in this study accessed news websites at specific times of the day which could be utilized for predicting the participant's news access patterns and prefetching news headlines based on that information.

6.5 Conclusion

This thesis explores smartphone traffic characteristics and analyzes the relationship between smartphone usage and user context. A user study was conducted for 5 weeks for data collection. A novel methodology was used during data collection where participants were given limited resources and network traffic was collected along with various sensor data (Accelerometer, GPS, bluetooth and Wi-Fi). Heterogeneity is found to be the main theme of the obtained results. Strong heterogeneity has been observed among usage patterns of different participants. Heterogeneity has also been identified with respect to network, location and proximity. Location and time-of-day dependent access patterns are noticed for individual participants. Results obtained from the analysis suggests that caching/prefetching should be tailored for habits of individual users, rather than using a 'one size fits all' approach.

References

[1] Atul Adya, Paramvir Bahl, and Lili Qiu. Analyzing the browse patterns of mobile clients. In *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, IMW '01, San Francisco, California, USA, pages 189–194, November 2001.

[2] Hari Balakrishnan, Mark Stemm, Srinivasan Seshan, and Randy Katz. Analyzing stability in wide-area network performance. In *Proceedings of the 1997 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '97, Seattle, Washington, USA, pages 2–12, June 1997.

[3] Naimul Basher, Aniket Mahanti, Anirban Mahanti, Carey Williamson, and Martin Arlitt. A comparative analysis of web and peer-to-peer traffic. In *Proceeding of the 17th International Conference on World Wide Web*, WWW '08, Beijing, China, pages 287–296, April 2008.

[4] Pierre. Borgnat, Guillaume Dewaele, Kensuke Fukuda, Patrice Abry, and Kenjiro Cho. Seven years and one day: sketching the evolution of Internet traffic. In *Proceedings of the 28th IEEE Conference on Computer Communications*, INFOCOM'09, Rio De Janeiro, Brazil, pages 711 –719, April 2009.

[5] Lee Breslau, Cao Pei, Fan Li, Graham Phillips, and Scott Shenker. Web caching and zipf-like distributions: evidence and implications. In *Proceedings of the 18th Annual Joint Conference of the IEEE Computer and Communications Societies*, INFOCOM '99, New York, New York, USA, pages 126–134, March 1999.

[6] Tom Callahan, Mark Allman, and Vern Paxson. A longitudinal view of HTTP traffic. In *Proceedings of the 11th International Conference on Passive and Active Measurement*, PAM '10, Zurich, Switzerland, pages 222–231, April 2010.

[7] Selchuk Candan, Wen-Syan Li, Qiong Luo, Wang-Pin Hsiung, and Divyakant Agrawal. Enabling dynamic content caching for database-driven web sites. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, SIGMOD '01, Santa Barbara, California, USA, pages 532–543, May 2001.

[8] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, San Diego, California, USA, pages 1–14, October 2007.

[9] Yanqing Cui and Virpi Roto. How people use the web on mobile devices. In *Proceeding of the 17th International Conference on World Wide Web*, WWW '08, Beijing, China, pages 905–914, April 2008.

[10] Jeffrey Erman, Alexandre Gerber, Mohammad Hajiaghayi, Dan Pei, Subhabrata Sen, and Oliver Spatscheck. To cache or not to cache: The 3g case. *IEEE Internet Computing*, 15(2):27–34, March 2011.

[11] Jeffrey Erman, Alexandre Gerber, Mohammad T. Hajiaghayi, Dan Pei, and Oliver Spatscheck. Network-aware forward caching. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, Madrid, Spain, pages 291–300, April 2009.

[12] Jeffrey Erman, Alexandre Gerber, and Subhabrata Sen. HTTP in the home: it is not just about pc's. *SIGCOMM Comput. Commun. Rev.*, 41(1):90–95, January 2011.

[13] Hossein Falaki, Dimitrios Lymberopoulos, Ratul Mahajan, Srikanth Kandula, and Deborah Estrin. A first look at traffic on smartphones. In *Proceedings of the 10th Annual Conference on Internet Measurement*, IMC '10, Melbourne, Australia, pages 281–287, November 2010.

[14] Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, and Deborah Estrin. Diversity in smartphone usage. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, San Francisco, California, USA, pages 179–194, June 2010.

[15] Aaron Gember, Ashok Anand, and Aditya Akella. A comparative study of handheld and non-handheld traffic in campus Wi-Fi networks. In *Proceedings of the 12th Passive and Active Measurement Conference*, PAM '11, Atlanta, Georgia, USA, pages 173–183, March 2011.

[16] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. Youtube traffic characterization: a view from the edge. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, San Diego, California, USA, pages 15–28, October 2007.

[17] Lei Guo, Enhua Tan, Songqing Chen, Zhen Xiao, and Xiaodong Zhang. The stretched exponential distribution of Internet media access patterns. In *Proceedings of the 27th ACM Symposium on Principles of Distributed Computing*, PODC '08, Toronto, Ontario, Canada, pages 283–294, August 2008.

[18] Martin Halvey, Mark Keane, and Barry Smyth. Predicting navigation patterns on the mobile-internet using time of the week. In *Special Interest Tracks and Posters of the 14th International Conference on World Wide Web*, WWW '05, Chiba, Japan, pages 958–959, May 2005.

[19] Martin Halvey, Mark Keane, and Barry Smyth. Time based patterns in mobile-internet surfing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, Montreal, Quebec, Canada, pages 31–34, April 2006.

[20] Martin Halvey, Mark T. Keane, and Barry Smyth. Mobile web surfing is the same as web surfing. *Commun. ACM*, 49(3):76–81, March 2006.

[21] Mohammad Hashemian, Kevin Stanley, Dylan Knowles, Jonathan Calver, and Nathaniel Osgood. Human network data collection in the wild: the epidemiological utility of micro-contact and location data. In *Proceedings of the International Health Informatics Symposium*, IHI '12, Miami, Florida, USA, page to appear, January 2012.

[22] Sunghwan Ihm and Vivek S. Pai. Towards understanding modern web traffic. In *Proceedings of the International Measurement Conference*, IMC '11, Berlin, Germany, pages 295–308, November 2011.

[23] Utku Irmak and Torsten Suel. Hierarchical substring caching for efficient content distribution to low-bandwidth clients. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, Chiba, Japan, pages 43–53, May 2005.

[24] Kenichi Ishii. Internet use via mobile phone in japan. *Telecommunications Policy*, 28(1):43–58, January 2004.

[25] Shaun Kane, Amy Karlson, Brian Meyers, Paul Johns, Andy Jacobs, and Greg Smith. Exploring cross-device web use on pcs and mobile devices. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I*, INTERACT '09, Uppsala, Sweden, pages 722–735, August 2009.

[26] Amy Karlson, Shamsi Iqbal, Brian Meyers, Gonzalo Ramos, Kathy Lee, and John Tang. Mobile taskflow in context: a screenshot study of smartphone usage. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, CHI '10, Atlanta, Georgia, USA, pages 2009–2018, April 2010.

[27] Antero Kivi. Diffusion and usage of mobile browsing in finland 2005-2006. In *Proceedings of 4th International CICT Conference*, CICT '07, Copenhagen, Denmark, pages 85–101, November 2007.

[28] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. *Wireless Networking*, 11(1-2):115–133, January 2005.

[29] Thomas Kunz, Thomas Barry, James Black, and Hugh Mahoney. WAP traffic: description and comparison to www traffic. In *Proceedings of the 3rd ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, MSWIM '00, Boston, Massachusetts, USA, pages 11–19, August 2000.

[30] Wei Li, Andrew Moore, and Marco Canini. Classifying HTTP traffic in the new age. In *Proceedings of the ACM SIGCOMM conference Poster Session*, SIGCOMM '08, Seattle, USA, August 2008.

[31] Gregor Maier, Fabian Schneider, and Anja Feldmann. A first look at mobile hand-held device traffic. In *Proceedings of the 11th International Conference on Passive and Active Measurement*, PAM'10, Zurich, Switzerland, pages 161–170, April 2010.

[32] Zink Michael, Suh Kyoungwon, Gu Yu, and Kurose Jim. Characteristics of youtube network traffic at a campus network - measurements, models, and implications. *Computer Networks*, 53(4):501 – 514, March 2009.

[33] Earl Oliver. The challenges in large-scale smartphone user studies. In *Proceedings of the 2nd ACM International Workshop on Hot Topics in Planet-Scale Measurement*, HotPlanet '10, San Francisco, California, USA, pages 5:1–5:5, June 2010.

[34] Bong-Won Park and Kun Chang Lee. The effect of users characteristics and experiential factors on the compulsive usage of the smartphone. *Communications in Computer and Information Science*, 151(1):438–446, June 2011.

[35] Vern Paxson. Bro: a system for detecting network intruders in real-time. *Computer Networks*, 31(23-24):2435 – 2463, December 1999.

[36] Rastin Pries, Florian Wamser, Dirk Staehle, Klaus Heck, and Phuoc Tran-Gia. On traffic characteristics of a broadband wireless Internet access. In *Proceedings of the 5th Euro-NGI Conference on Next Generation Internet Networks*, NGI '09, Aveiro, Portugal, pages 184–190, July 2009.

[37] Sungyul Ryoo, Hyojin Kim, and Soohyun Jeon. An empirical study on mobile web browsing service adoption in korea. In *Proceedings of 8th International Conference on Mobile Business*, ICMB '09, Dalian, China, pages 324–328, June 2009.

[38] Grischa Schmiedl, Markus Seidl, and Klaus Temper. Mobile phone web browsing: a study on usage and usability of the mobile web. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, Bonn, Germany, pages 70:1–70:2, September 2009.

[39] John Schneidawind. Big blue unveiling. USA Today, page 2B, November 23 1992.

[40] Fabian Schneider, Sachin Agarwal, Tansu Alpcan, and Anja Feldmann. The new web: characterizing ajax traffic. In *Proceedings of the 9th International Conference on Passive and Active Network Measurement*, PAM '08,Cleveland, USA, pages 31–40, April 2008.

[41] Popular Science. Best of what's new. page 60, December 1999.

[42] Zubair Shafiq, Lusheng Ji, Alex Liu, and Jia Wang. Characterizing and modeling Internet traffic dynamics of cellular devices. In *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '11, San Jose, California, USA, pages 305–316, June 2011.

[43] Clayton Shepard, Ahmad Rahmati, Chad Tossell, Lin Zhong, and Phillip Kortum. Livelab: measuring wireless networks and smartphone users in the field. *SIGMETRICS Performance Evaluation Rev.*, 38(3):15–20, January 2011.

[44] Ionut Trestian, Supranamaya Ranjan, Aleksandar Kuzmanovic, and Antonio Nucci. Measuring serendipity: connecting people, locations and interests in a mobile 3g network. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, IMC '09, Chicago, Illinois, USA, pages 267–279, November 2009.

[45] Hannu Verkasalo. Analysis of smartphone user behavior. In *Proceedings of 9th International Conference on Mobile Business/Global Mobility Roundtable*, ICMB/GMR '10, Athens, Greece, pages 258–263, June 2010.

[46] Hannu Verkasalo, Carolina Lapez-Nicola, Francisco Molina-Castillo, and Harry Bouwman. Analysis of users and non-users of smartphone applications. *Telematics and Informatics*, 27(3):pages 242 – 255, August 2010.

[47] Qiang Xu, Jeffrey Erman, Alexandre Gerber, Morley Mao, Jeffery Pang, and Shobha Venkataraman. Identify diverse usage behaviors of smartphone apps. In *Proceedings of the 11th Annual Conference on Internet Measurement*, IMC '11, Berlin, Germany, pages 329–344, November 2011.

[48] Chun Yuan, Yu Chen, and Zheng Zhang. Evaluation of edge caching/offloading for dynamic content delivery. In *Proceedings of the 12th International Conference on World Wide Web*, WWW '03, Budapest, Hungary, pages 461–471, 2003.