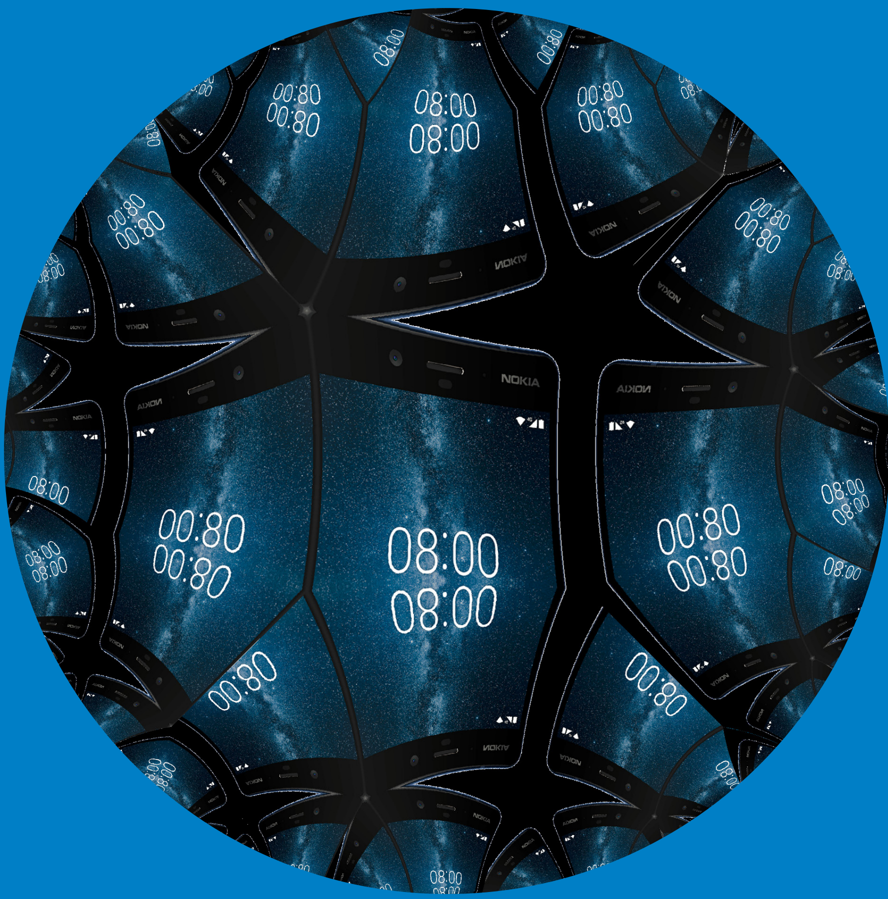
Department of Communications and Networking

User Experience and Usage of Mobile Services in Novel Contexts

Benjamin Finley



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 173/2017

User Experience and Usage of Mobile Services in Novel Contexts

Benjamin Finley

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall AS1 of the school on October 13th, 2017 at 12:00.

Aalto University
School of Electrical Engineering
Department of Communications and Networking
Network Economics

Supervising professor
Professor Heikki Hämmäinen, Aalto University, Finland

Thesis advisor
Dr. Kalevi Kilkki, Aalto University, Finland

Preliminary examiners
Professor Jussi Kangasharju, University of Helsinki, Finland
Professor Tobias Hoßfeld, University of Duisburg-Essen, Germany

Opponent
Dr. Sebastian Egger-Lampl, Austrian Institute of Technology, Austria

Aalto University publication series
DOCTORAL DISSERTATIONS 173/2017

ISBN 978-952-60-7609-6 (printed)
ISBN 978-952-60-7608-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-7608-9

Unigrafia Oy
Helsinki 2017

Finland

NORDIC ECOLABEL
441 697
Printed matter

Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Benjamin Finley

Name of the doctoral dissertation
User Experience and Usage of Mobile Services in Novel Contexts

Publisher School of Electrical Engineering

Unit Department of Communications and Networking

Series Aalto University publication series DOCTORAL DISSERTATIONS 173/2017

Field of research Network Economics

Manuscript submitted 20 June 2017 **Date of the defence** 13 October 2017

Permission to publish granted (date) 16 August 2017 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

Mobile devices and services are tightly woven into many aspects of modern society and culture. These devices and services fulfill diverse needs such as communication, information, and entertainment with an anywhere anytime philosophy. Despite this importance and ubiquity, mobile research has not thoroughly examined mobile usage and experience in all mobile contexts. Specifically, several understudied mobile contexts are novel and/or present practical difficulties for studying.

This thesis explores three such contexts: multiple mobile devices (multidevice), multiple mobile networks, and long-term mobile services. Towards this goal, the thesis research leverages two unique empirical datasets collected from the USA and Finland and several modern simulation and statistical analysis methods including generalized logistic regression modeling and agent-based modeling.

Given the diverse nature of the contexts, the research results are also diverse and primarily include context-specific insights. For example, for users with both smartphones and tablets, multidevice usage already represents a significant fraction (~50%) of all device usage time, therefore, illustrating the prevalence of such usage. Additionally, multidevice usage sessions show significant diversity thus emphasizing the need for multidevice apps to be highly personalized.

In the multiple mobile network context, the user quality of experience benefits of fast switching between multiple networks are significant in most situations, whereas the benefits of using multiple networks simultaneously are more limited due to inefficient resource allocation. Finally, in the context of long-term mobile services, users satisfaction with such services at a given point in time depends partly on complex temporal phenomena spanning the service time frame such as the peak-end effect. Though even accounting for these phenomena only explains a fraction of the variation in user satisfaction thus motivating future research.

Overall, the results present initial data points for these contexts that hopefully spur additional research and allow for more robust theory creation. Furthermore, several of the individual results are interesting for mobile ecosystem players such as consumers, national regulators, mobile network operators, and device vendors.

Keywords Mobile QoE, novel mobile contexts, multiple mobile devices, multiple mobile networks, long term mobile service

ISBN (printed) 978-952-60-7609-6 **ISBN (pdf)** 978-952-60-7608-9

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2017

Pages 174 **urn** http://urn.fi/URN:ISBN:978-952-60-7608-9

Preface

This thesis is the culmination of work in the research group of Prof. Heikki Hämmäinen at Aalto University. Therefore, I would foremost like to thank Heikki for the opportunity to perform the thesis in his team. I would also especially like to thank my instructor Dr. Kalevi Kilkki for significant guidance along the way and for initially encouraging me to apply to the Communications Ecosystem program. In addition, I appreciate the noteworthy effort of my pre-examiners Prof. Jussi Kangasharju and Prof. Tobias Hoßfeld and my distinguished opponent Dr. Sebastian Egger-Lampl.

Next, I thank my co-authors, specifically Dr. Tapio Soikkeli, Dr. Arturo Basaure, Dr. Kalevi Kilkki, Prof. Heikki Hämmäinen, Prof. Jukka Manner, Prof. Antti Oulasvirta, and Eren Boz. Also, I want to thank Dr. Colin Rose for pointing out an error in one of the publications of this thesis.

In addition, co-workers in the two main research projects that I participated in were of great help and always encouraging. Specifically, I thank Dr. Antti Riikonen, Dr. Tapio Soikkeli and the MoMIE project[1] partners (Accenture, DNA, Elisa, Liikenne- ja viestintäministeriö, Nokia, Sanoma, Tekes, and TeliaSonera) and Dr. Arturo Basaure, Dr. Kalevi Kilkki, Prof. Heikki Hämmäinen, Prof. Jukka Manner, Prof. Antti Oulasvirta, Dr. Sebastian Sonntag, Eren Boz, Joonas Lindh and the Emergent project[2] partners (Airbus, VIRVE Tuotteet ja Palvelut Oy, Goodmill, Liikenne- ja viestintäministeriö, YLE, Tekes). The Aalto ELEC Doctoral School also provided funding for my research, and I am grateful.

Also, I want to especially thank Dr. Timo Smura, Dr. Tapio Soikkeli, Dr. Matti Aksela, and Dr. Hannu Verkasalo from Verto Analytics[3] for providing research data and guidance over many meetings.

A good research team and atmosphere can make the thesis process much more enjoyable, and I can say that my research team was truly great. My so far

[1] http://momie.comnet.aalto.fi/

[2] http://emergent.comnet.aalto.fi/

[3] http://www.vertoanalytics.com/

unmentioned team members (including current and former) that deserve thanks include Dr. Mike Katsigiannis, Dr. Tapio Levä, Dr. Henna Suomi, Dr. Nancy Zhang, Dr. Thomas Casey, Dr. Juuso Karikoski, Prof. Juuso Töyli, Alexandr Vesselkov, Jaume Benseny, Levent Kartal, Jaspreet Walia, and Pekka Kekolahti.

My professors at Milwaukee School of Engineering including Prof. Christopher Taylor and Prof. Darrin Rothe also merit thanks and provided me with a solid software foundation that has been valuable every step of the way.

My friends that gave me so much support and whom I owe a debt of gratitude include Lur Eguiluz, Enrico Roverato, Caroline Moinel, Nastassja Favale, Corinne Isler, Dr. Kha Nguyen, and Johanna, Maria, and Elina Ritala.

Finally, I want to thank my family including Elaine, Dr. Mike M., Mike F., Joan, Robert, Reid, Sean, my father Patrick, and my mother Anita, to whom I dedicate this thesis.

September 12, 2017,

Benjamin J. Finley

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I B. Finley, K. Kilkki. Exploring Empirical Rank-Frequency Distributions Longitudinally through a Simple Stochastic Process. *PLoS ONE*, 9(4): e94920, April 2014.

II B. Finley, T. Soikkeli, K. Kilkki. Mobile Application Usage Concentration in a Multidevice World. In *Proceedings of the 13th International Joint Conference on e-Business and Telecommunications*, Lisbon, Portugal, pp. 40-51, July 2016.

III B. Finley, T. Soikkeli. Multidevice mobile sessions: A first look. *Pervasive and Mobile Computing*, 39C, pp. 267–283, August 2017.

IV B. Finley, E. Boz, K. Kilkki, J. Manner, A. Oulasvirta, H. Hämmäinen. Does network quality matter? A field study of mobile user satisfaction. *Pervasive and Mobile Computing*, 39C, pp. 80–99, August 2017.

V B. Finley, A. Basaure. Benefits of Mobile End User Network Switching and Multihoming. Submitted to *Computer Communications*, April 2017.

Author's Contribution

Publication I: "Exploring Empirical Rank-Frequency Distributions Longitudinally through a Simple Stochastic Process"

Kilkki conceived and designed the experiments. Finley performed the experiments and Finley and Kilkki analyzed the data. Finley wrote the paper.

Publication II: "Mobile Application Usage Concentration in a Multidevice World"

Finley conceived the idea, performed the analysis, and wrote the paper. Soikkeli and Kilkki provided feedback.

Publication III: "Multidevice mobile sessions: A first look"

Finley and Soikkeli jointly conceived the idea, performed the analysis, and wrote the paper.

Publication IV: "Does network quality matter? A field study of mobile user satisfaction"

All authors jointly conceived the idea and designed the questionnaire. Boz implemented the questionnaire code. Finley extracted the data, performed the analysis, and wrote the paper.

Publication V: "Benefits of Mobile End User Network Switching and Multihoming"

Finley and Basaure jointly conceived the idea, performed the simulations, and wrote the paper.

List of Acronyms and Abbreviations

3G	Third Generation
3GPP	3rd Generation Partnership Project (GSM, 3G, LTE, etc.)
ABM	Agent Based Modeling
AIC	Akaike Information Criterion
App	Application
CDF	Cumulative Distribution Function
DGBD	Generalized Discrete Beta Distribution
E-SIM	Embedded SIM
ECDF	Empirical Cumulative Distribution Function
ER	Equal Resource
Freq	Frequency
FT	Fracturing
GFLOPS	Billion (Giga) Floating Point Operations per Second
GSM	Global System for Mobile Communications
HCI	Human Computer Interaction
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
LTE	Long Term Evolution
MH	Multihoming
MLE	Maximum Likelihood Estimation
MPTCP	Multipath Transmission Control Protocol
MOS	Mean Opinion Score
NS	Network Switching
OS	Operating System
PC	Personal Computer
PDF	Probability Density Function
Perf	Performance
PPI	Pixels per Inch

PPO	Partial Proportional Odds
Pub	Publication
Pubs	Publications
QoE	Quality of Experience
QoS	Quality of Service
RMSE	Root Mean Squared Error
RP	Research Problem
RQ	Research Question
RTT	Round Trip Time
SIM	Subscriber Identity Module
SINR	Signal to Interference plus Noise Ratio
TCP	Transmission Control Protocol
TW	Time Window
US	United States
VMNO	Virtual Mobile Network Operator
Wifi	Marketing term for technology based on IEEE 802.11 standards
WiMax	Worldwide Interoperability for Microwave Access

List of Symbols

ψ	A stochastic process function
t	An index (often a discrete time step)
ζ	A random variable of FT process
x	A single value in the FT process state
L_t	FT process state at time step t
L_1	Initial FT process state
C_{FT}	A constant of the FT process
F	Final outcome of the FT process
p_1	A random variate of ζ at $t = 1$ for FT process
γ	Fitting parameter in FT process PDF
$U_{[0,1]}$	Standard uniform random variable
M	The number of categories of an ordinal dependent variable
X_i	The independent variable i in ordinal logistic regression
α_j	Constant in ordinal logistic regression model
β_j	Independent variable j coefficient in ordinal logistic regression
$\hat{p}$	Approximate permutation test p-value
m_p	The number of permutations for approximate permutation test
δ	Full permutation test error rate level
R^2	Correlation coefficient

1. Introduction

This chapter provides an introduction to the thesis including motivation, research problem and questions, research approach, and structure.

1.1 Background and motivation

Mobile devices[1] and services[2] are quickly becoming a pervasive and ubiquitous part of modern daily life across industrialized nations. For example, 77% of American adults own a smartphone, and these users spend on average 1.5 hours a day using their smartphone [63, 87]. Similarly, 51% of American adults own a tablet, and these users spend on average 40 minutes a day using their tablet [63, 87].

Given this ubiquity, many different societal stakeholders are interested in understanding this mobile phenomenon. Naturally, players within the mobile device and service ecosystem (such as device vendors, platform creators, and application (app) developers) are interested as they have a direct business stake. However, even non-related businesses, non-profit organizations, and governments now want to understand how to engage with customers or constituents through novel mobile services[3].

Therefore, significant research into mobile service user experience[4] and usage has been recently published[5]. These studies range in topic from optimizing mobile touchscreen keyboards [58] to applying mobile apps and devices to aid with

[1] The definition of a mobile device is found in Section 1.3.

[2] The definition of a mobile service is found in Section 1.3.

[3] For example, as of October 2015, 52% of large companies (>2500 employees) already implemented a mobile app strategy and 90% planned to increase mobile app investment [70].

[4] For this thesis, user experience refers to the concepts of the quality of experience (QoE) domain rather than concepts of the related human-computer interface focused user experience domain. Refer to [94] for a discussion of the difference between these domains.

[5] See Section 3, Tables 2 and 3 in [76], [53], and journals and conferences such as *MobileHCI* (`http://mobilehci.acm.org`)

health monitoring [75]. As expected, such mobile studies originate from many different research communities such as wireless networking, behavioral science, and human-computer interaction (HCI). Thus, the field can be considered both large and multidisciplinary.

However, despite this large and multidisciplinary approach, several novel mobile contexts have not been thoroughly studied, particularly from a user experience and usage viewpoint. These include user usage of mobile services in the context of multiple mobile devices[6], user experience of mobile services in a long-term context, and user experience of mobile services in the context of multiple mobile networks. The reason for the lack of significant study in each of these contexts is likely partly related to novelty and partly related to practical difficulties. Several difficulties inherent in studying the user experience and usage of these contexts are briefly detailed.

In the case of multiple mobile devices, the collection of usage data across users' devices is difficult in practice for several reasons. For example, the collection of detailed device usage data often requires device based rather than network or server based usage monitoring[7]. However, many device based monitoring apps only support Android[8] or only support a subset of functionality on iOS[9] thus limiting the extent of monitoring or forcing complex workarounds. Additionally, users must be sufficiently incentivized to install the monitoring apps on each of their devices. Given that monitoring multiple devices raises privacy concerns and necessitates more user effort in installing and maintaining the monitoring apps, such incentivization can be challenging and costly. For example, Verto Analytics (the provider of the data in Pubs. II and III) pays each panelist approximately 110 USD per year[10].

In the case of long-term mobile services, the collection of user data over long time periods ($\geq$ 1 year) is also difficult since many users drop out of data collection efforts over time[11] thus reducing the sample size and statistical power. Additionally, to evaluate, for example, user experience the data collection typically requires user surveys which often have low response rates thus further

[6] Multiple mobile devices is also referred to as multidevice.

[7] See Sections 1.3 and 4.1.1 for definitions and further discussion of device, network, and server based methods.

[8] For example, Device Analyzer [92] and Menthal [2] only support Android.

[9] For example, Aware [26] does not support app usage monitoring on iOS.

[10] See `https://vertosmart.com/#about`

[11] The rate that users drop out is known as the attrition rate. The reported annual attrition rates of commercial Internet-based survey panels vary widely but are typically around 5-25% depending on the lifetime of the panel [69, 11, 4, 9]. Whereas for mobile device-based monitoring panels the rates are generally not publicly available; though [56] suggests that such rates could be relatively high.

reducing sample size. For example, the response rate for the survey in Pub. IV was only about 15%.

Finally, in the case of multiple mobile networks, the use of multiple mobile networks by current day users is limited thus restricting the possibility for large-scale empirical studies. This particularly applies to using multiple networks simultaneously, known in this thesis as multihoming. For example, users currently cannot use two LTE data connections simultaneously (in other words aggregate their bandwidth) since this would require multiple LTE radio stacks in the mobile device and such devices do not exist. Additionally, although simultaneous usage of an LTE data connection and a WiFi data connection is technically possible with current devices such usage is not widespread nor supported by device vendors. Thus, this necessitates small scale empirical studies or non-empirical analysis such as simulations. However, most simulation efforts related to multiple mobile networks have been technical simulations rather than user-centric simulations likely due to the more technical nature of the mobile network community (as opposed to, for example, the HCI community).

In addition to the general ubiquity of mobile services, research of these specific novel contexts is important to many stakeholders for more pointed reasons. For example, given the growing prevalence of multidevice users (see Section 2.2.1); ecosystem players interested in how users distribute their time and attention (such as mobile advertisers) must understand multiple mobile device contexts to gain a holistic user view. Whereas multiple mobile network contexts are likely to be important for mobile regulators in advocating for certain regulatory policies (such as policies for reducing network switching costs) and for future mobile customers in making purchasing decisions (whether to purchase multiple network subscriptions).

Therefore, given this overall view, this thesis aims to provide several data points that help spur further research and push forward the understanding of user experience and usage in these novel contexts.

1.2 Objectives and scope

The thesis attempts to address the broad research problem (RP) detailed below by taking up the individual research questions (RQ) 1-4 also detailed below. The thesis publications partially attend to different parts of these questions as shown in the mapping of RQ to publications in Table 1.1.

Research Problem: How to understand, characterize, and model end user experience and usage of mobile services in diverse and novel contexts?

RQ1: How do users use mobile services in the context of multiple mobile devices?

RQ2: How do users experience mobile services in a long-term ($\geq$ 1 year) context?

RQ3: How would users experience mobile services in the context of multiple mobile networks?

RQ4: What distributions, processes, and models can characterize the user usage and experience of these contexts?

Table 1.1. Relationships between publications and research questions

Pub.	Pub. Title	RQ1	RQ2	RQ3	RQ4
I	Exploring Empirical Rank-Frequency Distributions Longitudinally through a Simple Stochastic Process				✓
II	Mobile Application Usage Concentration in a Multidevice World	✓			✓
III	Multidevice mobile sessions: A first look	✓			
IV	Does network quality matter? A field study of mobile user satisfaction		✓		✓
V	Benefits of Mobile End User Network Switching and Multihoming			✓	✓

1.3 Definitions

Several important terms used in this thesis are defined below:

Mobile device: Any device that is portable and can connect to a mobile network (specifically a 3GPP or IEEE WiFi network).

Mobile service: Any service enabled by a mobile device including both basic telecommunication services (such as voice, sms, and data) and complex services such as smartphone applications that are enabled by these basic services.

Network based measuring: A method for collecting data through software running on the mobile network.

Server based measuring: A method for collecting data through software running on a remote server.

Mobile device based measuring: A method for collecting data through software running directly on the end user's mobile device (the result is **mobile device based measurements**).

The following definitions are specific types of mobile device based measuring:

Mobile device based usage monitoring: A method for collecting mobile device usage data through software running directly on the end user's mobile device.

Mobile device based user surveying: A method for collecting user data through surveys that appear on the end user's mobile device.

Mobile device based network monitoring: A method for collecting mobile network data through software running directly on the end user's mobile device.

1.4 Research approach

In the spirit of related studies [76], this thesis research can be classified or placed into a general research approach taxonomy so as to view the research in terms of the larger scientific landscape. This placement exercise also helps identify other potential approaches that could have been used in this research. Järvinen et al. [36] provides such a general taxonomy that includes many different research approaches with each approach itself encompassing many research methods. The taxonomy is structured as a tree with the leaves being the research approaches. Figure 1.1 illustrates the taxonomy.

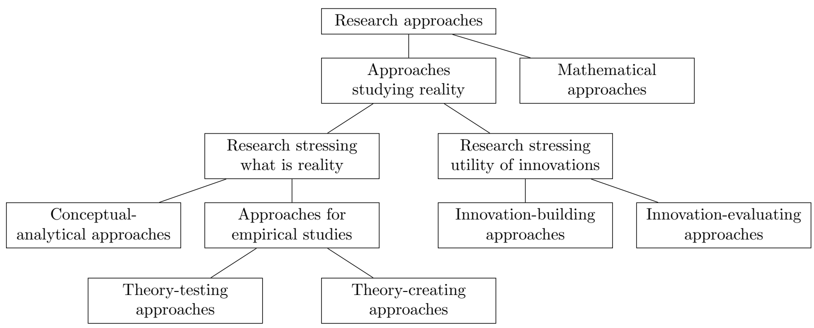


Figure 1.1. Research Approach Taxonomy [36]

In a top-down manner, all research in this thesis falls into approaches studying reality rather than general mathematical approaches that attempt to prove, for example, theorems from fundamental laws. Within these reality-based approaches, the studies in this thesis apply various approaches to differing degrees as summarized in Table 1.2.

Table 1.2. Relationships between publications and research approaches

Pub.	Primary research approach(s)
I	Innovation-building/evaluating approach
II	Theory-creating/testing approach
III	Theory-creating/testing approach
IV	Theory-creating/testing approach
V	Conceptual-analytical approach

In Pubs. II, III, and IV the theory creating approach related to finding observations from empirical data that provide a data point over which meta-analysis can allow strong theory creation. Whereas, the theory testing approach related to finding patterns that support an existing theory or allow the generalization or extrapolation of an existing theory (such as the generalization of some mobile usage theories to the multidevice context).

In Pub. I, the innovation building approach related to the proposal of a stochastic process (known as the fracturing (FT) process) that can simulate longitudinal rank-frequency distributions, whereas the evaluation approach related to comparing this process with related theoretical distributions on empirical data.

Finally, in Pub. V, the conceptual-analytical approach related to the analysis of conceptual (hypothetical) end user network switching and multihoming scenarios through detailed agent-based modeling. Thus, such analysis improves the understanding of the potential user experience given end user network switching and multihoming in the overall complex system (of autonomous agents, cellular infrastructure, and environment).

The main approach not used in this thesis work was the mathematical approach which arguably is not as well suited to, for example, the empirical user studies of this thesis. The main reason being that complex behavioral patterns rely on so many varying and difficult to quantify factors that deriving strict mathematical laws is not feasible and instead empirical (statistical) laws or models often prove more useful in practice.

1.5 Thesis structure

The remaining thesis is structured as follows. Chapter 2 provides theoretical background including brief descriptions of the three novel contexts and several additional fundamental concepts. Chapter 3 summarizes related work again divided into the sections emphasizing each novel context. Chapter 4 presents descriptions of the most prominently employed methods and empirical data. Chapter 5 reports the research results and finally, Chapter 6 discusses the implications of the results in both theoretical and practical terms.

2. Theoretical background

This chapter briefly introduces several topics related to the thesis research including mobile quality of experience, mobile service usage, and the three novel contexts described in RQ1-RQ3. The descriptions focus on concepts that help in understanding the research results.

2.1 Mobile quality of experience and service usage

2.1.1 Mobile quality of experience

Quality of experience (QoE) is broadly defined as "the degree of delight or annoyance of a person whose experiencing involves an application, service, or system" [68]. Furthermore, QoE stems "from the person's evaluation of the fulfillment of his or her expectations and needs with respect to the utility and/or enjoyment in the light of the person's context, personality and current state" [68]. Thus, QoE, as presently defined, is a rather wide term that depends on many technical, human, and environmental factors. The idea of mobile QoE simply narrows this definition to experiencing that occurs in a mobile app, service, or system context but maintains this dependence on diverse factors.

Historically, in the mobile networking community, service quality research began with studies of voice service quality in original GSM networks in the 1990s. The research attempted to understand the relationship between quantitative network parameters and perceived voice quality. However, such studies were quite technically orientated with a service provider point of view (and thus classified as quality of service (QoS) research). As the need for a more human-centric approach became apparent, the focus of research shifted and QoE[1] research became prevalent in the 2000s. Finally, with the advent of computer-like smartphones, QoE research drew the attention of the HCI community who

[1] See [86] for further details on the difference between QoS and QoE.

applied their own techniques to QoE research problems.

An important aspect of QoE research is the scale for measuring the experience of a user. The most well-known measurement scale is the Mean Opinion Score (MOS), which dates to fixed network QoS research of the 1970s [94]. The MOS scale ranges over the integers from 1-5 with each integer given a descriptive label to help the user in their evaluation. These labels are typically (from 1 to 5) Bad, Poor, Fair, Good, and Excellent. Though a single quantitative value is not always appropriate for experience evaluation, especially when evaluating many different aspects of a service, the MOS scale is widely used in practice [94]. A primary reason is that having a well-known quantitative measure allows easy modeling with MOS as a dependent variable.

Beyond MOS, recent studies have also applied qualitative methods such as semi-structured interviews or alternative quantitative metrics (beyond the basic mean and std. dev.), such as acceptability ratios or quality quantiles, to gain a more holistic[2] understanding of user experience [34].

In this thesis Pubs. IV and V both applied QoE concepts extensively and Pub. V used MOS as the main QoE metric.

2.1.2 Mobile service usage

Mobile service usage can be defined as the utilization of services enabled by a mobile network connected device. Thus, the study of mobile service usage is a research area concerned with understanding how and why users use such services. For example, researchers have studied the temporal and spatial distributions of mobile service usage [74]. Naturally, mobile service usage is fundamentally related to mobile QoE as users with poor quality will often prematurely abandon usage sessions[3]. For example, early video session abandonment correlates with many network quality indicators [73].

The mobile network community often studies mobile service usage through analysis of large datasets collected via network based measuring; for example, operator call data records or mobile network traces in the radio access or core networks [53]. Such research provides a view of high-level aggregate usage patterns (such as diurnal calling patterns). These patterns are then used to optimize overall mobile network performance.

Whereas the HCI community studies mobile service usage primarily through

[2] In the case of, for example, quality quantiles, a holistic understanding means a better understanding of the full quality distribution over users.

[3] In fact, this fundamental relation often allows researchers to link these domains in simulations and models. Refer, for example, to Section 3.5.1 of Pub. V and [50]

analysis of datasets collected via direct observation (typically in laboratory settings), user interviews, user completed usage diaries, and device based usage monitoring[4]. Such research can often provide granular usage patterns and contextual information (such as why the user uses an individual service). These usage patterns are primarily applied to improve the experience of mobile services such as mobile apps. For example, significant diversity between users in the usage of a service might be a justification for increasing the amount of personalization the service allows. Device based usage monitoring has become particularly prevalent as smartphones allow the widespread monitoring and storing of usage data directly on the device.

In this thesis Pubs. II and III both studied mobile service usage through analysis of data collected via device based usage monitoring. Though the publications have different focuses.

2.2 Novel contexts

The three novel contexts of RQ1-RQ3 can be viewed intuitively as axes of a user-centric research space. In such a space the user can be placed at the origin with three axes representing the number of mobile network connections, the number of mobile devices, and the analysis period respectively. Figure 2.1 depicts such a space. Current research is primarily limited to assumptions of one network connection, one mobile device, and a (relatively) short analysis period (on order of seconds to days). In other words, current research is limited to several planar surfaces near the origin. Contrastingly the work in this thesis lies further out on at least one axis in each case. And eventually, research should be generalizable to an arbitrary length of n on all axes.

As an additional note, a fourth axis should denote the number of users, but regarding this axis, the thesis studies are not significantly different than other studies (both are generally on the order of tens to hundreds).

The three novel contexts of RQ1-RQ3 are further briefly described in the Sections 2.2.1-2.2.3.

2.2.1 Multiple mobile devices

The mobile phone (and the subsequent successor the smartphone) has been the dominant device type in mobile networks since such networks inception. The relatively compact nature of the mobile phone form factor naturally matched

[4]Table 3.2 of Section 3.1 compares several of these data collection methods.

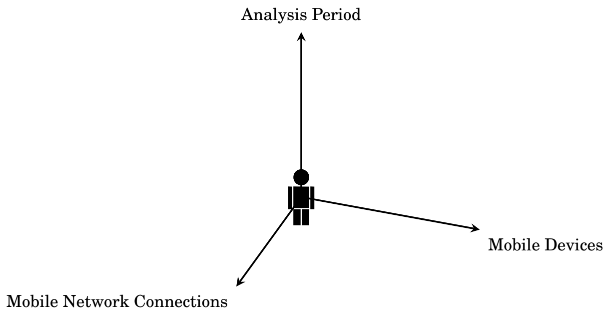


Figure 2.1. User Centric Research Space

the large focus of mobile networks on mobility. Recently, the number of mobile network connected devices has increased significantly in both the consumer and enterprise spaces. In the enterprise world, early examples include mobile payment terminals while later examples include connected smart energy meters (and other Internet of things (IoT) devices). In the consumer space, tablets have become the primary second mobile connected device, but also smart watches and connected cars are now broadly available. Statistically, the number of mobile-connected devices per capita in the US and Europe is expected to be approximately 2.8 by 2020 [14].

Even in the subset of user-centric devices, different device types have different form factors, screen sizes, processing capabilities, etc. and thus usage across device types is different. For example, Figure 2.2 illustrates the percentages of sessions of smartphones, tablets, and PCs for each hour of the day from the data from Pub. II. This heterogeneity has led to research into how users use and experience multiple mobile devices. Refer to the related work summarized in Section 3.1.

In this thesis Pubs. II and III both studied mobile service usage in multiple mobile device contexts. Pub. III studied smartphones and tablets, while Pub. II studied smartphones, tablets, and personal computers.

2.2.2 Multiple mobile networks

Multiple mobile networks have coexisted in many mobile markets since the widespread telecommunication market liberalization of the 1980s and 1990s. This liberalization essentially broke up government monopolies and allowed for multiple distinct network operators to build out their networks. In addition to multiple GSM-based networks, the popularization of alternative (non-3GPP) wireless technologies such as WiFi and WiMax also allowed for multiple mobile networks in many areas.

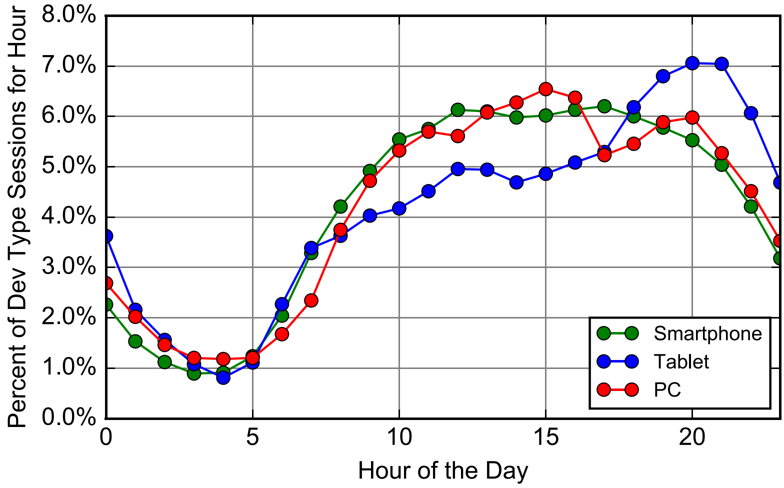


Figure 2.2. Percentages of sessions of smartphones, tablets, and PCs for each hour of the day. Adapted from data of Pub. II

The primary consumer benefit of multiple mobile network operators is increased market competition and thus lower prices. Additionally, though, the building out of distinct mobile networks implies a degree of spatial heterogeneity in network coverage and quality. Thus, a user that is unsatisfied with the network quality of their current network provider (for example due to low signal strength at their home) can potentially switch to a different provider with better quality. This heterogeneity is primarily driven by operator's desire to differentiate themselves in terms quality or coverage by densifying or building out their network beyond competitors[5].

As mentioned, theoretically, mobile network users can exploit this network heterogeneity by switching between networks at various locations to always connect to the best quality network at a given location. Additionally, in locations with multiple mobile networks, users can connect to several networks simultaneously and aggregate the bandwidth of these networks to achieve better quality (than would be possible with any single network). Refer to [31] for a broad survey of bandwidth aggregation techniques.

In this thesis, the general situation in which a user connects to and uses multiple networks simultaneously is known as multihoming. Also throughout the thesis, the examined network switching and multihoming mechanisms are end user mechanisms in that both mechanisms are assumed to occur entirely on

[5] Additionally, this heterogeneity is affected by regulatory requirements for spectrum owners that they cover a proportion of the population with their own network infrastructure. For example, Sweden requires spectrum owners to cover 30% of the population [8].

the end user device without requiring explicit MNO support (for the switching or multihoming).

In this thesis, Pub. V studied the potential user QoE benefits of both end user network switching and multihoming through detailed agent-based modeling.

2.2.3 Long-term mobile services

Long-term services are services that typically last for days, weeks, months, years, or longer[6]. Mobile network services including data and voice are examples of typical long-term mobile services. The user experience and evaluation of such services have unique dynamics in contrast to one-time or infrequent services. Specifically, such long-term services are typically multi-episodic, and the user's evaluation of the service is likely affected by the quality of each episode as well as the temporal integration processes of human cognition [30]. Unfortunately, these temporal processes are not yet well understood, though recent research suggests a few basic principles.

For example, research suggests that a retrospective evaluation that occurs months or years after an event[7] is influenced more by semantic memory[8] rather than mental snapshots of the episode [28]. Additionally, the most recent episode and the episode with the largest negative experience might have greater influence (than other episodes) on the evaluation per the peak-end effect [39]. Though again, this peak-end effect might be lessened or disappear if semantic memory becomes the dominant influence in the evaluation [28]. In the mobile service area, general recency effects have been observed for multi-episodic services[9] on the timescale of a few weeks [30, p. 57-72]. No research exists for longer timescales.

In this thesis, Pub. IV studied the mobile service satisfaction of Finnish users (which have quite long-term service as the annual mobile churn rate is only about 17%) through logistic regression modeling of mobile device based network measurements and satisfaction survey responses.

[6] In this thesis, long-term services are defined to include services lasting one year or longer as RQ2 indicates.

[7] In long-term multi-episodic services this will be the case for some but not all episodes.

[8] Semantic memory is conceptual memory that does not relate to any specific episode; this contrasts with episodic memory that describes a specific past episode [82]. Imagine the difference between the concept of dog versus a particular living dog.

[9] Refer to [30, p. 5-26] for more detailed background on retrospective evaluations and multi-episodic QoE.

3. Related research

This chapter briefly summarizes related work in areas of the three novel contexts described in RQ1-RQ3 and in the area of RQ4.

3.1 Multiple mobile device studies

A variety of multiple mobile device studies have been published over the past decade. Such studies vary significantly concerning factors including the data collection method, number of analyzed users, and analyzed devices types. Table 3.1 compares several multiple mobile devices studies on these factors and includes for reference Pubs. II and III from this thesis.

Perhaps the most fundamental of these factors is the data collection method since this method is driven by the objectives of the study which also relates to the domain of the research (i.e. HCI vs. mobile networking). To provide additional insight Table 3.2 details a variety of data collection methods for multiple mobile device studies and includes example studies and ratings of several characteristics as defined by the author. The important insight is that no single method dominates but instead the methods can complement each other. Refer to [16] for further details including practical information on several of these methods.

Several HCI studies have emphasized the problems with or enablers of multi-device usage. For example, [72] used semi-structured interviews and observation to identify the inadequacies of current methods for handling parallelism and personal data fragmentation in multidevice cases. Similarly, [48] used semi-structured interviews and observation to identify the different advantages of certain device types in certain contexts (such as the low initial access time for smartphones compared to PCs when checking messages or emails).

Other studies have quantified and characterized multidevice usage among current users. For example, Ref. [52] used user completed electronic diaries to

Table 3.1. Comparison of multiple mobile device studies

Citation	Method[a]	Users	Devices[b]	Year of Data
[59]	I	11	-[c]	2005
[22]	I	27	S,PC,D,M,O	2007
[48]	I	21	S,PC	2009
[29]	D	1611	S,T,PC,O	2012
[93]	S	39081	S,T,PC	2012
[72]	I	22	S,T,PC,E	2013
[49]	I,D	3586	S,T,PC,O	2013
[52]	D	176	S,T	2013
[51]	S	1675272	S,T,PC,O	2014
[33]	U	1588	S,T	2014
[37]	I,D	14	S,T,PC,D,M,O	2015
[20]	N	32581	S,T,PC	2015
Pub. II	U	398[d]	S,T,PC	2015
Pub. III	U	65	S,T	2015
[89]	U	307	S,W	2016

[a] Data collection method: I (Interviews and Direct Observation), D (User Completed Electronic Diaries), S (Server Based Usage Monitoring), N (Network Based Usage Monitoring), U (Device Based Usage Monitoring)

[b] Included device types: S (Smartphone, Mobile Phone), T (Tablet), PC (Laptop, Desktop), W (Smartwatch), E (E-Reader), D (Digital Camera), M (Music Player), O (Other)

[c] The full extent of included device types was not specified.

[d] This represents an upper bound based on the combination of the three subsets 4,6,7 in Table 4.3 of Section 4.4.

illustrate usage patterns of primary and secondary activities on smartphone and tablet devices. Likewise, [29] also used user diaries from a large US panel to quantify their daily multidevice interactions across a variety of devices. Refs. [51] and [93] used server based usage monitoring to examine the multidevice (known as cross-device in their terminology) searches of over 1.6M users performed on the search service of a large US Internet company. Whereas, [20] used network based usage monitoring of a university WiFi network to characterize the multidevice usage and traffic of about 30000 users. These studies are all closely related in philosophy to Pub. III but differ in data collection methodology.

Finally, closely related in data collection methodology to both Pubs. II and III, [33] used mobile device based usage monitoring to compare session properties

Table 3.2. Data collection methods of multiple mobile device studies

Data Collection Method	Method Ratings[a]		Example Studies
Interviews and direct observation	Contextual Holistic Accurate Scalable	● ● ● ● ● ● ● ● ● ● ● ● ● ○ ○ ● ○ ○ ○ ○	[22, 59, 72, 48]
User completed electronic diaries	Contextual Holistic Accurate Scalable	● ● ● ● ○ ● ● ● ● ○ ● ● ○ ○ ○ ● ● ○ ○ ○	[52, 29]
Server based usage monitoring	Contextual Holistic Accurate Scalable	● ● ○ ○ ○ ● ○ ○ ○ ○ ● ● ● ● ○ ● ● ● ● ●	[51, 93]
Network based usage monitoring	Contextual Holistic Accurate Scalable	● ● ○ ○ ○ ● ● ● ○ ○ ● ● ● ● ○ ● ● ● ● ○	[20]
Mobile device based usage monitoring	Contextual Holistic Accurate Scalable	● ● ● ○ ○ ● ● ● ● ○ ● ● ● ● ○ ● ● ● ○ ○	[33, 89], Pub. II, Pub. III
Combination of methods	Depends on individual methods		[37, 49]

[a] Contextual means the ability to capture a variety of contextual information such as location, the motivation for usage, experienced quality, etc. Holistic means the ability to capture the full range and diversity of usage (including, for example, different services and usage information). Accurate means the ability to capture accurately objective quantitative data such as usage times. Scalable means ability to cheaply and efficiently scale the method to more users.

of smartphones and tablets. However, due to limitations of the data (driven by privacy concerns), [33] was not able to link together smartphones and tablets of the same user. Therefore [33] did not study, for example, individual multidevice sessions but only aggregate device type characteristics. Similarly, [89] used device based usage monitoring to compare session properties of smartphones and smartwatchs (though again not for sessions of the same user).

3.2 Multiple mobile network studies

Published multiple mobile network studies have been primarily technical studies driven by the need for novel technologies to overcome the dominant single network connection paradigm inherent in many devices and networks. These multiple mobile network studies can be roughly classified on how the multiple networks are used (network switching vs. multihoming) and the types of the multiple networks (IEEE WiFi vs. 3GPP). Table 3.3 details this classification including example studies, also Pub. V is included for reference. Furthermore, given that Pub. V of this thesis focused on mechanisms that are controlled by the end user, the related work also focuses on such studies.

Table 3.3. Classification of multiple mobile network studies[a] by types of networks and how the networks are used

	3GPP & 3GPP	3GPP & WiFi	WiFi & WiFi
Network Switching	[44], Pub. V	[23, 55]	[40, 12]
Multihoming	[67], Pub. V	[23, 19, 55]	[40, 81]

[a] Studies that cover multiple classes appear more than once.

Significant focus has been placed on cases with a single 3GPP network and a single IEEE WiFi network as virtually all modern smartphones include both a 3GPP radio and a WiFi radio. For example, [23] addressed 3GPP and WiFi network switching and multihoming through a general optimization framework known as Delphi. The framework leverages MPTCP for multihoming. The overall system was validated on commodity smartphones in situations where both LTE and WiFi connectivity were available. Also, [55] compared the empirical network and energy performance of network switching versus multihoming over LTE and WiFi. The multihoming analysis included both MPTCP and an optimal data partitioning method. The results indicated that multihoming through MPTCP is not energy optimal. Ref. [19] performed a system level simulation of multiple multihoming users connected to both a 3GPP and WiFi network. The simulation tested several data partitioning techniques and identified a similar inefficient resource allocation problem as identified in Pub. V. This work is the most closely related, in both philosophy and method, to Pub. V.

Regarding cases with two 3GPP networks, in a fundamental work [44] proposed a network switching platform and algorithm that uses online machine learning for the network switching decisions. The VMNO overlay model of Google Fi[1] allowed the multiple mobile network subscriptions.

Finally, for cases with two IEEE networks, [40] introduced FatVAP as an IEEE WiFi driver for PCs that performs both network switching and multihoming over multiple WiFi networks. The driver performs active network probes to assess the capacity of different networks, and the data partitioning algorithm then uses the probe results for partitioning. Similarly, [12] detailed Multinet as an IEEE

[1] `https://fi.google.com/`

WiFi driver for PCs that virtualizes the network interface card and thus allows OS transparent and fast network switching between multiple WiFi networks.

3.3 Long-term mobile service studies

Very few long-term mobile service studies have been published. Though, several studies in related areas prove helpful.

Studies that use consumer surveys to assess general user satisfaction with mobile services[2] [60, 41, 83] are generally related. The surveys of these studies were generally longer and more detailed than the survey from Pub. IV. Also, these studies did not utilize any supplemental network measurement data as in Pub. IV.

Ref. [60] used factor analysis of such a consumer survey to identify factors affecting customer satisfaction. The results illustrated that service availability significantly affected customer satisfaction. Though the definition of service availability in Ref. [60] differs from that in Pub. IV as it includes both the concepts of network speed and availability. Similarly, [41] used a similar methodology and found that service quality factors including system reliability (similar to network availability in Pub. IV) and connection speed (equivalent to network speed in Pub. IV) significantly affected customer satisfaction. Finally, [83] again used a similar methodology and found that perceived service quality (which includes service reliability as a component) significantly affected customer satisfaction.

In QoE terms, studies that examine temporal QoE with so-called long-term character [43] (i.e. perceived service reliability over time) or multi-episodic QoE are also related. Specifically, [30, p. 57-72] studied multi-episodic QoE of telecommunication services and found significant recency effects. Ref. [30, p. 73-85] also found that the overall service evaluation could be modeled through a weighted[3] average over the individual episode evaluations. Such a modeling approach is similar to the logistic regression model in Pub. IV that included both median and most recent throughput. Though the timeframe of [30, p. 57-72] was two weeks as compared to one year for Pub. IV. Refer to [30, p. 5-26] for further related work on retrospective evaluations and multi-episodic QoE.

In the broader and related domain of behavioral decision making, studies have elucidated some of the temporal dynamics of longer term evaluations [28]

[2] In other words, given the assumption that such a general service evaluation is essentially a long-term service evaluation rather than an episodic evaluation.

[3] Weights decreased linearly with respect to time before the final overall evaluation.

though such research remains sparse.

3.4 Distribution, process, and model studies

Many studies have performed distribution fitting on empirical mobile usage and experience data; however, this section focuses on rank-frequency distributions as these are widely illustrated in Pubs. I and II.

Several studies have proposed various theoretical distributions for modeling empirical rank-frequency distributions. Ref. [42] suggested the stretched exponential distribution family as an alternative to the common power law distribution for many rank-frequency distributions. While [47] similarly proposed the generalized discrete beta distribution (GDBD) as another alternative. Ref. [47] even explained the wide empirical applicability of the GDBD through a relationship of the GDBD to multiplicative hierarchical processes. In other words, processes similar to the FT process of Pub. I. However, in contrast to Pub. I, these studies did not explore rank-frequency distributions over time (in other words longitudinally).

Additionally, other work has examined processes very similar to the FT process but with different focuses. For example, [6] proved that a very similar fracturing process follows Benfords law. In other words, the first digits of the lengths of the resulting intervals follow the logarithmic distribution.

In terms of specific mobile usage distributions, [24] found that for most users their mobile application popularity (usage time) distributions can be well fit by an exponential distribution. Comparatively, Pub. II found these distributions were well fit by either a log-normal or stretched exponential distribution depending on the user.

4. Methods and datasets

This chapter describes the major methods and datasets used in the publications of this thesis. Several methods and datasets are used over several publications as detailed in Tables 4.1, 4.2, and 4.4.

Table 4.1. Summary of major data collection and analysis methods used in publications

Method	**Application**	**Pub.**
Mobile device based measuring	Collecting device usage, survey, and network perf data	II, III, IV, V
Agent-based modeling	Modeling end user network switching and multihoming scenarios	V
Stochastic process simulation	Simulating longitudinal rank-frequency distributions	I
Generalized ordinal logistic regression modeling	Modeling user satisfaction with network speed and availability	IV
Robust statistical testing	Testing for significance of differences in app usage	II, III
Distribution fitting	Fitting theoretical distributions to empirical data	I, II

4.1 Data collection methods

4.1.1 Mobile device based measuring

Mobile device based measuring was used to collect datasets 1 and 2 (see Section 4.4) which were used in Pubs. II III IV and to a small extent in Pub. V.

Mobile device based measuring is a method in which data is collected through software running on the actual end user mobile device. In other words, the measurement point is the mobile device itself. This method contrasts with network or server based measuring where the measurement point is further away from the end user.

In terms of specific benefits, mobile device based measuring generally provides more granular data (primarily on user behavior) than network or server based measuring. For example, mobile device based measuring (on Android) can monitor both app usage time and app network traffic whereas network based measuring can typically only monitor app network traffic[1]. Additionally, many apps use network encryption (HTTP over SSL) or non-specific HTTP user agents thus making network based identification of individual apps more difficult.

In terms of the specific challenges of mobile device based measuring, firstly, as previously mentioned, many device based monitoring apps only support Android or only support a subset of functionality on iOS thus limiting the extent of monitoring or forcing complex workarounds[2]. Secondly, the broad array of mobile device models and operating system versions (even within Android) means that measuring errors, exceptions, and incompatibilities are unavoidable[3]. Therefore, data cleaning and validation is a substantial task. Thirdly, since mobile device based measuring often collects sensitive information, privacy is a significant concern. Whereas in network based measuring this information is typically already encrypted; and in server based measuring the service associated with the server (for example mobile banking) in any case already has the specific sensitive information. As mentioned in Section 1.1, methods for alleviating such privacy concerns include incentivizing users with rewards and providing a clear privacy policy[4]. Finally, the lack of complete control of mobile devices by the measuring entity creates additional difficulties. For example, updating the measurement software might require user intervention if automatic updates

[1] Though, app network traffic can be used in a probabilistic method for estimating app usage (see [2]) or app network traffic can be directly taken as an imperfect proxy for app usage (for example, see [20]).

[2] For example, identifying the current foreground app in iOS through iOS APIs is not possible, and identification must be made through other methods. A common method is through machine learning models that are trained apriori to identify the foreground app through distinct data traffic patterns. Therefore, the identification is probabilistic.

[3] This is based on personal and research team experience.

[4] The privacy policy of the panel from which dataset 1 (see Section 4.4) was collected can be found at `https://smartpanel.io/en-US/about/privacy-policy/`, whereas the privacy policy of the app from which dataset 2 was collected can be found at `https://s3-eu-west-1.amazonaws.com/nettitutkadev/assets/documents/Netradar-privacy-v1.0.pdf`

are not enabled. Thus, different versions of the measurement software might be running simultaneously on different devices. Furthermore, users might purposefully or accidentally kill the measurement process thus creating temporal gaps in the measurements.

In general, partly due to the mentioned challenges, mobile device based measuring is significantly more expensive per user than network and server based measuring. Therefore, mobile device based measuring is often seen as complementing rather than competing with or replacing network and server based measuring. Specifically, a company might perform mobile device based measuring on a smaller sample of users and network and server based measuring on the entire population (census). Then the population results can be used to calibrate the sample results[5].

4.1.2 Mobile device based user surveying

When mobile device based measuring includes user surveys, as in dataset 2 (see Section 4.4), then the unique challenges inherent in surveying users come into play. These challenges include, for example, low response rates (again the response rate for the survey of dataset 2 was only about 15%), survey inconsistency, and false/fake answers. Additionally, in the mobile case, surveys should arguably be even shorter and more self-explanatory than in the PC case to boost response rates[6] which are lower in mobile surveys compared to PCs surveys [17]. Ref. [17] provides a further review of the issues prevalent in mobile surveys.

4.2 Simulation methods

4.2.1 Agent-based modeling

Agent-based modeling (ABM) was used in Pub. V to study the potential end user QoE given end user mobile network switching and multihoming.

ABM is a general modeling methodology that uses simple micro level rule-based agents that interact in a predefined environment. The interaction of the agents with each other and the environment often facilitate the observation of complex emergent macro level behavior [32] [46]. Such emergent behavior is

[5] See, for example, the methodology of Verto Analytics at `http://www.vertoanalytics.com/methodology/`

[6] Response rate generally decreases linearly with survey length [25].

typically seen in many complex systems thus making ABM particularly useful for understanding such complex systems. For example, researchers often model infectious disease spreading among people over realistic environments through agent-based models. A downside of agent-based models is that given their rule-based and interactive nature, such models are generally not analytically tractable (in contrast to other types of models).

Many general ABM platforms are open source and freely available including MASON[7], FLAME[8], and Repast[9]. These platforms provide a variety of utility and visualization methods that make model building easier. Additionally, most also allow for model creation in a high-level domain language or diagram method thus negating the need to have extensive coding skills. Though the use of object orientated languages in many of these platforms in any case matches to ABM since each agent can be a simple instantiation of an object class.

The motivation for using ABM in Pub. V was that the focus of ABM on agent behavior well matches the study focus on user (agent) benefit (and especially QoE). Additionally, rule-based agents are a good fit for simple feedback based models, such as the QoE feedback based usage model of Pub. V (see Section 3.5 of Pub V), since base rules can easily describe such feedback systems. Regarding the specific modeling platform, the Repast Symphony framework was utilized because Repast Symphony supports development in Java and Java has many third-party math and scientific libraries that can be leveraged in development.

4.2.2 Stochastic process simulation

Stochastic process simulation was used to illustrate the FT process proposed in Pub. I.

A stochastic process (also known as a random process) can be defined as a collection of random variables that share a common probability space and are indexed by a discrete or continuous set known as an index set [61]. In many common stochastic processes, this index set is continuous and denotes time.

In formalized terms, a stochastic process $\psi(t,\zeta)$ is an ensemble of functions where t is an index from the index set and ζ is a random variable. If t varies but ζ is fixed, then ψ provides a *sample* (a collection of outcomes over t) of the given stochastic process. If t is fixed but ζ varies then ψ provides a random variable that describes the *state* of the stochastic process at a given index. And if t and ζ are both fixed then ψ provides a single *outcome* [61].

[7] http://cs.gmu.edu/ eclab/projects/mason/

[8] http://flame.ac.uk/

[9] http://repast.github.io

In this thesis, the term stochastic process simulation means obtaining many *outcomes* of a given stochastic process. Then these outcomes can be used to illustrate the general behavior and variability of the process.

The FT process of Pub. I can be formalized (in a somewhat simplified form) as in Equation 4.1 (recurrence relation in set builder notation). Where L_t represents the process *state* at index t, x represents a value in the *state*, ζ represents a random variable (PDF shown in Equation 5.1), and L_1 represents the initial condition of the recurrence relation with a constant C_{FT} (in Pub. I, $C_{FT} = 1$). Also of note is that the *outcome* of this process (in other words a single *outcome*) is a list of interval sizes rather than a single numeric value.

$$\begin{aligned} L_t &= \{x * \zeta, x - (x * \zeta) | x \in L_{t-1}\} \\ L_1 &= \{C_{FT}\} \end{aligned} \tag{4.1}$$

Furthermore, the final interval size list F (which itself is of size 2^N where N is the final index t) can be formalized as in Equation 4.2 where p_1 is the random variate at $t = 1$, p_2 is the random variate at $t = 2$ and so on.

$$\begin{aligned} F_1 &= C p_1 p_2 p_4 \cdots p_{2^{N-2}} p_{2^{N-1}} \\ F_2 &= C p_1 p_2 p_4 \cdots p_{2^{N-2}} (1 - p_{2^{N-1}}) \\ &\vdots \\ F_{2^{N-1}} &= C(1 - p_1)(1 - p_3)(1 - p_7) \cdots (1 - p_{2^{N-2}-1}) p_{2^{N-1}-1} \\ F_{2^N} &= C(1 - p_1)(1 - p_3)(1 - p_7) \cdots (1 - p_{2^{N-2}-1}) 1 - p_{2^{N-1}-1} \end{aligned} \tag{4.2}$$

Many textbooks including [61] provide further background on stochastic processes.

4.3 Analysis and statistical modeling methods

4.3.1 Generalized ordinal logistic regression modeling

Generalized ordinal logistic regression modeling was used to model user satisfaction as a function of network and non-network variables in Pub. IV.

Generalized ordinal logistic regression modeling is a modeling method for cases where the dependent variable is not numeric but does have a natural order (in other words is ordinal). In generalized ordinal logistic regression models (as in all logistic regressions) the log odds of the dependent variable are modeled as a linear combination of the independent variables. The difference between general and ordinary ordinal logistic regression relates to the proportional odds assumption (also known as parallel lines assumption). The ordinary case requires this assumption while the generalization does not. Furthermore, the assumption can be relaxed for only a subset of all independent variables thus allowing for a partial proportional odds (PPO) model (that falls between the ordinary and general extremes). The generalized and PPO models can be fitted in Stata through the user defined `gologit2` [96] command.

More specifically, the generalized ordinal logistic regression model can be parametrized as detailed in Equation 4.3 where M is the number of categories of the ordinal dependent variable, X_i are the independent variables, α_j is a constant, and β_j are the independent variable coefficients [96]. Thus for $M = 5$, if $j = 1$ then $P(Y_i > 1)$ relates the categories {1} vs {2,3,4,5}, and if $j = 2$ then $P(Y_i > 2)$ relates the categories {1,2} vs {3,4,5}, and if $j = 3$ then $P(Y_i > 3)$ relates the categories {1,2,3} vs {4,5}, and so forth.

$$P(Y_i > j) = g(X\beta_j) = \frac{\exp(\alpha_j + X_i\beta_j)}{1 + \{\exp(\alpha_j + X_i\beta_j)\}}, j = 1, 2, \ldots, M-1 \quad (4.3)$$

In this parametrization the proportional odds assumption can be related to whether the β coefficients are the same for all $P(Y_i > 1)$, $P(Y_i > 2)$, $P(Y_i > 3)$, etc. Thus, the advantage of the assumption holding for all or even some of the independent variables is a more parsimonious model. Further treatment of generalized ordinal logistic regression modeling can be found in Ch. 5 of [45].

The motivation for using Generalized ordinal logistic regression in Pub. IV was that the mobile survey that collected user satisfaction levels used a Likert scale (i.e. from strong disagree to strongly agree) which is considered an ordinal scale. Furthermore, generalized regression was required since some but not all the independent variables met the proportional odds assumption.

4.3.2 Robust statistical testing

Robust statistical tests were used to test for the statistical significance of usage patterns from highly non-normal data with many outliers in Pubs. II and III.

Robust statistical test methods are methods that can handle data problems such as non-normality, heterogeneity of variances, and significant outliers.

For example, Wilcox's trimmed mean percentile bootstrap method [95] can be used for testing for significant differences between the trimmed mean of two distributions even given the mentioned problems. Specifically, the outlier problem is solved by using the trimmed mean which is a centrality measure robust to outliers. Whereas the non-normality and heterogeneity of variances problems are solved by using a bootstrap method that does not rely on confidence intervals that assume a specific underlying distribution (i.e. normal distribution).

The bootstrap philosophy relies on creating confidence intervals through repeated random resampling (with replacement) of the distribution. For example, imagine a distribution of size n and a trimmed mean centrality measure, for X iterations the original distribution is sampled (with replacement) n times thus giving X new distributions of size n. The trimmed mean of each of these new distributions thus gives a distribution of trimmed means. This distribution of trimmed means can then itself be used to estimate a confidence interval for testing against the confidence interval of another distribution. (In other words, a two-sample trimmed means test.)

Usefully, this percentile bootstrap method applies in cases of both paired and non-paired distributions. Ref. [95] provides further information on robust statistical testing.

The permutation test is a related robust statistical test that can be used for testing, among other things, the significance of linkages between bipartite sets or groupings. Imagine an undirected bipartite graph where each edge's weight is a function of the two nodes the edge connects. A permutation test can test whether the median edge weight of such an observed (empirical) graph is significantly different than would be expected in a random degree-preserving graph (denoted as a dp-graph). In other words, would the observed median edge weight likely occur simply by chance?

Specifically, the permutation test creates dp-graphs by permuting the edges of the observed graph with the constraint that the number of edges per node (and thus the total number of edges) remains the same as in the observed graph (hence the degree-preserving description). This permutation method is repeated for all possible permutations to acquire a full distribution of dp-graphs. Then the median edge weight of each of these graphs is calculated and a confidence interval for median edge weight is found. The observed median edge weight can be compared against the confidence interval to test for a significant difference.

In practice, using all permutations is typically computationally infeasible. For example, the number of potential dp-graphs in the case of a 1-regular balanced bipartite graph[10] is $\frac{n}{2}!$ where n is the number of nodes[11]. Thus, approximate permutation tests (also known as Monte Carlo permutation tests) typically use only a small random fraction (often 10^4 or 10^5) of all possible dp-graphs. This

[10] This problem equates to finding the number of perfect matchings of an n node complete balanced bipartite graph.

[11] For the approximate permutation test of Section 5.2 in Pub. II, the total number of dp-graphs is thus $\frac{114}{2}! \approx 4.05 \times 10^{76}$.

approximation means that the percentiles of the distribution of median edge weights themselves have some potential margin of error (that depends on the number of dp-graphs used) and must also be considered in making p-value and confidence interval comparisons with the observed median edge weight.

For example, the type I error rate of the approximate permutation test of Section 5.2 in Pub. II can be calculated through Equation 4.4 where $\hat{p}$ is the approximate permutation test p-value, m_p is the number of permutations (specifically the number of dp-graphs used), and δ is the error rate level (typically 0.05, 0.01, 0.001, etc.) [65]. Thus, for the test with $m_p = 100000$ and $\delta = 0.0001$ the actual error rate is 0.00011 rather than 0.0001 due to this approximation. (The erroneous value of 0.0001 was reported in Pub. II, though as we note in the errata this does not change any conclusions of Pub II).

$$P(\hat{p} \leq \delta) = \frac{\lfloor m_p \delta \rfloor + 1}{m_p + 1} \tag{4.4}$$

4.3.3 Distribution fitting

Distribution fitting was used in II to determine the best fitting theoretical distributions for empirical data and in Publication I to compare several theoretical distributions and the FT process.

Fitting in Pub. II

The distribution fitting method in Pub. II was based on methods developed by [15] and [1]. Specifically, each candidate theoretical distribution was fit to the empirical data through maximum likelihood estimators of the distribution parameters. Then the candidate with the highest Akaike weights[12] was selected as the best-fit candidate. This best-fit candidate was then compared against all other candidates through Vuong likelihood ratio tests [90] to determine whether the best candidate fit was significantly better than all other fits. In other words, several different distributions maybe equivalently good fits in a statistical significance sense.

Fitting in Pub. I

Contrastingly, the distribution fitting method in Pub. I was based on linear regression (least squares) fitting on log transforms of the distributions. The primary reason for utilizing the linear regression methodology was that the

[12]Akaike weights are related straightforwardly to Akaike information criterion (AIC) by $w_i = \frac{e^{-(A_i - \min(A_s))/2}}{\sum_{r=1}^{R} e^{-(A_r - \min(A_s))/2}}$ where w_i is the Akaike weight for candidate i, A_i is the AIC for candidate i, A_s is set of AICs for all candidates, and R is the total number of candidates.

original work describing the GDBD [47] (the main distribution compared to the simulation) used the methodology. Furthermore, the GDBD PDF does not have a known analytical form, and thus maximum likelihood estimation would be difficult since the log likelihood function also does not have a known analytical form[13]. The distribution fits were compared through adjusted[14] and weighted[15] coefficients of determination (R^2).

4.4 Mobile device based measurement datasets

Two distinct mobile device based measurement datasets were used; Table 4.2 summarizes these datasets.

Table 4.2. Summary of mobile device based measurement datasets used in publications

Collection Period	Users	Data Collected	Pub.	Abbr.
February 2015	999[a]	Device Usage	II, III, V	Dataset 1
Summer 2014-2015	2224	Network Perf & Survey	IV	Dataset 2

[a] The different publications analyze different subsets of the total panel of 999 users. Table 4.3 details these subsets.

Dataset 1 consisted of mobile device based usage measurements collected for one month (February 2015) from a large on-going user panel in the US organized by Verto Analytics. Potential panelists were initially recruited online through advertisements and other means and then given a recruitment survey to collect demographic and technographic information. Potential panelists were then accepted to or rejected from the panel based on the demographic and technographic match between the currently accepted panelists and the target population (US consumers). This method is known as quota sampling. All panelists were paid for participation. The measurements include device usage information such as app sessions and visited URLs for all of user's reported devices including smartphone, tablet, and PC. Table 4.3 summarizes the different subsets of this dataset analyzed or used in different publications.

In contrast, dataset 2 consisted of mobile device based measurements collected from a large group of users of the mobile app Netradar. Netradar is available for both Android and iOS and is relatively popular (with well over 20K installations).

[13] Though the PDF and log likelihood function in the case of the GDBD where the two distribution parameters are equal (also known as the Lavalette distribution) do have analytical forms [27]

[14] The adjusted R^2 uses a penalty based on the number of distribution parameters to avoid R^2 inflation.

[15] The weighted R^2 was weighted inversely proportional to rank, thus giving more weight to higher ranks.

Table 4.3. Summary of subsets of dataset 1 used in publications

Subset Criteria[a]	Users	Publications	Abbreviation
Active[b] S	562	II, V	Subset 1
Active T	125	II	Subset 2
Active PC	630	II	Subset 3
Active S, T	77[c]	II	Subset 4A
Active S, T	65[c]	III	Subset 4B
Active S, ¬ T[d]	496	III	Subset 5
Active S, PC	269	II	Subset 6
Active T, PC	52	II	Subset 7

[a] S (Smartphone), T (Tablet), PC (Laptop, Desktop)

[b] Active denotes in all cases that the length between the first and last usage of the device was at least 23 days.

[c] The difference in the number of users between the two subsets relates to slightly different preprocessing of the raw device event data. Therefore, the subset of Pub. II is slightly more reliable.

[d] The negation ¬ indicates that the user did not report owning a tablet in their initial recruitment survey.

The app allows users to perform network performance measurements both manually and in an automated fashion. For example, the network speed test uses a bulk data transfer method over TCP to estimate network throughput. The app collects a large variety of network performance parameters including upload and download throughput and latency. Additionally, the app occasionally (about every three months) prompts users to complete a short survey about network satisfaction.

Netradar initially acquired users through word of mouth and several high profile national news articles[16]. In terms of user motivation, motivations were primarily intrinsic. Anecdotally, many users installed the application to know the network performance at their home or office (simply for curiosity), to test network performance at troublesome locations, or to help advance university research. In contrast to dataset 1, no direct extrinsic reward (i.e. payment) was provided to participants. Extrinsically and intrinsically motivated users present different trade-offs in terms of the collected data [35]. Refer to [35] for further discussion about user motivations in crowdsourcing.

The dataset itself consisted of network measurements and survey responses from Finland-based Netradar users that completed the network satisfaction survey in summer 2015 and had at least five network measurements in the one

[16]http://www.hs.fi/talous/art-2000002682298.html

year before survey completion. Thus, the dataset consisted of both objective network performance measurements and subjective user survey responses.

4.5 Public rank distribution datasets

Several diverse public rank distribution datasets were used to illustrate the fitting of the FT process of Pub. I. Table 4.4 summarizes these datasets. Several of the datasets were longitudinal datasets that included data for each year over a several year time frame. The collection methods and properties of these datasets are not discussed in detail but can be found in the given citations of Table 4.4.

Table 4.4. Summary of public datasets used in Pub. I.

Dataset	**Collection Period**	**Citation**
French Book Sales	2003-2007	[71]
US Theatrical Earnings	2002-2012	[10]
Census Last Name Frequency	2000	[85]
Audioscrobbler Artist Play Freq	2005	[7]
UMASS[a] YouTube Video Request Freq	2007-2008	[84]
US Magazine Circulation Revenues	2000-2012	[80]

[a] University of Massachusetts Amherst

5. Results

This chapter presents the results of this thesis in the same framework as the research questions and thus in four distinct sections. The first three sections describe the results pertaining to the three contexts: multiple mobile devices, multiple mobile networks, and long-term mobile service. The final section relates to the identification, modeling, and simulation of several of the distributions that characterize these contexts.

5.1 Multiple mobile device context

To illustrate the empirical usage of mobile services in multiple mobile device contexts, a multidevice session definition and construction framework was proposed, a statistical analysis of multidevice sessions was performed using subset 4B of dataset 1, and a statistical analysis of app usage concentration of multiple device types was performed using subsets 1-4A, 6 and 7 of dataset 1.

5.1.1 Multidevice session construction

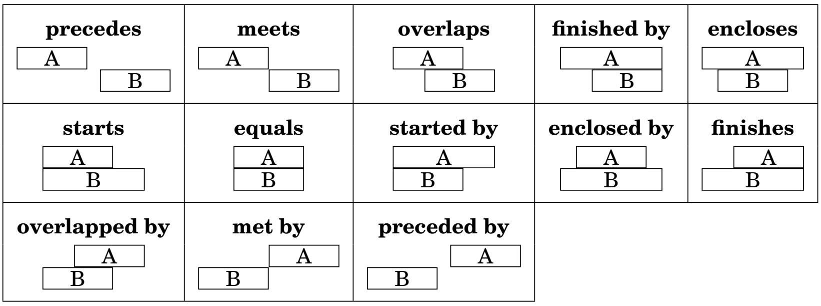
As a precursor to an analysis of multidevice sessions, a framework was required to define the notion of a multidevice session as previous works did not provide a formal and consistent definition. First, framework relates the idea of an app session to a temporal interval (as in previous work [77]). Specifically, the framework defines an app session as the temporal interval covering the time from the app entering the foreground[1] of the device to leaving the foreground. Then the framework applies an extension of the well-known Allen's Interval Algebra to these temporal intervals.

Allen's interval algebra defines a complete set of possible relationships between two temporal intervals. For example, an interval from $t = 1$ to $t = 4$ *precedes* an interval from $t = 6$ to $t = 8$, whereas an interval from $t = 1$ to $t = 4$ *overlaps*

[1] In the case of PC, the definition relates to the app window currently in focus on the screen.

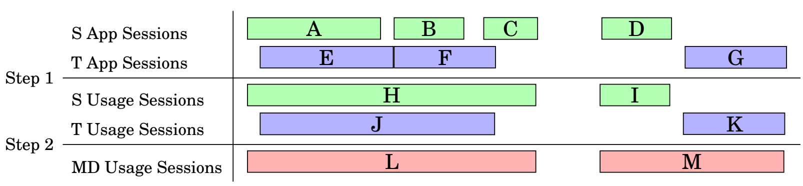
an interval from $t = 3$ to $t = 6$. Table 5.1 illustrates the complete set of these algebraic relationships. The framework also adds two additional relationships *precedes within time window* and *preceded by within time window* given a time window (TW) t. This addition essentially breaks up the course *precedes* and *preceded by* relations into two cases depending on the time between the sessions. This is important because mobile app sessions do not occur in a randomly distributed manner but instead are highly bursty and this time window helps in grouping sessions given this burstiness.

Table 5.1. Allen's time interval relations. Adapted from Pub. III



Finally, given this relational algebra, the framework defines a two-step process for constructing multidevice sessions. First for each device up to device N, the process constructs single device usage sessions from app sessions by grouping app sessions into a single usage session if the sessions *meet*, *metby*, *precedes within TW*, or *preceded by within TW*. Additionally, if an app session is isolated, it also becomes a single device usage session. Second, the process constructs multidevice sessions by grouping single device usage sessions of different devices into a multidevice usage session if the sessions partly overlap (all relationships except the precedes and preceded by) or the sessions *precedes within TW*, or *preceded by within TW*. Table 5.2 illustrates this process for an example case of smartphone and tablet app sessions.

Table 5.2. Multidevice usage session construction. Step 1: Apply usage session definition to each device's app sessions (merge A, B, C to H and E, F to J; D to I and G to K are degenerate single app session cases), Step 2: Collapse simultaneous, meeting and preceding within time-window usage sessions into multidevice usage sessions (merge H, J to L and I, K to M). Adapted from Pub. III



5.1.2 Multidevice session analysis

Next, a statistical analysis of the multidevice sessions derived from subset 4B of dataset 1 was performed.

In comparison to the single device usage sessions, the multidevice usage sessions had significantly different dynamics (even given that the sessions came from the same users). Specifically, multidevice usage sessions were about 10 times longer and had five times as many app sessions as single device usage sessions. However, multidevice usage sessions were also about 14 times less frequent than, for example, smartphone usage sessions. Therefore, interestingly, the overall time spent by users in these multidevice sessions compared to single device sessions was approximately equal (about 180 minutes per day in each case). Given this prevalence in everyday use, multidevice sessions are likely already ubiquitous for multidevice users.

The temporal patterns within and diversity between individual multidevice sessions were characterized by clustering multidevice sessions through a matrix similarity measure known as Interval-Based Sequence Matching. Given a group of prototype matrices, the multidevice sessions were clustered such that each session was assigned to one of the 256 prototype groups. The two most common groups (comprising over 35% of multidevice sessions) were sessions with one dominant device and the infrequent usage of the secondary device. For example, a typical tablet dominant session consists of a tablet game app and occasional use of smartphone notification driven social network and news apps. Furthermore, an examination of the specific types of apps in these groups indicated that inferring the type of app used on another device based on the app used on a given device may be possible. Such inference could be useful in providing contextual suggestions that help ease multidevice usage.

Regarding diversity, over 25 different groups had frequencies over 1% thus indicating significant diversity in the temporal patterns of multidevice sessions. Furthermore, only 8% of multidevice sessions had exactly 2 app sessions with the mean number of app sessions being 22. Thus, the base case of using a single app on a device and then a single app on another device is relatively rare in practice, and complex sessions are the norm.

Finally, in temporal terms, diurnal patterns of multidevice sessions were driven by the usage patterns of tablets since tablet usage was more limited and more concentrated than smartphone usage which was more evenly distributed throughout the day.

5.1.3 Multidevice app concentration analysis

The multidevice session analysis did not describe how users divide their time between distinct apps across devices. Therefore, an analysis of the concentration of app usage between device types was also performed. This analysis was performed using subsets 1-3 of dataset 1 (notice the inclusion of PC data). Also, due to the use of different app names on iOS and Android platforms, the analysis looked at device type-platform combinations rather than simply different device types.

Figure 5.1 illustrates the rank-frequency distributions of overall normalized app usage time for different device type-platform combinations. All device types had very high concentration with Gini indexes[2] above 0.90. PC devices had the highest concentration with three clearly dominant apps (as seen in the first three ranks in Figure 5.1) which were all web browsers. Such high concentration emphasizes the dominance of the web browser on PCs.

High concentrations were also found on the individual device level with mean Gini indexes above 0.80. In other words, for both smartphones and tablets, users tend to spend most of their time on only a tiny subset of heavily used apps.

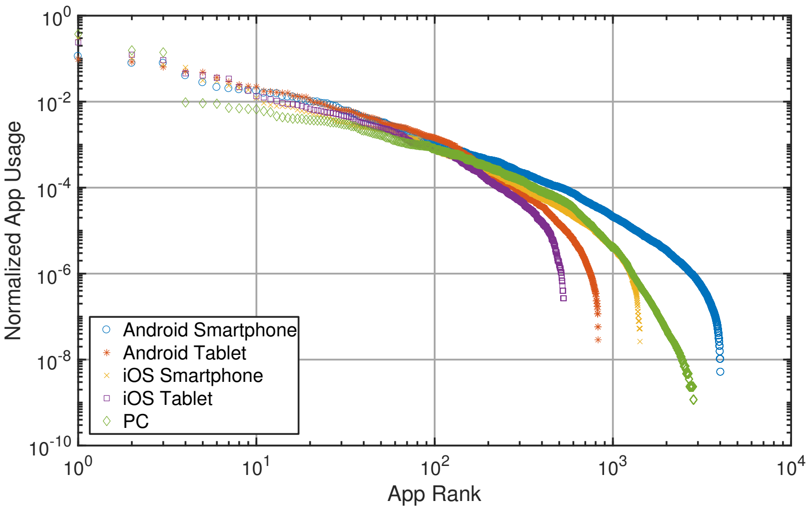


Figure 5.1. Rank-frequency distributions of normalized app usage time for device type-platform combinations. From Pub. II.

However, these high overall concentrations hid differences in concentration between individual app categories for different device types. Figure 5.2 illustrates normalized app usage times vs. Theil indexes[3] for various app categories by

[2]Gini index is a common concentration measure with a range of $[0,1]$.

[3]Theil index is another common concentration measure with a range of $[0,1]$ but with an information theory basis and more sensitivity to large tail values than Gini Index [18].

device type for Android devices. Interestingly, app categories that theoretically have large network effects (such as social networking) showed higher concentration than app categories without such a theoretical underpinning (such as games).

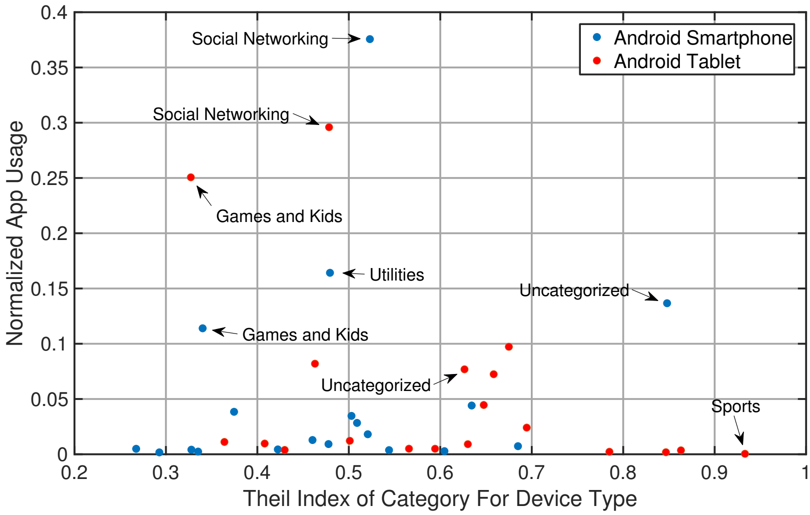


Figure 5.2. Normalized app usage vs. Theil indexes for app categories by device type (Android only). From Pub. II.

Finally, the nature of the data means that analyzing the concentration of app usage time (or other app usage measures) across multiple devices of the same user was possible. This analysis was performed on subsets 4A, 6, and 7 of dataset 1. Table 5.3 details the correlations of Theil indexes and the number of used apps across device type ownership combinations. Surprisingly, the results showed no significant correlations in terms of app concentration across any combination. Thus, understanding usage phenomena such as app concentration across all of a user's devices will likely require monitoring all of a user's devices since making accurate inferences from a single device appears infeasible.

Interestingly, the correlation between tablet and PC in terms of the number of used apps was both negative and significant. This relationship suggests that the more tablet apps a user uses, the fewer PC apps they will use. In other words, the relationship suggests an amount of device substitution. In further support, the correlation between this combination for total usage time was also negative and significant (-0.242, $p < 0.05$).

Given the non-significance of correlations between smartphones and tablets, a further examination of the relationship between the apps of a user's smartphone and the apps of their tablet was performed. Specifically, the Jaccard similarity[4]

[4]The Jaccard similarity between two sets is defined as the cardinality of the intersection of two

Table 5.3. Pearson correlation coefficients (including significance levels[a]) for Thiel indexes and total number of utilized apps for different device type combinations. Adapted from Pub. II

Device Type Combin[b]	Corr (Theil Index)	Corr (Num Apps)
S, T (subset 4A)	0.054	0.146
S, PC (subset 6)	0.022	-0.024
T, PC (subset 7)	0.042	-0.305*

[a] Significance levels * : 5%, ** : 1%, *** : 0.1%

[b] S (Smartphone), T (Tablet), PC (Laptop, Desktop)

between the sets of used smartphone apps and used tablet apps for each user was calculated. Interestingly, the similarity was surprisingly small with a median user similarity of just 0.169 (or 16.9%). Therefore, similar to the previous observations, the knowledge of apps used on a given device of a user does not necessarily imply that the same apps are used on a different device of that user. Though, a permutation test showed that this median similarity is still highly statistically significant ($p < 0.001$). Permutation tests, in general, are further detailed in Section 4.3.2.

5.2 Multiple mobile network context

To illustrate the potential QoE of users of mobile services in multiple mobile network contexts, an analysis of end user network switching (NS) and multihoming (MH) scenarios was performed.

The analysis used an agent-based modeling approach such that each user's behavior was modeled in detail. Section 4.2.1 further details agent-based modeling in general as a method. The baseline scenario consisted of a group of 300 network users moving in a 0.09 km^2 area that includes both indoor and outdoor locations and indoor and outdoor base stations from three mobile networks. Network and user parameters of the model (especially those relating to NS and MH) were then varied to explore the result space. The main performance metrics were mean user download throughput and mean user MOS.

Figures 5.3 and 5.4 illustrate mean user throughput and MOS for the different network (base station) layouts and user types. The analysis indicates that for NS cases as the number of available networks (to switch between) increases both throughput and MOS also significantly increase. This observation emphasizes

sets divided by the cardinality of the union of those sets, for sets A and B the formulation is $\frac{|A \cap B|}{|A \cup B|}$.

the value of NS for improving end user QoE. The increase was especially evident in the case of scenario 4 that includes an indoor only operator[5] since NS allows the indoor operator's customers to receive much better quality (from an alternative network) while outdoors (rather than relying on the highly attenuated indoor BS signals).

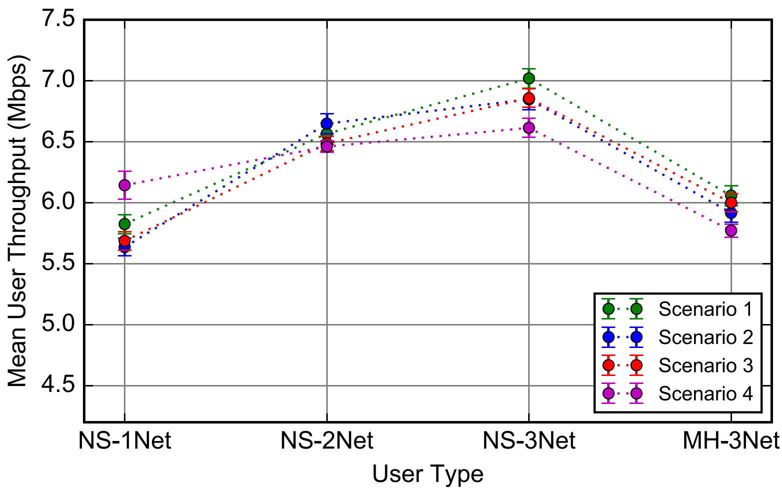


Figure 5.3. Throughput values for different scenarios and user types (all users are a single type) given ER radio resource allocation scheme. Adapted from Pub. V

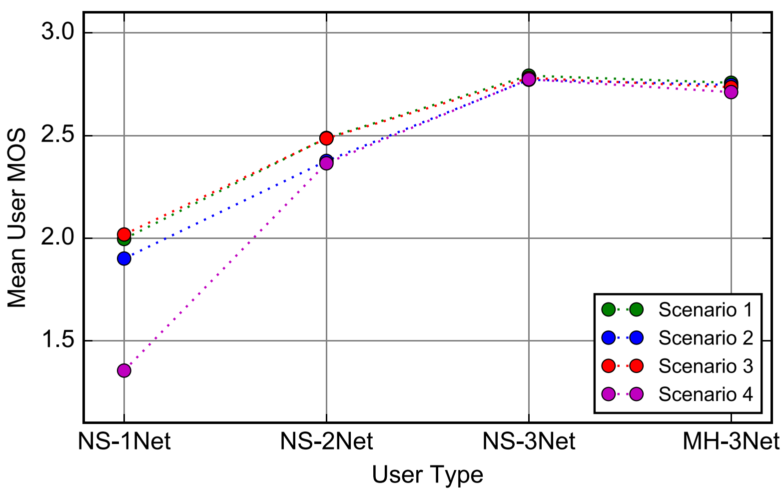


Figure 5.4. MOS values for different scenarios and user types (all users are a single type) given ER radio resource allocation scheme. Adapted from Pub. V

Interestingly the performance of MH (with 3 networks) was less regarding throughput and equal regarding MOS compared to NS (with 3 networks). The

[5]An indoor operator is an operator with base stations only in indoor locations.

relatively poor performance of MH was the result of inefficient radio resource allocation due to networks not accounting for the MH nature of users (specifically, MH users are over-provisioned).

However, in the case of a small number of users (and thus low congestion and little competition for resources), MH did significantly improve throughput. Though, MOS was not improved because the already high throughput in these low congestion situations was adequate for most mobile services. Figures 5.5 and 5.6 illustrate throughput and MOS values for different user types (assuming that all users of a single type) given different user densities.

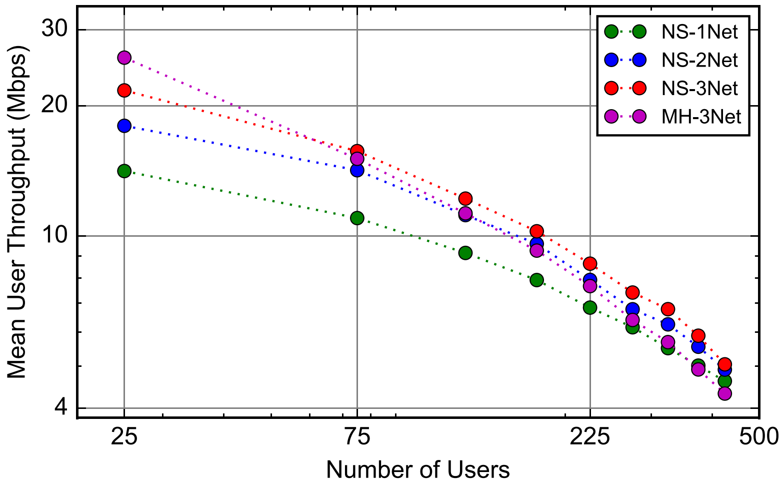


Figure 5.5. Throughput values for different numbers of users and NS and MH user types given ER radio resource allocation scheme and scenario 1. Adapted from Pub. V

Finally, in the case where a fraction of users were MH (with 3 networks) and the remaining fraction of users were NS (with 3 networks), interesting dynamics were identified. Specifically, when the proportion of MH users was small (relative to NS users) each MH user made a large gain while each NS user made a small loss, but when the proportion of MH users was large each MH user made a small gain while each NS users made a large loss. In other words, large gains from MH were only available when the fraction of users MH was small; such dynamics further limit the applicability of end user MH.

Given this overall analysis, user experience in multiple mobile network contexts (as quantified through MOS) would likely be significantly improved through end user NS (especially in certain cases such as indoor only operators) but the additional improvement through end user MH would be limited.

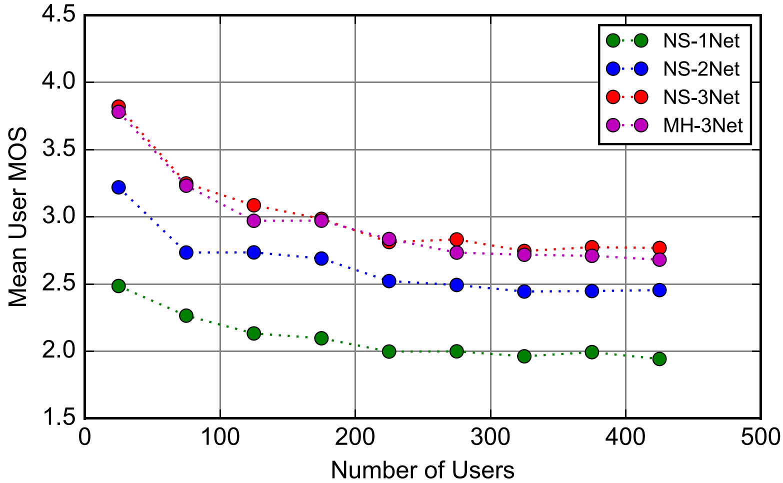


Figure 5.6. MOS values for different numbers of users and NS and MH user types given ER radio resource allocation scheme and scenario 1. Adapted from Pub. V

5.3 Long-term mobile service context

To illustrate the experience of users of mobile services in long-term mobile service contexts, a modeling-based analysis of dataset 2 was performed.

Specifically, a variety of network and non-network features were regressed against the responses to each network satisfaction question from the survey so that the significant factors affecting user's responses could be identified. The main questions of focus dealt with user satisfaction of network speed and availability. All the created regression models were ordinal logistic partial proportional odds models. Section 4.3.1 further details generalized ordinal logistic regression modeling as a method. Table 5.4 details the significance levels for features with significance $p < 0.05$ for at least one of the speed and availability questions.

The significance of the median, most recent (last), and minimum download throughput values on network speed satisfaction suggests the influence of several temporal phenomena including the peak-end effect [39] and aggregation over several network measurement results.

The effect sizes of the significant features were illustrated through predicted probability functions that show the change in the outcome variable as the significant feature is varied and all other variables are kept at mean levels. Figure 5.7 illustrates these predicted probability functions for four significant features of the speed satisfaction model (note that the outcome Likert scale is dichotomized to a satisfaction vs. non-satisfaction binary variable). The TCP

Table 5.4. Feature significance levels[a] for models of availability and speed questions. Adapted from Pub. IV.

Feature	Availability	Speed
Median TCP Download	*	**
Last TCP Download		***
Min TCP Download	**[b]	***[b]
Last RTT	*	
Max RTT	*	
Number Frequently Measured Locations	**	*
Number Invalid User-Initiated Measurements		*
Device Platform (Android or iOS)	*	*
Mobile Network Operator	**	
Device Type (Tablet or Smartphone)	**	

[a] $*$: 5%, $**$: 1%, $***$: 0.1%

[b] Feature did not satisfy parallel lines assumption, reported significance level is for baseline of (1,2) vs (3,4,5).

download functions illustrate slight diminishing returns as download speeds increase.

The lack of significance of latency features in network speed satisfaction indicates that either users are not using latency in their satisfaction evaluations[6] or that the current latency levels are good enough not to affect their evaluations or likely both. For example, most mobile apps do not require latencies below 100ms [91], and 90% of respondents have median RTTs below this 100ms threshold. Furthermore, the lack of significance of upload features suggests similar possibilities for upload as for latency.

5.4 Distribution methods

To illustrate the distributions, processes, and models that can characterize this user usage and experience the stochastic FT process was proposed and empirical fitting of several mobile usage distributions was performed. The FT process can simulate empirical rank-frequency distributions including the longitudinal variations of the distributions.

[6] latency is likely a more abstract concept to respondents since latency is not as often advertised or discussed.

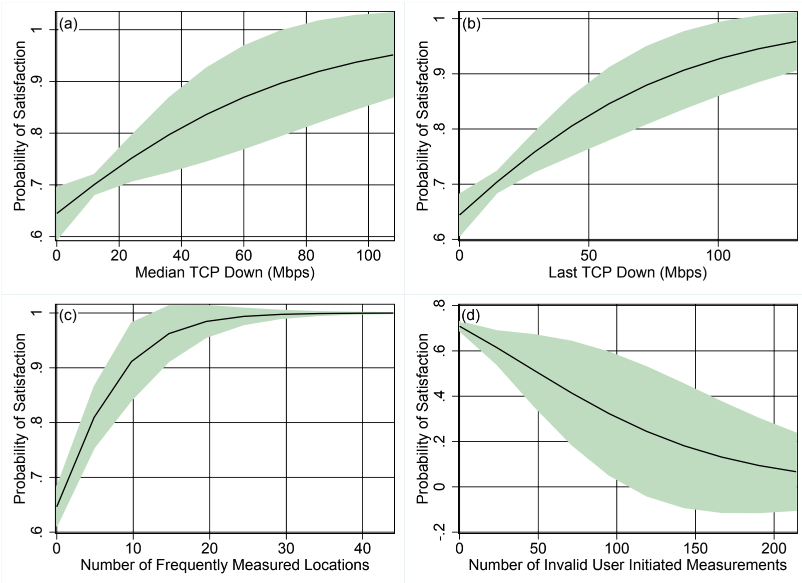


Figure 5.7. Predicted probability functions for several significant features of speed satisfaction question: (a) Median TCP download, (b) Last TCP download, (c) Number of frequently measured locations, (d) Number of invalid user initiated measurements. From Pub. IV.

5.4.1 Simulating empirical rank-frequency distributions

A simple stochastic process, known as the FT process, for simulating empirical rank-frequency distributions was proposed. Section 4.2.2 further details stochastic processes, in general. The FT process works by recursively splitting a unit interval into smaller intervals until the number of final intervals matches the number of empirical ranks. Each splitting occurs at a point given by a random variate from a random variable with probability density function (PDF) as in Equation 5.1. The fitting parameter of the FT process is γ in the PDF. Figure 5.8 illustrates the process for 2 time steps (thus resulting in 2^t final intervals given t time steps (which implies $2^t - 1$ individual splittings and random variates)).

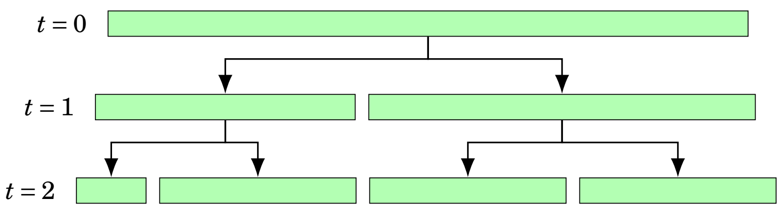


Figure 5.8. A visualization of the FT process after 2 time steps ($t = 2$) and thus 4 resultant intervals. Adapted from Pub. I.

$$\text{PDF} = \left\{ \begin{array}{ll} \frac{\gamma}{\gamma-1} & \left(0 < U_{[0,1]} < \frac{1-\gamma}{2}\right) \\ \frac{1}{\gamma} - 1 & \left(\frac{1-\gamma}{2} \leq U_{[0,1]} < \frac{1+\gamma}{2}\right) \\ \frac{\gamma}{\gamma-1} & \left(\frac{1+\gamma}{2} \leq U_{[0,1]} < 1\right) \end{array} \right\} \tag{5.1}$$

The FT process can well simulate a variety of rank-frequency distributions, especially when the fitting method uses a weighted metric (such as weighted R^2) with larger ranks (in other words the first-ranked, second-ranked and so on) given more weight[7]. The stochastic nature of the process allows for simulating rank-frequency distributions longitudinally by simply repeating the method with different random number seeds. For example, Figure 5.9 illustrates that the FT process well simulates the longitudinal variation in the ranks of French book sales.

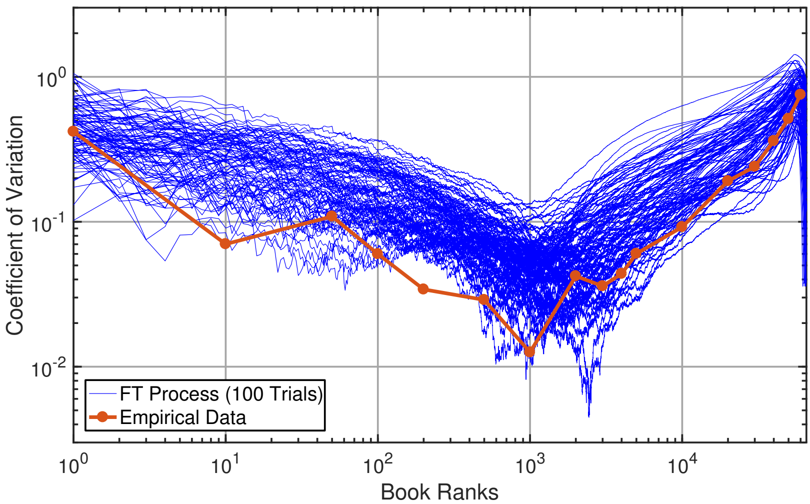


Figure 5.9. Coefficients of variation of ranks for the 2003–2007 French book sales dataset and simulated coefficients for 100 FT process trials. (Note that the French book dataset consists of only 17 disparate ranks.) Adapted from Pub. I.

As detailed in Section 4.2.2, the FT process can be well represented through a recursive definition and thus the practical coding implementation of the process is also well suited to recursion. Such a recursive implementation in Perl was provided as supporting material at the online version of Pub. I.

The FT process was also compared to several other commonly proposed theoretical distributions for simulating empirical rank-frequency distributions such as the power law distribution and the DGBD [54]. The FT process was shown to perform well on rank-frequency distributions with a concave shape and reciprocal head and tail slopes. Whereas, the DGBD performed better on less concave, power law like bodies. Both methods performed better than the pure power

[7] This weighting scheme is intuitive given that emphasis is often place on these large ranks.

law distribution due to the empirical truncated tails that result from finite size effects. Section 4.3.3 details the fitting procedure for these common theoretical distributions.

5.4.2 Further fitting empirical distributions

To further illustrate the fitting of empirical distributions, several common long tail distributions were fit to empirical app usage distributions. Section 4.3.3 details the fitting method for these distributions. Table 5.5 details the best-fit distributions for the aggregate app usage of different device types.

Such distribution fitting should help future modeling efforts by allowing calibration of models. For example, the app session length distributions used in the user model of Pub. V were derived from distribution fitting of app session length data from dataset 1.

Table 5.5. Parameter estimates and CDF error for aggregate app usage distribution fitting. From Pub. II.

Device Type	Best Fit[a]	Param Est	RMSE[b](%)
Smartphone (Android)	LN	$\mu = 6.624$ $\sigma = 2.703$	1.94
Smartphone (iOS)	LN	$\mu = 6.392$ $\sigma = 2.512$	1.71
Tablet (Android)	LN	$\mu = 7.092$ $\sigma = 2.832$	1.82
Tablet (iOS)	LN	$\mu = 6.381$ $\sigma = 2.481$	1.41
PC (Windows)	SE	$\lambda = 0.002$ $\beta = 0.173$	1.78

[a] LN (Log-Normal), SE (Stretched Exponential)

[b] Root mean squared error between empirical and predicted cumulative distribution functions.

5.5 Results summary

The main results can be summarized such that only the most essential parts of the thesis are emphasized. Table 5.6 both restates the research questions and summarizes the main thesis results in relation to each research question.

Table 5.6. Summary of research questions, publications, and main thesis results

Research Questions			
RQ1: How do users use mobile services in the context of multiple mobile devices?	**RQ2:** How do users experience mobile services in a long-term (≥ 1 year) context?	**RQ3:** How would users experience mobile services in the context of multiple mobile networks?	**RQ4:** What distributions, processes, and models can characterize the user usage and experience of these contexts?
Publications			
II,III	IV	V	I, II, IV, V
Main Results			
• Analysis of multidevice sessions illustrating the diversity and ubiquity of such sessions for users with smartphone and tablet devices. • Analysis of app usage concentration illustrating differences in usage concentration between multiple devices types including for different devices of the same user.	• Analysis illustrating that users experience of long-term mobile services includes complex temporal phenomenon including the peak-end effect and integrating over multiple episodes. • Analysis illustrating that users experience can only be weakly explained by that user's network measurement episodes thus motivating further work.	• Analysis of multiple network scenarios illustrating that end user network switching provides significant benefits in both throughput and QoE in many scenarios. • Analysis also showing adoption of end user multihoming does not provide comparatively large benefits due to lack of inter-operator radio resource allocation. Though if multihoming user fraction is small then benefits are large.	• Simple stochastic process that can simulate longitudinal rank-frequency distributions. Such empirical rank-frequency distributions are also found in mobile app concentration. • Analysis illustrating that many mobile app usage and concentration distributions can be modeled by common long tail distributions including log-normal and stretched exponential. • Analysis showing agent-based modeling can be profitably applied to study user experience of end user network switching and multihoming.

6. Discussion

This chapter discusses the research implications including implications for specific ecosystem players, the generalizability of the research results, and paths forward in future work.

6.1 Research implications

Overall, due to the diverse and novel nature of the three primary research questions (RQ 1-3), the results span many contexts and the main research benefit is to provide initial data points and methods that allow for further research of these contexts. This objective contrasts with research in well-established areas where the value is in providing a definitive conclusion to the given questions. In other words, this work helps build the basis or foundation rather than putting in the final keystone.

An example of such a foundational result is the development of a multidevice session definition and construction framework based on Allen's Interval Algebra. This framework helps bring formalization to the construction and definition of multidevice sessions.

Though additionally, the results extend several current research results or observations to these new contexts. For example, the multidevice session results expand the previous observation of significant usage diversity between users (i.e. [33]) to the multidevice session context. Similarly, the high concentration of social networking apps on tablets expands the previous observation of high concentration of social network apps on smartphones [38]. Therefore, illustrating potential network effects on both device types.

6.2 Practical implications for ecosystem players

Beyond the pure research implications, the results also have practical implications for several mobile ecosystem players. These implications are briefly summarized.

6.2.1 Mobile network operators

The implications of the results for mobile network operators are several-fold with the main results deriving from the multiple mobile networks and long-term mobile service research.

Specifically, the multiple mobile network results indicate that, given the potential benefits of end network switching, national regulators might pressure operators to help enable such switching or at least help support an equally beneficial alternative mechanism. For example, regulators could push operators to support the addition of automated SIM profile switching to the current E-SIM functionality. Therefore, operators should have a strategy for dealing with such pressure and with the future implications of such network switching mechanisms. Such a strategy should detail, for example, how future investment decisions would change given such mechanisms.

Additionally, the illustrated dynamics of end-user multihoming suggests that operators might need inter-operator coordination to provide efficient scheduling of such multihoming users. In fact, the lack of such coordination might even result in lower QoE for non-multihoming users. Hence understanding the dynamics of such coordination is important.

Whereas, the long-term mobile service research suggests that mobile network operators need to consider several temporal effects to understand a user's evaluation of their mobile network experience. For example, given the observed peak-end effect operators may want to optimize networks to avoid very poor service for any given user even at the expense of slightly better average service levels. Additionally, in the case of users that measure their own network performance, factors such as upload speed and latency do not seem to affect their evaluations, potentially because users do not normally consider these factors or these factors are already good enough for most users. Therefore, operators could primarily concentrate their focus on meeting users download throughput expectations.

6.2.2 Mobile app developers

The implications of the results for mobile app developers primarily derive from the multiple mobile device research.

Specifically, the ubiquity of multidevice mobile sessions suggests that app developers and designers should already be thinking in terms of their app sessions spanning multiple devices types. Towards this end app developers could utilize existing multidevice interaction frameworks such as the 4C framework [78] to help in such multidevice thinking. Therefore, these results further emphasize the importance of adopting such tools. Additionally, the diversity within multidevice sessions indicates that adaptation and customization of multidevice interaction for mobile apps is also important, especially given the increasing number of different and specialized device types (for example smartwatches [66]).

6.2.3 Mobile content providers and advertisers

The implications of the results for mobile content providers and advertisers also primarily derive from the multiple mobile device research. For example, mobile advertisers primarily rely on the number of impressions as a measure of ad inventory. Given that the ads are typically loaded at app start-up and thereafter refreshed at a fixed interval (often 30 seconds), the number of sessions and amount of time spent on different apps and device types informs the division of total ad inventory between these apps and device types.

6.3 Generalizability

A major part of almost all scientific studies in understanding the limitations of the study with regard to generalizability. In other words, in what other contexts or populations with the study results hold? This is an important concern in all studies of this thesis but especially for the studies based mainly on empirical data (Pubs. II, III, and IV).

A common strategy for improving the generalizability of a given study is to use a sample that is somehow representation of the larger population for which the results should hold. In terms of publications that used dataset 1, the analyzed subsets of dataset 1 were groups of US based users that, though not completely representative, did roughly match the US smartphone user population on many demographic parameters. Table 6.1 details the demographic of several subsets compared to the demographics of US smartphone users from a representative survey by Pew Research [62] performed only a few months before the collection of Dataset 1. This type of comparison allows the limitation in terms of generalizability to be more easily understood. For example, the comparison illustrates that the subsets favored female users with lower incomes and education levels compared to US smartphone users in general. Therefore, generalizations to groups with significantly different demographic profiles should be done with caution.

Table 6.1. Demographic comparison of several dataset 1 subsets with representative sample of US smartphone users. Adapted from Pubs. II, III.

Demographic	Subset 1	Subset 4B	Subset 5	US S Users[a]
Mean Age (Years)[b]	37.08 (12.54)	38.53 (11.91)	37.06 (12.63)	41.30 (15.08)
Gender (% Male)	26.42	23.08	26.61	50.08
Education (% Some College or Less)	62.18	56.92	63.10	53.04
Marital Status (% Married)	41.45	44.62	41.33	48.96
Household Income (% <50K USD)	64.08	47.69	65.93	40.72
Mean Household Size	2.96 (1.51)	2.97 (1.54)	2.95 (1.50)	3.05 (1.61)
Mean Children in Household	0.91 (1.19)	0.94 (1.06)	0.93 (1.21)	0.74 (1.27)
Race (% White)	70.98	72.31	70.97	71.53

[a] US smartphone user demographic data is from Pew Research survey (June-July 2015, sub-population with smartphones n=1327) [62]. The survey utilizes weighting to population parameters of census data to create nationally representative results (see [64]). We note that Verto Analytics also performs its own national surveys, we utilize the Pew Research survey only for brevity.

[b] All mean values also include standard deviations.

In terms of Pub. IV, dataset 2 was a crowdsourced dataset from the freely available Netradar app[1]. Therefore, collecting demographic data was more difficult due to the lack of financial incentives for users to give such data, and even asking for such data might raise privacy concerns for users. Therefore, instead of demographic data, technographic data that can be automatically collected can act as a proxy. Table 6.2 details the technographics of the 13 most popular smartphones in dataset 2 compared to the 13 most popular smartphones from a census of smartphones on Finnish mobile networks [88]. The comparison illustrates that the users in dataset 2 were likely to be more technologically advanced users with higher end devices than Finnish mobile network users in general. Therefore, generalizations based on the results should consider that results might be different for less advanced users.

Overall, the lack of complete representativeness is less important in a more general theory building context. In other words, as previously stated, these works represent only single data points of a specific sample at a particular point in time. Therefore, as both [13] and [5] point out, the value of these studies

[1] https://www.netradar.org/

Table 6.2. Techographic comparison of top 13 device models of dataset 2 and top 13 device models of Finnish mobile network users. Adapted from Pub. IV.

Technographic	**Dataset 2**	**Finnish Users**
Mean Screen Size (inches)	4.70	4.19
Mean Pixel Density (PPI)	346.00	280.85
Mean Total Freq (GHz)	4.80	2.92
Mean GPU Perf (GFLOPs)	71.61	34.74

primarily lie in being points in a general theory building effort that includes meta-analysis and comparisons over time and between different samples. This was also emphasized in the Theory-creating research approach of [36] discussed in Section 1.4.

In fact, in many circumstances even replication/reproduction[2] of a given study can be valuable when the results are used for theory building in human behavior. The reason is that even in cases where behavior is thought to be well established there has been difficulty in reproducing/replicating many studies. For example, [57] reperformed 100 previous psychological studies and found that in less than half of the studies were the results reproducible (specifically, similar significances and effect sizes). Similar results might be found in related areas such as cognitive neuroscience [79].

6.4 Future research

The future research of the studied contexts can follow several tracks again depending on the specific research community performing the research.

For multiple mobile device studies, future work in the HCI community should focus on multiple methods studies such as combining device based measurements and device based pop-up surveys. Such a combination can allow for the benefits of both objective quantitative usage data along with contextual subjective survey data. This combination is important since user reported statistics of, for example, mobile app usage have been found to be inaccurate [21]. Towards this goal, the recent introduction of free and open source frameworks (such as Aware [26]) that allow such studies can significantly reduce study costs.

Whereas in the mobile network community, future work should follow a similar approach but with a stronger emphasis on the effect of mobile network usage on the user experience of the many device types. For example, given small device

[2] In a replication changes to the study design and implementation are not allowed whereas in a reproduction some changes are allowed [5].

types like smartwatches understanding the tradeoff between the frequency of network connectivity and battery life in terms of total QoE is important. These kinds of tradeoffs only become more complex as users acquire multiple different interacting device types. Maybe a user has adequate QoE if one of their multiple devices still has battery life; therefore such tradeoffs can span multiple devices types.

For multiple mobile network studies, future work could consider the mechanisms of end user network switching and multihoming from a business viewpoint. For example, studies could analyze the potential types of pricing schemes for consumers with end user network switching and multihoming. Users might have a primary contract with a given operator that provides unlimited data and a cheaper secondary backup contract with another operator that provides only 1GB of data. Therefore, switching or multihoming algorithms would need to take these contractual constraints into account. Similarly, the potential effects of end user mechanisms on operator behavior in terms of investment incentive and competition are largely unexplored.

In terms of long-term mobile services, future work should just concentrate on identifying and replicating even basic temporal phenomena given that current research is so sparse. Though, as previously mentioned, performing QoE studies at timescales of years will remain challenging due to issues with retention of users in such long studies. Perhaps emerging user retention strategies such as community engagement and gamification may help in retention. For example, the mobile energy saving recommendation app Carat[3] (part of a project by UC Berkeley and University of Helsinki) includes a ranking of each user's energy savings relative to the all other users thus creating a competition aspect for the users. A survey of Carat users shows that longer term users are especially interested in this user ranking [3].

Overall, as emphasized in Sections 2.2 and 6.1, there is significant room for additional research in all three of the studied contexts.

[3]http://carat.cs.berkeley.edu/

References

[1] J. Alstott, E. Bullmore, and D. Plenz. powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE*, 9(1):e85777, 01 2014.

[2] I. Andone, K. Blaszkiewicz, M. Eibes, B. Trendafilov, C. Montag, and A. Markowetz. Menthal: A framework for mobile data collection and analysis. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, pages 624–629, 2016.

[3] K. Athukorala, E. Lagerspetz, M. von Kügelgen, A. Jylhä, A.J. Oliner, S. Tarkoma, and G. Jacucci. How carat affects user behavior: Implications for mobile battery awareness applications. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, pages 1029–1038, 2014.

[4] R.P. Baker, M. Callegaro, A.S. Goritz, and J. Bethlehem. The untold story of multimode (online and mail) consumer panels. In M. Callegaro, R. Baker, J. Bethlehem, A.S. Göritz, J.A. Krosnick, and P.J. Lavrakas, editors, *Online Panel Research*, pages 104–126. Wiley, 2014.

[5] N. Banovic. To replicate or not to replicate? *GetMobile: Mobile Comp. and Comm.*, 19(4):23–27, mar 2016.

[6] T. Becker, T.C. Corcoran, A. Greaves-Tunnell, J.R. Iafrate, J. Jing, S.J. Miller, J.D. Porfilio, R. Ronan, J. Samranvedhya, and F.W. Strauch. Benford's law and continuous dependent random variables, 2013.

[7] J. Bergstra. Audioscrobbler data. `http://www-etud.iro.umontreal.ca/~bergstrj/audioscrobbler_data.html`.

[8] J. Björkdahl and E. Bohlin. 3g network investments in sweden, 2003. Swedish National Post and Telecom Agency.

[9] A.G. Blom, M. Bosnjak, A. Cornilleau, A. Cousteaux, M. Das, S. Douhou, and U. Krieger. A comparison of four probability-based online and mixed-mode panels in europe. *Social Science Computer Review*, 34(1):8–25, 2016.

[10] Box Office Mojo. Yearly box office imdb. `http://boxofficemojo.com/yearly/`.

[11] CentERdata. Composition and response. `https://www.lissdata.nl/lissdata/about-panel/composition-and-response`.

[12] R. Chandra, P. Bahl, and P. Bahl. Multinet: connecting to multiple ieee 802.11 networks using a single wireless card. In *IEEE INFOCOM 2004*, volume 2, pages 882–893 vol.2, March 2004.

[13] K. Church, D. Ferreira, N. Banovic, and K. Lyons. Understanding the challenges of mobile phone usage data. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pages 504–514, New York, NY, USA, 2015. ACM.

[14] Cisco. Cisco vni global mobile data traffic forecast, 2015-2020. `http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/vni-inforgraphic.html`.

[15] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[16] S. Consolvo, F.R. Bentley, E.B. Hekler, S.S. Phatak, and M. Satyanarayanan. *Mobile User Research: A Practical Guide*. Synthesis Lectures on Mobile and Pervasive Computing. Morgan & Claypool Publishers, 2017.

[17] M.P. Couper, C. Antoun, and A. Mavletova. *Mobile Web Surveys*, pages 133–154. John Wiley and Sons, Inc., 2017.

[18] F.A. Cowell and E. Flachaire. Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141(2):1044–1072, 2007.

[19] G. Dandachi, S.E. Elayoubi, T. Chahed, and N.C. Taher. Performance evaluation of user centric multihoming strategies in lte/wifi networks. In *2016 IEEE Wireless Communications and Networking Conference*, pages 1–6, April 2016.

[20] A.K. Das, P.H. Pathak, C. Chuah, and P. Mohapatra. Characterization of wireless multidevice users. *ACM Trans. Internet Technol.*, 16(4):29:1–29:25, dec 2016.

[21] M. De Reuver, H. Bowman, N. Heerschap, and H. Verkasalo. Smartphone measurement: Do people use mobile applications as they say they do? In *International Conference on Mobile Business*, Atlanta, GA, USA, 2012. Association for Information Systems.

[22] D. Dearman and J.S. Pierce. It's on my other computer!: Computing with multiple devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 767–776, 2008.

[23] S. Deng, A. Sivaraman, and H. Balakrishnan. Delphi: A software controller for mobile network selection. `http://web.mit.edu/anirudh/www/delphi-tr.pdf`, 2016.

[24] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin. Diversity in smartphone usage. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, pages 179–194, New York, NY, USA, 2010. ACM.

[25] W. Fan and Z. Yan. Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26(2):132–139, 2010.

[26] D. Ferreira, V. Kostakos, and A.K. Dey. Aware: Mobile context instrumentation framework. *Frontiers in ICT*, 2:6, 2015.

[27] O. Fontanelli, P. Miramontes, Y. Yang, G. Cocho, and W. Li. Beyond zipf's law: The lavalette rank function and its properties. *PLOS ONE*, 11(9):1–14, 09 2016.

[28] X. Geng, Z. Chen, W. Lam, and Q. Zheng. Hedonic evaluation over short and long retention intervals: The mechanism of the peak-end rule. *Journal of Behavioral Decision Making*, 26(3):225–236, 2013.

[29] Google. The new multi-screen world: Understanding cross-platform consumer behavior. `https://think.withgoogle.com/databoard/media/pdfs/the-new-multi-screen-world-study_research-studies.pdf`, 2012.

[30] D. Guse. *Multi-episodic perceived quality of telecommunication services*. PhD thesis, Technischen Universität Berlin, 2016.

[31] K. Habak, K.A. Harrasb, and M. Youssef. Bandwidth aggregation techniques in heterogeneous multi-homed devices: A survey. *Computer Networks*, 92, Part 1:168 – 188, 2015.

[32] D. Helbing. *Social Self-Organization*, chapter Agent-Based Modeling, pages 25–70. Springer, 2012.

[33] D. Hintze, R.D. Findling, S. Scholz, and R. Mayrhofer. Mobile device usage characteristics: The effect of context and form factor on locked and unlocked usage. In *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia*, MoMM '14, pages 105–114. ACM, 2014.

[34] T. Hoßfeld, P.E. Heegaard, M. Varela, and S. Möller. Qoe beyond the mos: an in-depth look at qoe via better metrics and their relation to mos. *Quality and User Experience*, 1(1):2, Sep 2016.

[35] T. Hoßfeld, C. Keimel, and C. Timmerer. Crowdsourcing quality-of-experience assessments. *Computer*, 47(9):98–102, 2014.

[36] P. Järvinen and A. Järvinen. *On research methods*. Opinpajan Kirja, 2004.

[37] T. Jokela, J. Ojala, and T. Olsson. A diary study on combining multiple information devices in everyday activities and tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3903–3912, New York, NY, USA, 2015. ACM.

[38] J. Jung, Y. Kim, and S. Chan-Olmsted. Measuring usage concentration of smartphone applications: Selective repertoire in a marketplace of choices. *Mobile Media & Communication*, 2(3):352–368, 2014.

[39] D. Kahneman. Objective happiness. In M. Kubovy, D. Kahneman, E. Diener, and N. Schwarz, editors, *Well-being: The Foundations of Hedonic Psychology*, pages 3–25. Russel Sage, 1999.

[40] S. Kandula, K.C. Lin, T. Badirkhanli, and D. Katabi. Fatvap: Aggregating ap backhaul capacity to maximize throughput. In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, NSDI'08, pages 89–104. USENIX Association, 2008.

[41] Y.F. Kuo, C.M. Wu, and W.J. Deng. The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services. *Computers in Human Behavior*, 25(4):887 – 896, 2009.

[42] J. Laherrère and D. Sornette. Stretched exponential distributions in nature and economy: fat tails with characteristic scales. *The European Physical Journal B - Condensed Matter and Complex Systems*, 2(4):525–539, 1998.

[43] P. Le Callet, S. Möller, and A. Perkis. Qualinet white paper on definitions of quality of experience. Technical report, European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, 03 2013.

[44] Y. Li, H. Deng, C. Peng, Z. Yuan, G.H. Tu, J. Li, and S. Li. icellular: Device-customized cellular network access on commodity smartphones. In *Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation*, NSDI'16, pages 643–656, 2016.

[45] X. Liu. *Applied Ordinal Logistic Regression Using Stata: From Single-Level to Multilevel Modeling*. SAGE Publications, Incorporated, 2015.

[46] M.C. Macal and J.M. North. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3):151–162, 2010.

[47] G. Martinez-Mekler, R.A. Martinez, M. Beltran del Rio, R. Mansilla, P. Miramontes, and G. Cocho. Universality of rank-ordering distributions in the arts and sciences. *PLOS ONE*, 4(3):1–7, 03 2009.

[48] T. Matthews, J. Pierce, and J. Tang. No smartphone is an island: The impact of places, situations, and other devices on smartphone use, 2009. IBM Research Report RJ10452.

[49] Microsoft. Cross-screen engagement. `http://advertising.microsoft.com/es-xl/WWDocs/User/display/cl/researchreport/1932/global/Cross_ScreenWhitepaper.pdf`, 2013.

[50] C. Moldovan and F. Metzger. Bridging the gap between qoe and user engagement in http video streaming. In *2016 28th International Teletraffic Congress (ITC 28)*, volume 01, pages 103–111, Sept 2016.

[51] G.D. Montanez, R.W. White, and X. Huang. Cross-device search. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 1669–1678, New York, NY, USA, 2014. ACM.

[52] H. Müller, J.L. Gove, J.S. Webb, and A. Cheang. Understanding and comparing smartphone and tablet use: Insights from a large-scale diary study. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, pages 427–436, 2015.

[53] D. Naboulsi, D. Fiore, S. Ribot, and R. Stanica. Large-scale mobile traffic analysis: A survey. *IEEE Communications Surveys Tutorials*, 18(1):124–161, 2016.

[54] G.G. Naumis and G. Cocho. Tail universalities in rank distributions as an algebraic problem: The beta-like function. *Physica A: Statistical Mechanics and its Applications*, 387(1):84–96, 2008.

[55] A. Nika, Y. Zhu, N. Ding, A. Jindal, Y.C. Hu, X. Zhou, B.Y. Zhao, and H. Zheng. Energy and performance of smartphone radio bundling in outdoor environments. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 809–819. International World Wide Web Conferences Steering Committee, 2015.

[56] T. Oltman. Attrition in mobile ratings panels. `http://www.mapor.org/confdocs/absandpaps/2014/1C2_Oltman_slides.pdf`, 2014.

[57] Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*, 349(6251), 2015.

[58] A. Oulasvirta, A. Reichel, W. Li, Y. Zhang, M. Bachynskyi, K. Vertanen, and P.O. Kristensson. Improving two-thumb text entry on touchscreen devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2765–2774, New York, NY, USA, 2013. ACM.

[59] A. Oulasvirta and L. Sumari. Mobile kits and laptop trays: Managing multiple devices in mobile information work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1127–1136, 2007.

[60] A. Özer, M.T. Argan, and M. Argan. The effect of mobile service quality dimensions on customer satisfaction. *Procedia - Social and Behavioral Sciences*, 99:428 – 438, 2013. The Proceedings of 9th International Strategic Management Conference.

[61] A. Papoulis and Pillai S. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Series in Electrical and Computer Engineering. McGraw-Hill, 4th edition, 2002.

[62] Pew Internet and American Life Project. June 10-july 12, 2015 – gaming, jobs and broadband. `http://www.pewinternet.org/datasets/june-10-july-12-2015-gaming-jobs-and-broadband/`, 2015.

[63] Pew Research Center. Mobile fact sheet. `http://www.pewinternet.org/fact-sheet/mobile/`.

[64] Pew Research Center. Our survey methodology in detail. `http://www.pewresearch.org/methodology/u-s-survey-research/our-survey-methodology-in-detail/`, 2016.

[65] B. Phipson and G.K. Smyth. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Stat Appl Genet Mol Biol*, 9(1), 2010.

[66] S. Pizza, B. Brown, D. McMillan, and A. Lampinen. Smartwatch in vivo. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5456–5469. ACM, 2016.

[67] A. Qureshi and J. Guttag. Horde: Separating network striping policy from mechanism. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, MobiSys '05, pages 121–134, 2005.

[68] A. Raake and S. Egger. Quality and quality of experience. In S. Möller and A. Raake, editors, *Quality of Experience: Advanced Concepts, Applications and Methods*, T-Labs Series in Telecommunication Services, pages 11–34. Springer International Publishing, 2014.

[69] RAND Corporation. Completion rates and attrition. `http://www.rand.org/labor/alp/panel/completion-rates.html`.

[70] Redhat. Mobile maturity survey. Technical report, Redhat, 2016.

[71] Peltier S. and Moreau F. Looking for the long tail: Evidence from the french book market. In *Proceedings of 16th ACEI International Conference*, 2010.

[72] S. Santosa and D. Wigdor. A field study of multi-device workflows in distributed workspaces. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 63–72, 2013.

[73] M.Z. Shafiq, J. Erman, L. Ji, A.X. Liu, J. Pang, and J. Wang. Understanding the impact of network dynamics on mobile video user engagement. In *The 2014 ACM International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '14, pages 367–379. ACM, 2014.

[74] M.Z. Shafiq, L. Ji, A.X. Liu, J. Pang, and J. Wang. Geospatial and temporal dynamics of application usage in cellular data networks. *IEEE Transactions on Mobile Computing*, 14(7):1369–1381, July 2015.

[75] B.M.C. Silva, J.J.P.C. Rodrigues, I. de la Torre Díez, M. López-Coronado, and K. Saleem. Mobile-health: A review of current state in 2015. *Journal of Biomedical Informatics*, 56:265–272, 2015.

[76] T. Soikkeli. *End user context in analyzing mobile device and service usage*. PhD thesis, Aalto University, 2016.

[77] T. Soikkeli, J. Karikoski, and H. Hämmäinen. Characterizing smartphone usage: Diversity and end user context. *International Journal of Handheld Computing Research*, 4(1):15–36, 2013.

[78] H. Sørensen, D. Raptis, J. Kjeldskov, and M.B. Skov. The 4c framework: Principles of interaction in digital ecosystems. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pages 87–97. ACM, 2014.

[79] D. Szucs and J.P.A. Ioannidis. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, 15(3):1–18, 03 2017.

[80] The Association of Magazine Media. Circulation trends and data. `http://www.magazine.org/insights-resources/research-publications/trends-data/magazine-industry-facts-data/circulation-trends`.

[81] N. Thompson, G. He, and H. Luo. Flow scheduling for end-host multihoming. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pages 1–12, April 2006.

[82] E. Tulving. Episodic and semantic memory. In E. Tulving and W. Donaldson, editors, *Organization of Memory*, pages 381—-402. Academic Press, 1972.

[83] O Turel and A. Serenko. Satisfaction with mobile services in canada: An empirical investigation. *Telecommunications Policy*, 30(5-6):314 – 331, 2006.

[84] University of Massachusetts Amherst. Umass trace repository. `http://traces.cs.umass.edu/index.php/Network/Network`.

[85] US Census Bureau. Genealogy data: Frequently occurring surnames from census 2000. `http://www.census.gov/genealogy/www/data/2000surnames/names.zip`.

[86] M. Varela, L. Skorin-Kapov, and T. Ebrahimi. Quality of service versus quality of experience. In S. Möller and A. Raake, editors, *Quality of Experience: Advanced Concepts, Applications and Methods*, T-Labs Series in Telecommunication Services, pages 85–96. Springer International Publishing, 2014.

[87] H. Verkasalo. Metrics that matter: Two new approaches from huawei connect europe. http://www.vertoanalytics.com/metrics-that-matter-two-new-approaches-from-huawei-connect-europe/.

[88] A. Vesselkov and H. Hämmäinen. Mobile handset population in finland 2005-2015. Technical report, Aalto University, Department of Communications and Networking, 01 2016.

[89] A. Visuri, Z. Sarsenbayeva, N. van Berkel, J. Goncalves, R. Rawassizadeh, V. Kostakos, and D. Ferreira. Quantifying sources and types of smartwatch usage sessions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3569–3581, New York, NY, USA, 2017. ACM.

[90] Q.H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333, 1989.

[91] K. Wac, G. Pinar, M. Gustarini, and J. Marchanoff. Smartphone users mobile networks quality provision and volte intend: Six-months field study. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*, pages 1–9, June 2015.

[92] D.T. Wagner, A. Rice, and A.R. Beresford. Device analyzer: Understanding smartphone usage. In *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pages 195–208. Springer, 2014.

[93] Y. Wang, X. Huang, and R.W. White. Characterizing and supporting cross-device search tasks. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 707–716, New York, NY, USA, 2013. ACM.

[94] I. Wechsung and K. De Moor. Quality of experience versus user experience. In S. Möller and A. Raake, editors, *Quality of Experience: Advanced Concepts, Applications and Methods*, T-Labs Series in Telecommunication Services, pages 35–54. Springer International Publishing, 2014.

[95] R.R. Wilcox. *Introduction to Robust Estimation and Hypothesis Testing*. Elsevier, 3rd edition, 2012.

[96] R. Williams. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, 6(1):58–82(25), 2006.

Errata

Publication I

The PDF_{X_i} detailed in the section *PDF of a Single Interval over Independent FT Process Trials* was incorrect due to a coding error. The corrected function can be found below. Subsequently, the $PDF_{X_i^2}$ detailed in Supporting Text S1 was also incorrect (due to the same error). The corrected function can also be found below. The corrected $PDF_{X_i^2}$ is less complex but still a piecewise function with 7 unique sub-functions (note that some sub-functions and sub-domains are present in both parts of the function). The $PDF_{X_i^3}$ case is computationally tractable but again splits the result into an even larger number of piecewise functions. Thus, as mentioned, any analytical form is unlikely to be of practical use. Thanks to Dr. Colin Rose of the Theoretical Research Institute (Sydney) and MathStatica for pointing out this error.

$$PDF_{X_i} = \left\{ \begin{array}{cc} \frac{\alpha}{\alpha-1} & \left(0 < y < \frac{1-\alpha}{2}\right) \\ \frac{1}{\alpha} - 1 & \left(\frac{1-\alpha}{2} \leq y < \frac{1+\alpha}{2}\right) \\ \frac{\alpha}{\alpha-1} & \left(\frac{1+\alpha}{2} \leq y < 1\right) \end{array} \right\}$$

For $(-2+\sqrt{5}) < \alpha < 1$

$$PDF_{x_i{}^2} = \left\{ \begin{array}{cc} \frac{(4-8\alpha)ArcTanh[\alpha]-\alpha^2\log[y]}{(-1+\alpha)^2} & \left(0 < y < \frac{1}{4}(\alpha-1)^2\right) \\ \frac{1}{(-1+\alpha)^2\alpha^2}\left((-1+\alpha)^4\log\left[\frac{4y}{(-1+\alpha)^2}\right]+2\alpha^4\log\left[\frac{2}{1+\alpha}\right]-2(-1+\alpha)^2\alpha^2\log\left[-\frac{4y}{-1+\alpha^2}\right]\right) & \left(\frac{1}{4}(\alpha-1)^2 < y < \frac{1}{4}\left(1-\alpha^2\right)\right) \\ \frac{1}{(-1+\alpha)^2\alpha^2}\left(-2\alpha^4\log\left[-\frac{2y}{-1+\alpha}\right]-(-1+\alpha)^4(\log[4]+\log[y]-2\log[1+\alpha])+2(-1+\alpha)^2\alpha^2\log\left[-\frac{4y}{-1+\alpha^2}\right]\right) & \left(\frac{1}{4}\left(1-\alpha^2\right) < y < \frac{1-\alpha}{2}\right) \\ \frac{-\log[4]+\alpha\log[16]-(-1+\alpha)^2\log[y]+(2-4\alpha)\log[1+\alpha]}{\alpha^2} & \left(\frac{1-\alpha}{2} < y < \frac{1}{4}(\alpha+1)^2\right) \\ \frac{\alpha^2\log\left[\frac{4y}{(1+\alpha)^2}\right]}{(-1+\alpha)^2}-2\log\left[\frac{2y}{1+\alpha}\right] & \left(\frac{1}{4}(\alpha+1)^2 < y < \frac{1+\alpha}{2}\right) \\ -\frac{\alpha^2\log[y]}{(-1+\alpha)^2} & \left(\frac{1+\alpha}{2} < y < 1\right) \end{array} \right\}$$

For $0 < \alpha < (-2+\sqrt{5})$

$$PDF_{x_i{}^2} = \left\{ \begin{array}{cc} \frac{(4-8\alpha)ArcTanh[\alpha]-\alpha^2\log[y]}{(-1+\alpha)^2} & \left(0 < y < \frac{1}{4}(\alpha-1)^2\right) \\ \frac{1}{(-1+\alpha)^2\alpha^2}\left((-1+\alpha)^4\log\left[\frac{4y}{(-1+\alpha)^2}\right]+2\alpha^4\log\left[\frac{2}{1+\alpha}\right]-2(-1+\alpha)^2\alpha^2\log\left[-\frac{4y}{-1+\alpha^2}\right]\right) & \left(\frac{1}{4}(\alpha-1)^2 < y < \frac{1}{4}\left(1-\alpha^2\right)\right) \\ \frac{1}{(-1+\alpha)^2\alpha^2}\left(-2\alpha^4\log\left[-\frac{2y}{-1+\alpha}\right]-(-1+\alpha)^4(\log[4]+\log[y]-2\log[1+\alpha])+2(-1+\alpha)^2\alpha^2\log\left[-\frac{4y}{-1+\alpha^2}\right]\right) & \left(\frac{1}{4}\left(1-\alpha^2\right) < y < \frac{1}{4}(\alpha+1)^2\right) \\ \frac{-\alpha(8ArcTanh[\alpha]+\alpha\log[y])+\log\left[\frac{1}{(-1+\alpha)^2}\right]+2\log[1+\alpha]}{(-1+\alpha)^2} & \left(\frac{1}{4}(\alpha+1)^2 < y < \frac{1-\alpha}{2}\right) \\ \frac{\alpha^2\log\left[\frac{4y}{(1+\alpha)^2}\right]}{(-1+\alpha)^2}-2\log\left[\frac{2y}{1+\alpha}\right] & \left(\frac{1-\alpha}{2} < y < \frac{1+\alpha}{2}\right) \\ -\frac{\alpha^2\log[y]}{(-1+\alpha)^2} & \left(\frac{1+\alpha}{2} < y < 1\right) \end{array} \right\}$$

Publication II

For the permutation test of section 5.2 the result is $0.00011 \leq p$ rather than $0.0001 \leq p$ due to the use of an approximate permutation test rather than an exact permutation test. This does not change at all the conclusion of the test given the extremely high significance in either case.

Aalto-DD 173/2017

ISBN 978-952-60-7609-6 (printed)
ISBN 978-952-60-7608-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Communications and Networking
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS