MARMARA UNIVERSITY
INSTITUTE FOR GRADUATE STUDIES
IN PURE AND APPLIED SCIENCES

ENERGY AND DATA RATE MODELING IN USER PROVIDED NETWORKS WITH INCENTIVE MECHANISMS

KÜBRA ULUDAĞ

MASTER THESIS
Department of Computer Engineering

ADVISOR
Assist. Prof. Ömer Korçak

ISTANBUL, 2018

MARMARA UNIVERSITY
INSTITUTE FOR GRADUATE STUDIES
IN PURE AND APPLIED SCIENCES

ENERGY AND DATA RATE MODELING IN USER PROVIDED NETWORKS WITH INCENTIVE MECHANISMS

KÜBRA ULUDAĞ

(524115003)

MASTER THESIS

Department of Computer Engineering

ADVISOR

Assist. Prof. Ömer Korçak

ISTANBUL, 2018

MARMARA UNIVERSITY

INSTITUTE FOR GRADUATE STUDIES IN PURE AND APPLIED SCIENCES

Kübra ULUDAĞ, a Master of Science student of Marmara University Institute for Graduate Studies in Pure and Applied Sciences, defended her thesis entitled "**Energy and Data Rate Modeling in User Provided Networks with Incentive Mechanisms**", on June 21, 2018 and has been found to be satisfactory by the jury members.

Jury Members

Assist. Prof. Ömer KORÇAK (Advisor)
Marmara University ..... (SIGN).....

Prof. Dr. Çiğdem EROĞLU ERDEM (Jury Member)
Marmara University ..... (SIGN).....

Assoc.Prof. Didem GÖZÜPEK (Jury Member)
Gebze Technical University ..... (SIGN).....

APPROVAL

Marmara University Institute for Graduate Studies in Pure and Applied Sciences Executive Committee approves that Kübra ULUDAĞ be granted the degree of Master of Science in Department of Computer Engineering, Computer Engineering Program on 25.06.2018 (Resolution no: 2018/15-03)

Director of the Institute
Prof. Dr. Bülent EKİCİ

ACKNOWLEDGMENT

I would like to express my gratitude to my thesis supervisor, Assist. Prof. Ömer Korçak, for his guidance, support and encouragement throughout my graduate study and completion of this thesis. I would also thank to Prof. Dr. Çiğdem Eroğlu Erdem and Assoc. Prof. Didem Gözüpek for participating my thesis committee and their useful comments.

I would like to Ahmet Özlü, Berivan Şahin, Süleyman Eser, Ömer Faruk Mercan and Saadeddin Porhan for their helps during the measurements.I would like to thank Zuhal Öztürk, Serap Korkmaz, Gizem Süngü and Esra Nur Varlı for their friendship and support.

I gratefully acknowledge the financial support of The Scientific and Technological Research Council of Turkey (TUBITAK) through a research grant (Project Number 115E389).

Finally, I would like to thank my family for their patience, encouragement and support during my whole life.

June, 2018 **Kübra ULUDAĞ**

TABLE OF CONTENTS

ACKNOWLEDGMENT i
TABLE OF CONTENTS ii
ÖZET iv
ABSTRACT v
SYMBOLS vi
ABBREVIATIONS ix
LIST OF FIGURES x
LIST OF TABLES xii
1. INTRODUCTION 1
2. USER PROVIDED NETWORK SYSTEM WITH INCENTIVE MECHANISM 8
2.1. Network Model 8
2.2. System Architecture 9
2.2.1. Data collection module 9
2.2.2. Training module 10
2.2.3. Bargaining module 11
2.2.4. Protocol operation 18
3. ENERGY AND DATA RATE MODELING 21
3.1. Experimental Study 21
3.1.1. Tool and environment 21
3.1.2. Idle running experiments 25
3.1.3. Data download experiments 26
3.1.3.1. Measured signal metrics 27
3.1.3.2. Standalone experiments 28
3.1.3.3. Standalone versus Gateway experiments 37
3.1.3.4. Standalone versus Client experiments 41
3.2. Regression Analysis 44
3.2.1. Linear regression assumptions 44
3.2.2. Applied regression techniques 46
3.2.3. Evaluation criteria and standardization 47
3.2.4. Data preprocessing 48
3.2.5. Modeling 51
3.2.5.1. Models for standalone case 51

3.2.5.2. Models for gateway case 61
3.2.5.3. Results and discussion 66
3.3. Application for bargaining module 67
4. CONCLUSION AND FUTURE WORK 75
REFERENCES 77

ÖZET

TESVIK MEKANIZMALI KULLANICI SAĞLAYICILI AĞLARDA ENERJI VE VERI HIZI MODELLEMESI

Mobil veri trafiği, internet erişimine ihtiyaç duyan mobil cihazların çoğalması nedeniyle hızla artmaktadır, diğer taraftan, mobil cihazlar teknolojik gelişmelerle birlikte daha yetenekli hale gelmiştir. Bu durum, kullanıcıların İnternet bağlantılarını paylaşmasına dayanan Kullanıcı Sağlayıcılı Ağ kavramını ortaya çıkarmaktadır. Bu yaklaşıma göre, hücresel bağlantıya sahip kullanıcılar, sınırlı bağlantıya sahip olan veya hiç bağlantısı olmayan kişilerle kendi bağlantılarını paylaşırlar. Ancak, bu tür hizmetlerde, en önemli konulardan biri kullanıcıların katkıda bulunma istekliliğidir. Enerji tüketimi ve veri kullanım maliyetleri kritik öneme sahiptir ve Kullanıcı Sağlayıcılı Ağ hizmetlerini talep eden ve sunan kullanıcıların kararları üzerinde önemli etkileri vardır. Son zamanlarda bu amaç için bazı teşvik mekanizmaları önerilmiştir, fakat bunlar genellikle enerji maliyeti ve fayda modellemesi açısından gerçekçi parametrelere dayanmaz. Bu çalışmada, amacımız hem enerji-duyarlı hücresel veri çizelgeleme hem de işbirlikçi İnternet erişim uygulamaları için bazı gerçekçi enerji ve veri hızı modelleri sunmaktır. Bu amaç doğrultusunda, çeşitli deneyler yaptık ve çeşitli regresyon yöntemlerini kullanarak enerji ve veri hızı tahminleri için formüller sunduk. Elde edilen modeller, mevcut uygulamalarda kullanılabilmelerinin yanı sıra yaklaşan 5G ağları ve sonrası bağlamındaki yeni çalışmaların enerji verimliliğini karşılaştırmak için bir referans noktası oluşturacaktır.

June, 2018 **Kübra ULUDAĞ**

ABSTRACT

ENERGY AND DATA RATE MODELING IN USER PROVIDED NETWORKS WITH INCENTIVE MECHANISMS

Mobile data traffic is rapidly growing due to the proliferation of the mobile devices that need the Internet access, on the other hand, the mobile devices have become more capable together with technological improvements. This reveals the concept of User Provider Network which is based on sharing the users' Internet connections. According to this approach, users with cellular connection share their own connections with others with limited or no cellular connection. However, in such services, one of the key issues is the users' willingness to contribute. Energy consumption and data usage costs are of critical importance and have a large impact on users' decisions for both requesting and offering UPN services. Recently, some incentive mechanisms are proposed for this purpose but they are usually not based on realistic parameters in terms of energy cost and utility modeling. In this work, our goal is to introduce some realistic energy and data rate models for both energy-aware cellular data scheduling and collaborative Internet access applications. In accordance with this purpose, we performed various experiments and provided several formulas for energy and throughput estimations by applying various regression methods. Resulting models could be utilized in current applications as well as they would provide a benchmark to compare energy efficiency of new studies in the context of upcoming 5G networks and beyond.

June, 2018 **Kübra ULUDAĞ**

SYMBOLS

$\boldsymbol{\alpha_i}$: Sensitivity parameter for data rate of node i

$\boldsymbol{\beta_c}$: Sensitivity parameter for achieved data rate

$\boldsymbol{C_{con}}$: Energy cost of gateway caused by sharing its connection

$\boldsymbol{C_{ji}}$: Cost function of a gateway j serving for client i

$\boldsymbol{c_h}$: Counter for data below the minimum amount in the head state

$\boldsymbol{c_m}$: Counter for data below the minimum amount in the main state

$\boldsymbol{D_A}$: Amount of transferred data

$\boldsymbol{D_{avg}}$: Cumulative average amount of data transferred during the head state

$\boldsymbol{\delta_1}$: Discount factor for gateway g

$\boldsymbol{\delta_2}$: Discount factor for client c

$\boldsymbol{D_h}$: Data threshold for passing from the head state to the main state

$\boldsymbol{d_h}$: Minimum amount of data to be received/sent in head state

$\boldsymbol{D_{ji}}$: Transferred amount of data from gateway j to client i

$\boldsymbol{d_i}$: Data amount during a time slot i

$\boldsymbol{d_m}$: Minimum amount of data to be received/sent in the main state

$\boldsymbol{D_s}$: Maximum threshold value to stay in the silent s state

$\boldsymbol{D_t}$: Data threshold for passing from the tail state to the silent state

$\boldsymbol{E_g^{sta}}$: Standalone energy consumption function of a gateway g

$\boldsymbol{E_g^{oc}}$: Energy consumption function of a gateway while receiving from operator o and sending to client c

$\boldsymbol{E_i^{sta}}$: Standalone energy consumption function of node i

$\boldsymbol{e_{residual}}$: Percentage value of remaining battery energy of the gateway

$\boldsymbol{E_{go}^{r}}$: Energy consumption function of gateway g for receiving via operator o

E_{sta} : Standalone energy

IV^z : Standardized independent variable

L_{gc} : Link speed between gateway and client

N_r : Amount of bandwidth that gateway needs for itself

P_{ij} : Pie between client i and gateway j

θ_c : Sensitivity parameter for battery energy

R_0 : Minimum agreed data rate

R_1 : Maximum agreed data rate

R_A : Achieved data rate

R_{agr} : Agreed data rate

R_{gc}^{max} : Maximum average downlink data rate between client c and gateway g

R_g^{oc} : Received data rate of a gateway g from operator o while serving client c

R_g^{sta} : Standalone data rate of a gateway g

R_i^{min} : Minimum required data rate for client i

R_i^{sta} : Standalone data rate of a node i

R_{max} : Maximum provided data rate

R_{og}^{max} : Maximum average downlink data rate from operator o to a gateway g

R_{sta} : Standalone data rate

ρ : Normalization parameter for transferred data amount

T_{h1} : Time threshold for passing from the head state to the silent state

T_{h2} : Maximum waiting time in the head state

t_h : Time counter for the head state

t_i : i^{th} time slot

T_m : Time threshold for passing from the main state to the tail state

T_s : Time period

T_t : Time threshold for passing from the tail state to the silent state

t_t : Time counter for the tail state

U_{i0} : Standalone utility function (per-bit) of a client node i

U_{ij} : Utility function of a client i accessing Internet via a gateway j

V_{ij} : Virtual currency (per bit) transferring from client i to gateway j

W_{ss} : WiFi signal strength

ABBREVIATIONS

CI : Condition Index

CQI : Channel Quality Indicator

CV : Constant Variance

DV : Dependent Variable

HSPA+ : Evolved High Speed Packet Access

IoR : Independence of Residuals

IV : Independent Variable

LOWESS : Locally Weighted Scatterplot Smoothing

LTE : Long Term Evolution

MNO : Mobile Network Operator

MVNO : Mobile Virtual Network Operator

OLS : Ordinary Least Squares

QoS : Quality of service

RANSAC : RANdom SAmple Consensus

RSCP : Received Signal Code Power

RMSE : Root Mean Square Error

RSRP : Reference Signal Received Power

RSRQ : Reference Signal Received Quality

RSSI : Received Signal Strength Indicator

SINR : Signal to Interference plus Noise Ratio

TCP : Transmission Control Protocol

UMTS : Universal Mobile Telecommunication System

UPN : User Provided Network

WAN : Wide Area Network

WLAN : Wireless Local Area Network

WWAN : Wireless Wide Area Network

LIST OF FIGURES

Figure 2.1 Standalone and user provided connectivity scenarios 9
Figure 2.2 Proposed framework 9
Figure 2.3 Cost and Utility functions in terms of virtual currency per bit 14
Figure 2.4 Data transfer process state diagram 16
Figure 2.5 Protocol operation 20
Figure 3.1 Energy meter circuit scheme 23
Figure 3.2 The comparison of hardware and software measurement results 24
Figure 3.3 Possible data download scenarios 26
Figure 3.4 Average power and signal values while downloading 50 MB data file on HSPA+ network. Error bars represent mean and standard deviation of the power 28
Figure 3.5 Average energy and signal values while downloading 50 MB data file on HSPA+ network. Error bars represent mean and standard deviation of the energy. 29
Figure 3.6 Average data rate and signal values while downloading 50 MB data file on HSPA+ network. Error bars represent mean and standard deviation of the data rate. 30
Figure 3.7 Average power and signal values while downloading 50 MB data file on LTE network. Error bars represent mean and standard deviation of the power 31
Figure 3.8 Average energy and signal values while downloading 50 MB data file on LTE network. Error bars represent mean and standard deviation of the energy 32
Figure 3.9 Average data rate and signal values while downloading 50 MB data file on LTE network. Error bars represent mean and standard deviation of the data rate. 33
Figure 3.10 Relationship between power and LTE signal metrics 34
Figure 3.11 Relationship between energy and LTE signal metrics 35
Figure 3.12 Relationship between data rate and LTE signal metrics 36
Figure 3.13 Relationship of data rate to power and energy 37
Figure 3.14 The average values of LTE signal metrics during the tests for a standalone and a gateway node 38
Figure 3.15 The average values of WiFi signal strength and link speed according to the near and far client node 39
Figure 3.16 The average values of the data rate for a standalone and a gateway node 40
Figure 3.17 The average values of the energy consumption for a standalone and a gateway node 40
Figure 3.18 The average values of LTE signal metrics for the client node 41
Figure 3.19 The average values of WiFi signal strength and link speed 42
Figure 3.20 The average values of the data rate for a standalone and client node, and the average values of RSRP signal strength for the client and gateway nodes 43
Figure 3.21 The average values of the energy consumption for a standalone and a client node, and the average values of RSRP signal strength for the client and gateway nodes 43
Figure 3.22 Separation of head, main and tail parts of a data transfer period 49
Figure 3.23 Comparison of the separation techniques 50
Figure 3.24 Data distribution and relationships for the standalone case 53
Figure 3.25 Scatterplots of the standalone energy consumption models without data rate 55
Figure 3.26 Scatterplots of the standalone energy consumption models with data rate 57
Figure 3.27 Scatterplots of the standalone data rate models 59

Figure 3.28 Data distribution and relationships for the gateway case 63
Figure 3.29 Scatterplots of the gateway energy consumption and data rate models 64
Figure 3.30 Estimated and actual energy consumption values of the gateway 68
Figure 3.31 Estimated utility and cost functions with the received data rate of the gateway node while relaying the client 69
Figure 3.32 Illustration of 50 MB data download process according to the states 70
Figure 3.33 The comparison of the estimated and actual utility and cost values 71
Figure 3.34 The comparison of the estimated and actual virtual currency values 72
Figure 3.35 The estimated net profit of the nodes 73
Figure 3.36 The actual net profit of the nodes 73

LIST OF TABLES

Table 3.1 List of used devices....24
Table 3.2 Power consumption tests in idle state....26
Table 3.3 Descriptive statistics of data for the standalone case....51
Table 3.4 Correlations of data for the standalone case....52
Table 3.5 Model summary of the standalone energy consumption without data rate....54
Table 3.6 Model assumptions for the standalone energy consumption without data rate....54
Table 3.7 Model summary of the standalone energy consumption with data rate....56
Table 3.8 Model assumptions for the standalone the energy consumption with data rate....56
Table 3.9 Model summary of the standalone data rate....58
Table 3.10 Model assumptions for the standalone data rate....58
Table 3.11 Model summary of robust regression techniques for the standalone case....60
Table 3.12 Model summary of various regression techniques for the standalone case....60
Table 3.13 Descriptive statistics of data for the gateway case....61
Table 3.14 Correlations of data for the gateway case....62
Table 3.15 Model summary of the gateway energy consumption and data rate....64
Table 3.16 Model assumptions for the gateway energy consumption and data rate....64
Table 3.17 Model summary of robust regression techniques for the gateway case....65
Table 3.18 Model summary of various regression techniques for the gateway case....65
Table 3.19 LTE signal parameters of the nodes....68
Table 3.20 The values of the parameters used for separation of data transfer process into the states....70

1. INTRODUCTION

Global mobile data traffic is growing explosively such that it is expected to reach 49 exabytes by 2021 while it is 7.2 exabytes per month in 2016. Also, global mobile data traffic will grow twice as fast as fixed IP traffic from 2016 to 2021 [1]. In order to deal with this tremendous increase in mobile traffic load, mobile network operators (MNOs) need to make notable increase of capacity in their cellular networks. However, traditional capacity enhancement methods such as acquiring more spectrum licenses and expanding of network infrastructure are time-consuming and costly, and also they may not be always applicable due to limited resources such as spectrum. This leads MNOs to lower quality of service or higher pricing. Most of mobile network users are not able to receive cheap and/or high-capacity Internet services. It is obvious that innovative methods must be probed to solve this problem [2].

Mobile devices have become a part of our daily lives. As communication requirements increase, the number of mobile devices used is also increasing. By 2021, it is expected that there will be nearly 12 billion mobile-connected devices, approximately 1.5 per capita [1]. On the other hand, with recent technological advancements, hand-held user devices such as smartphones and tablets have very advanced features such as advanced network interfaces, advanced antenna arrays and high-end processors that can execute complicated networking tasks. However, these devices are mostly used as tools that only communicate with a base station and respond to the needs of their own owners, and their advanced features cannot be utilized sufficiently. Clearly, these devices can offer communication services to the nearby devices and act as micro-operators that serve each other according to their networking capabilities and demands. Therefore, it is considered that these enhanced capabilities of mobile devices can be exploited to address growing mobile traffic load. This gives rise to so-called User Provided Networks (UPN) [3]. In traditional infrastructure networks, users obtain network connectivity and services from network providers. The distinction between “provider” and “consumer” is obvious. However, in UPNs, users can be in both provider and consumer roles. In case users act as a provider, they provide direct connectivity to other users by sharing their available network connection. Thus, coverage and service of traditional providers are extended. UPNs have a great potential to reduce network congestion and network access costs. They

enable to improve user-perceived network performance by utilizing unused capabilities without requiring significant additional network infrastructure investment [3, 4].

Currently, there are several UPN models that differ in the architectures and services they offer to users. One of the first UPN solutions is FON [5] that is a centralized UPN model. It offers a wide WiFi sharing network with modems belonging to FON users. In the FON model, mobile users can access the Internet through residential WiFi APs of other users. Similarly, other FON users can also use the internet connections of FON home users while they are mobile. Lately, new UPN solutions that offer to use mobile devices as a mobile hotspot have emerged. These UPN models can be classified as autonomous services and network-assisted services. In autonomous services, mobile users connect directly to each other using their connections without intervention of network operators. On the other hand, in the network-assisted services the connectivity is provided by mobile network operators and also the operator may determine pricing rules and sharing policy. Open Garden [6] is an example of autonomous UPN services. It allows nearby mobile users to create a mesh network among them via WiFi or Bluetooth connections and it enables users in the network to access the Internet via WiFi or cellular connection. In this model, users can be in multiple roles such as a client node (consumer node), a relay node (bridge node between client and gateway) and a gateway node (node having direct WiFi/ cellular Internet connection). The usage of this kind of services can be restricted by some network operators because these services may exploit resources of operators without any payment. This has led to the introduction of network-assisted UPN solutions such as Karma [7] and M-87 [8].

In Karma, each subscriber can act as mobile WiFi hotspot via a portable device and share the Internet connection with nonsubscribers, in turn, she earns a free quota for every client she serves. Each subscriber pays a fixed price for the data she consumes. In M-87 solution, mobile devices are used as the hotspot and M-87 users can access the Internet over other users' devices, but each user uses her own quota although she accesses the internet via another user. Remote users can improve the quality of their connectivity by accessing the Internet via other users who have strong cellular connection. This is also profitable for base stations because serving a distant node is more costly in terms of signaling overhead and energy consumption. In products such as Open Garden and M87, sharing the connections of users is based on voluntariness. In this case, users may not be

willing to use UPN services or they only use such services among their relatives or friends, because sharing the resources of users brings a cost to them in terms of energy and bandwidth [9]. Therefore, some incentive mechanisms are needed to encourage the user to participate in the system.

Some incentive mechanisms have been proposed for providing widely and effectively usable UPN solutions [10]. Iosifidis et al. [9] propose an incentive mechanism for autonomous UPN services, which is modeled as a multi-hop, multi-path mesh network. It is based on a virtual currency system that encourages users to cooperate. Here users can utilize the service of other users in exchange for virtual currency or they can serve other users so as to collect virtual currency. In order to ensure the fair and efficient sharing of resources among users, they employ Nash Bargaining Solution (NBS) [11] that is computed in a distributed fashion. Nevertheless, the proposed decentralized scheme has some challenges and it is very difficult to implement it in case of users who are included in or left from a dynamic network topology.

There are also some proposed incentive mechanisms for network-assisted UPN services. For Karma model, Gao et al. [12] propose an incentive mechanism with usage-based pricing model instead of the fixed prize given by the mobile virtual network operator (MVNO). They employ two-stage Stackelberg game in order to model the interaction between MVNO and hosts. Firstly MVNO offers a pricing and reimbursing plan for each host, then every host decides the amount of data to be consumed by themselves and amount of data to be forward to clients. Changing the price and reimbursement affects the hosts' strategies, hence this also changes revenues of MNVOs. The proposed mechanism requires knowing data demand of clients and the decisions of both hosts and the MVNO to maximize the MVNO's revenue, which causes an additional signaling overhead. Khalili et al. [13] propose an alternative incentive mechanism based on two-stage Stackelberg game. In this study, mobile users can choose their roles such as a host, a client or not a subscriber. The prior study [12] focused on interactions between a particular host and its clients, but in the study, authors also characterized how the users' membership choices will affect each other's payoffs. Nevertheless, the variation of the connection qualities of hosts is not considered. In [14], a different incentive mechanism is proposed for network-assisted UPN solutions such as M-87, which is based on iterative bargaining scheme. In the considered model, participants that are customers of a

particular service provider can share their internet connections or they can improve the QoS by using connections of others. Utility and cost functions are defined considering provided data rate and energy consumption for client and gateway nodes. Rubinstein's bargaining model is employed in order to share the pie between utility and cost functions, and also a virtual currency system is introduced. The proposed model is beneficial for both network providers and participants and it is also adaptive to the network changes. In order to increase the success of such mechanisms, it is desirable to model the energy consumption and data rate in a more realistic manner. However, in the studies mentioned above, costs and utilities of users were not realistically modeled in terms of parameters such as energy and data rate. Studies in [9, 14] use very general energy consumption modeling proposed in [20], which did not consider actual signal parameters.

There exist several approaches to optimize the UPN services in term of energy consumptions. Ananthanarayanan et al. [15] propose a collaborative downloading system with a one-hop architecture where devices are within WLAN range but they combine their WWAN links to increase their available speed. In this system, nearby nodes form a collaboration group by using WLAN links and distribute traffic across their WWAN links for improving their network performance. Although it is worked on the cost of energy for users who share their cellular connections, it is primarily focused on performance rather than energy-efficiency. Sharma et al. [16] propose an energy-efficient architecture called Cool-Tether that optimizes the energy consumption of the WAN (GPRS/ EDGE/ 3G) and WiFi radios on smartphones. The study provides an on-the-fly WiFi hotspot that both energy-efficient and easy to use, by exploiting the cellular radio links of multiple mobile smartphones. Yu et al. [17] propose data sharing plan based on self-organizing networks and also they consider the dynamic characteristics of mobile terminals. A pricing scheme is applied in order to guarantee the cooperation among the users, and the management of the credits is done by a server. All of these studies only try to reduce energy costs while enhancing network performance, they do not provide any model for the energy consumption or throughput.

There are various studies for modeling energy and data rate in LTE networks. Huang et al. [18] evaluate the performance and power characteristics of LTE networks and provide a comparison with 3G/WiFi networks. They provide an empirically derived linear power model with respect to uplink/downlink data rates in addition to LTE states

and state transitions. In this study, they also investigated the impact of the configurations of DRX (a power management mechanism in LTE) parameters and tail timer on energy usage and signaling overhead. Pathak et al. [19] propose a power modeling approach based on tracing system calls, which captures utilization-based and non-utilization based power behavior of I/O components and network activities. It uses Finite State Machines to model the power states and state transitions of each. Here some states have a constant power consumption that represents non-utilization (system calls that do not imply utilization, i.e. file open/close) based power behavior and the other states use a linear regression model to capture the power consumption utilization based that generate workload. Schulman et al. [20] try to characterize the relationship between (3G) signal strength and power consumption on one hand, and between signal strength and data rate on the other hand. Via measurements, they observe that energy per bit can be as much as six times higher when the signal strength is low compared to communication with high signal levels. In order to reduce radio energy consumption for different application workloads, they present a framework that provides energy-aware cellular data scheduling by being cognizant of the signal strength. Thus, non-urgent communications can be deferred when the signal is low for energy-efficiency. However, neither this study nor the other mentioned studies provide a model with respect to signal characteristics.

In [21], it is examined the relationships among the signal parameters (such as SINR, RSRP, RSSI, and RSRQ) and evaluated the effect of SNR on throughput. However, it is not presented a model based on signal parameters. Cainey et al.[22] propose a linear model for estimating download throughput on LTE networks based on the crowdsourced information they collect via an Android application. They use LTE signal metrics (RSRP, RSRQ, and SINR) as well as time, device and operator information in the regression. They observe that no LTE signal metric is sufficient alone and the combined use of these metrics better explains the variability in download throughput. Also, they observe that SINR is more influential predictor. However, this model does not provide some linear regression assumptions such as homoscedasticity and normality of residuals. Violation of homoscedasticity and normality assumptions can cause distortion of the findings such as standard error, confidence interval and significance tests.

This thesis mainly focuses on energy and data rate modeling of a node which is standalone or involved in a UPN system. Firstly, we describe our network model and

propose a system architecture. Besides, we propose a data separation technique that divides data into states (such as silent, head, main, and tail) according to data per time while data is flowing. Then we make several data download tests for standalone case via our developed application that is also a sort of connectivity sharing service, and we model energy and data rate of a standalone node while downloading data according to LTE signal characteristics by using linear regression, robust regression, and other machine learning techniques. Before modeling, we divide the collected data into parts (such as head, main, and tail) according to energy per data so as to get more accurate information about energy and data rate. Also, we focus on modeling of energy and data rate of a node involved in UPN system. We examine the difference among standalone (download data to herself) node, gateway (download data to other) node, and client (download data via a gateway) node in term of energy usage and data rate and we model energy and data rate of a gateway node by using these experiment results. Lastly, based on the bargaining scheme proposed in [14], we examine the utility, cost and profit of users cooperated in a UPN system, by using resulting models and the data download tests.

Main contributions of our work can be summarized as follows:

- We introduce a framework that enables to estimate energy and data rate of a node while downloading or relaying data [23].
- We make data download experiments in different scenarios and point out how energy consumption and data rate changes when a node acts as a standalone, gateway or a client node.
- We propose two different data separation techniques that enable to divide data into states in order to get more consistent results for modeling. One of them makes separation according to energy per data while the other makes separation according to data per time and it is applicable for separation of flowing data.
- We use OLS, robust regression techniques and other machine learning techniques and present several energy and data rate models with respect to signal characteristics for a standalone node in LTE network [24]. To the best of our knowledge, our study is the first study that provides energy consumption modeling with respect to various LTE signal characteristics.

- Also, we provide energy and data rate models for a gateway node with respect to her standalone energy and her standalone data rate along with WiFi link speed.

The rest of the thesis is organized as follows: In Chapter 2, we describe the UPN's network model and present our proposed framework for energy and data rate modeling of a node. In Chapter 3, details and results of the experiments are given, and energy and data rate models are presented. Finally, Chapter 4 concludes the thesis and summarizes our work.

2. USER PROVIDED NETWORK SYSTEM WITH INCENTIVE MECHANISM

In this section, we describe our User Provided Network (UPN) model and propose a framework to design a UPN model with incentive mechanisms.

2.1. Network Model

In our network model, there are basically two types of nodes: client and gateway. A client is a consumer node which demands a mobile data and her Internet connectivity is limited or she has no connection at all. A gateway is a node which has directly connected to a base station and has strong network connectivity (3G-HSPA+ or 4G-LTE), which shares the Internet connection to the demanding neighboring nodes. Gateway node acts as a bridge that relay data coming from neighbor nodes to base station and data coming from base station to neighbors. In this network model, gateway and client nodes creates a network topology in which they are linked to each other by Bluetooth, WiFi Direct or LTE Direct technologies.

The standalone case and possible user-provided connectivity scenarios are shown in Figure 2.1. Part 0 represents standalone case in which a mobile user is directly connected to the Internet and downloads data for herself. Part 1 is the base scenario for user-provided connectivity in which a client tries to access the Internet via a gateway node with one hop. In Part 2, there exists one client node, one gateway and one (or more) relay nodes. This represents a multi-hop version of the user-provided network. Relay is a node that is used as a bridge to help the Internet access of a client. Relay node does not have to be directly connected to the base station. In this version, client node tries to access the Internet through two neighbor nodes. In our work, we firstly examine the standalone case scenario (Part 0) and then we focus on the basic scenario of the user provided network (Part 1). This is important for examining the difference between a standalone node and a node involved in a user-provided network in terms of energy consumption and data rate.

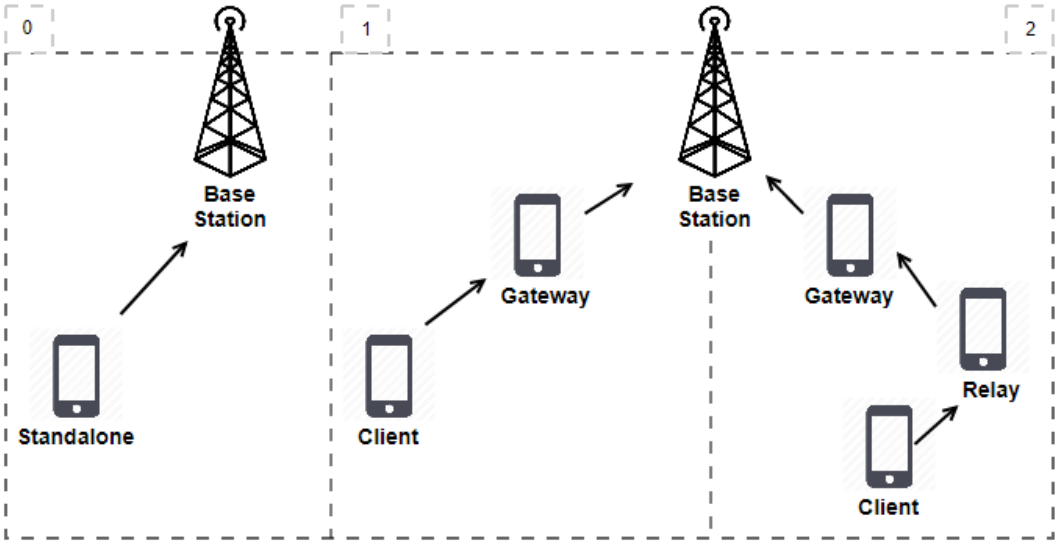


Figure 2.1 Standalone and user provided connectivity scenarios

2.2. System Architecture

We propose system architecture shown in Figure 2.2. It consists of three main modules, namely Data Collection module, Training module and Bargaining module. The details of the modules are explained in the next subsections.

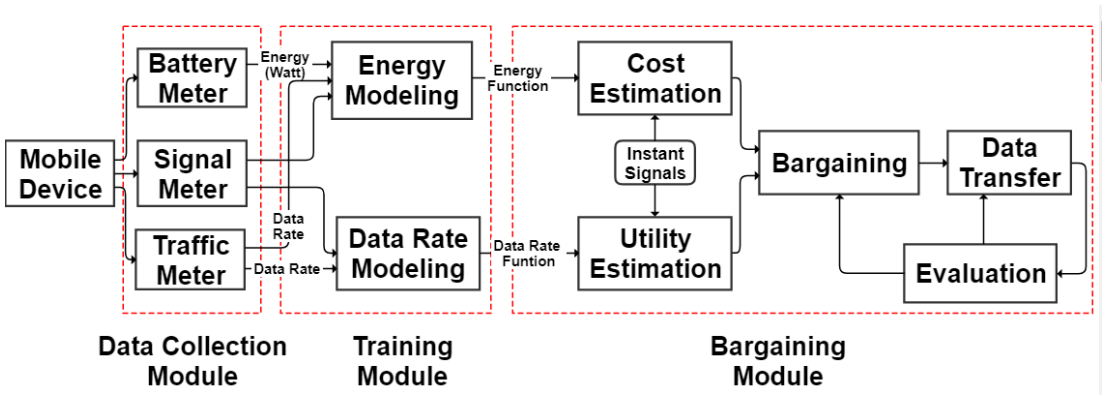


Figure 2.2 Proposed framework

2.2.1. Data collection module

This module is an experimental module that provides us to collect various statistics related to energy consumption, cellular/WiFi signal parameters, and mobile/WiFi data

usages on a smartphone. This module consists of three measurement units as Battery Meter, Signal Meter, and Traffic Meter, and these are described below.

Battery Meter is a software unit that measures voltage and current values and calculates the power spent and the energy consumed by a smartphone during data download or idle running. *Signal Meter* measures instantaneous signal strength and other signal parameters (such as signal quality, noise, modulation technique, WiFi link speeds and so on) on 3G/LTE and WiFi networks. *Traffic Meter* is a unit that periodically reads the amount of data transmitted/received over different network types (WiFi or 3G / LTE), and it also calculates data rate using collected statistics about data usage and download time.

In this module, our aim is to create a dataset, which consists of several parameters such as battery, signal and traffic statistics, by conducting a variety of data download experiments in different cases. The experiments that are made to create the dataset will be explained in Section 3.1.3. The created dataset will be used to model the energy consumption and the data rate on a smartphone (node) during downloading data for herself or others (detailed in Section 3.2.5).

2.2.2. Training module

In Training module, the dataset obtained from *Data Collection* module is analyzed, and the energy consumption and data rate of a node are modeled with respect to signal metrics. In this module, our aim is to determine the relationship among energy consumption, data rate, and signal metrics and to derive functions based on this relationship by using regression and other supervised learning techniques. This module consists of two sub-units: Energy Modeling and Data Rate Modeling.

In the *Energy Modeling* unit, we aim to derive an energy consumption function that estimates how much energy will be consumed by a smartphone during data download using both instant signal parameters and the data rate, or only instant signal parameters. This function aims to estimate how much energy will be consumed at which data rate, according to the instantaneous signal parameters. In the *Data Rate Modeling* unit, our purpose is to derive a general data rate function that relies on signal metrics-data rate relationship. The derived data rate function takes signal metrics as input and returns data

rate as an output. The derived functions are used in estimations of cost and utility on the *Bargaining module*. The details of energy and data rate modeling will be explained in Section 3.2.5.

2.2.3. Bargaining module

This module is based on the bargaining scheme proposed in [14]. A bargaining is made between the client and the gateway users according to the utility and cost estimations based on energy and data rate modelings. Before starting the bargaining, client estimates her own utility and gateway estimates her own cost. Then the client and the gateway start to bargain with each other using the utility and the cost estimations, if the parties come to an agreement, the gateway shares her available bandwidth with the client, in return client sends virtual currency to the gateway.

Utility. In [14], utility of client was expressed as accessing a better Internet connectivity than her available standalone Internet connection, so the utility was defined as a function of experienced data rate. The client will only decide to access the Internet over the gateway, when she expects higher data rate compared to the standalone case. Otherwise, the client may not be willing to be involved in the system since she will not make a gain. Therefore, she needs to estimate her utility before joining the system. The utility functions defined in [14] are given in Eq. (2.1) and Eq. (2.2). $U_{i0}(R_i^{sta})$ is the utility function experienced by the client node *i* itself (standalone). R_i^{sta} is the standalone data rate that is experienced by the client. $U_{ij}(R)$ is the utility function (per-bit) experienced by the client, when she reach the Internet via a gateway and receive data rate of *R*. If the experienced data rate *R* is less than R_i^{min}, which is the minimum data rate required by the client, then client receives zero utility. For elastic applications R_i^{min} is considered as zero, but it would be high for multimedia applications like streaming video.

$$U_{i0}(R_i^{sta}) = \begin{cases} \alpha_i \log(1 + R_i^{sta}) & \text{if } R_i^{sta} \geq R_i^{min} \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

and

$$U_{ij}(R) = \begin{cases} \alpha_i \log(1 + R) - U_{i0}(R_i^{sta}) & \text{if } R \geq R_i^{min} \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

where $\alpha_i \in [0,1]$ is for capturing the sensitivity to data rate.

Cost. If a gateway node shares her Internet connection with a client node, this creates a cost for the gateway. The cost was formulated in [14] as given in Eq. (2.3). In this formulation, C_{ji} is the cost function (per-bit) of a gateway *j* which is caused by serving a client *i* with a data rate *R*.

$$C_{ji}(R) = \frac{\beta_c}{R_{og}^{max} - N_g - R} + \frac{\theta_c(E_{go}^r(R_{og}^{max}) + E_{gc}^s(R_{gc}^{max}))}{(e_{residual})^2} \qquad (2.3)$$

The downlink data rate (average) between gateway *g* and client *c* which is maximum is represented with R_{gc}^{max} and the maximum average data rate of gateway *g* receiving from the operator *o* is represented with R_{og}^{max}. E_{go}^r represents a function of energy consumed for receiving over link (*g*,*o*) while E_{gc}^s indicates a function of energy consumed for sending over link (*g*,*c*). $\beta_c, \theta_c \in [0,1]$ are the normalization constants and they express the sensitivity of provided data rate and battery energy of gateway, respectively. N_g is amount of bandwidth which gateway allocates for herself needs. $e_{residual}$ represents remaining battery percentage value of the gateway, which gets values between 0 and 1. In this cost function, only downlinks have been considered, uplinks were ignored. In our study, we have referred the functions given above for Utility and Cost estimates, and we have simplified the cost function defined in [14] according to our needs. We have extracted $\frac{\beta_c}{R_{og}^{max} - N_g - R}$ part from the cost function because we assume that gateway can allocate as much bandwidth as her own needs (she only shares if she is not using for herself), so sharing the bandwidth of the gateway does not create a cost for herself in terms of data rate. We have modified the cost function with a more realistic equation given in Eq. (2.4). E_g^{oc} represents energy consumption function of a gateway node while receiving from operator *o* and sending to a client *c*.

$$C_{ji}(R) = \frac{\theta_c E_g^{oc}(R,S)}{(e_{residual})^2} \qquad (2.4)$$

R is not an estimated value, it represents a set of values from 0 to R_{max} which is maximum average downlink data rate. *S* represents instant LTE signal parameters. Both *R* and *S* are used for the estimation of the standalone energy consumption E_j^{sta} of gateway node *j*.

Bargaining. According to [14], if utility function is greater than the cost function, a cooperation area limited to R_0 and R_1 occurs where the client and the gateway can collaborate. This area was defined in Eq. (2.5) as a pie that will be shared between the client and the gateway according to data rate supplied by the gateway. The sharing of the pie is based on Rubinstein's bargaining scheme.

$$P_{ij}(R) = U_{ij}(R) - C_{ji}(R) \qquad (2.5)$$

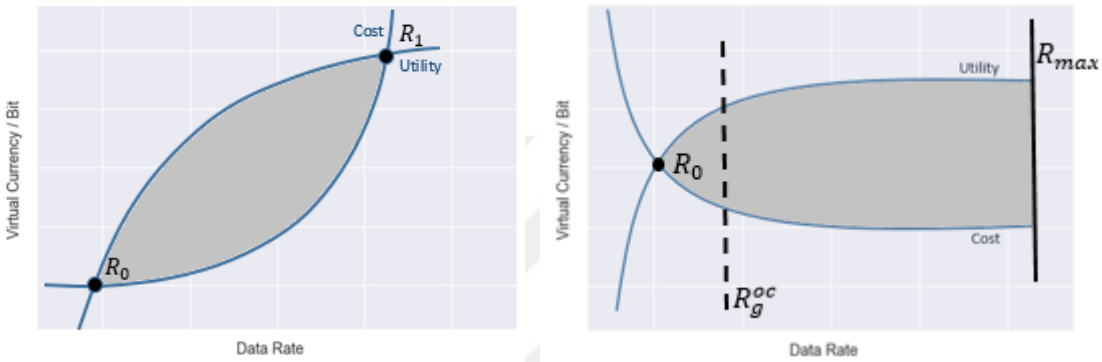
Rubinstein bargaining model [25] is a bargaining game that players take turns making offers over potentially infinitely many periods. In each period, a player makes an offer (about sharing of the pie). Game continues until one player accepts the other's offer. Every rejection leads to delay, which incurs a cost represented by discount factors $\delta_1, \delta_2 \in (0,1)$. δ_1 is the discount factor of gateway, while δ_2 is the discount factor of client. Share of one player changes depending on who makes an offer. For instance, if gateway offers first, gateway will get the share $\frac{1-\delta_2}{1-\delta_1\delta_2}P_{ij}$, and the amount of virtual currency (per bit) transferred from client *i* to gateway *j* is computed according to in Eq. (2.6). However, if client first makes an offer, the share of the gateway will be $\frac{\delta_1(1-\delta_2)}{1-\delta_1\delta_2}P_{ij}$, and the amount of virtual currency to be passed from client to gateway will be as in Eq. (2.7).

$$V_{ij} = C_{ji} + \frac{1-\delta_2}{1-\delta_1\delta_2}P_{ij} \qquad (2.6)$$

$$V_{ij} = C_{ji} + \frac{\delta_1(1-\delta_2)}{1-\delta_1\delta_2}P_{ij} \qquad (2.7)$$

In [14], client and gateway make an agreement if and only if there is a closed area between utility and cost functions, as illustrated in Figure 2.3.a. Also, amount of virtual currency to be transferred is computed by experienced mean data rate thus the participants do not pay more or less. If the experienced data rate is out of agreed range and it exceeds R_1, transferring virtual currency is done according to R_1. If the experienced data rate is less than R_0, then the amount of virtual currency to be transferred is equal to the client utility. In this way, it is ensured that net profit gained by the client is not negative. Different from [14], in our bargaining scheme, the utility and cost functions do not

intercept at any maximum data rate R_1 as seen in Figure 2.3.b. Because, cost function only depends on energy parameters, so the cost decreases while the utility increases as the data rate increases. However, it can still be assumed that there is a closed area because there is a maximum data rate threshold R_{max} that the gateway can provide. Client and gateway can only agree at a minimum data rate R_0 if and only if $U_{ij}(R) > C_{ji}(R)$ and $R_0 \leq R_g^{oc}$. R_g^{oc} is the received data rate of the gateway g from the operator o while relaying the client c.



a. According to the cost function in Eq. (2.3) **b.** According to the cost function in Eq. (2.4)

Figure 2.3 Cost and Utility functions in terms of virtual currency per bit

Data Transfer. After the bargaining**,** if the parties come to an agreement, the gateway shares her Internet connection (via WiFi) with the client and thus a connection is established between client and gateway nodes, and data transfer is done over the link. In general, data rate is too low at the beginning of data downloading process because the amount of downloaded data are low due to the TCP slow-start. Therefore, average data rate appears low during slow start period and the instantaneous data rate changes very quickly. Also, energy consumed per data is too high. After the slow-start phase, the amount of data transferred increases, energy consumed per data decreases, and the average data rate become more consistent. Towards the end of the data downloading, the data rate decreases rapidly again because the amount of downloaded data is too low but it is not finished completely yet. Because of all these reasons, it will be very difficult to stay in the feasible region during the whole data download process. Therefore, we propose a separation technique that divides data transfer process into four states as *head*, *main*, *tail,* and *silent,* depending on the amount of data downloaded per time and we apply different virtual currency policy according to these states in order to keep the negotiation.

Data transfer process is shown as a state diagram in Figure 2.4. *Head state* is considered as warm-up period in which data flow is slow due to the TCP slow-start. *Main state* is a state where the amount of data transferred is more than other states, and it gives more accurate information about data rate and energy usage, so we take into consideration only this state while modeling data rate and energy consumption, and also calculating achieved data rate. *Tail state* represents the last stage of data download process, where data download is almost completed but there are some negligible activities. *Silent state* is a state in which there is no data transfer, but there exists a connection between gateway and client nodes. The transitional rules among the states are described as follows (the rules are the same for both the client and the gateway nodes). The values of the parameters used for separation of data transfer process into the states are given in Table 3.20 in Section 3.3.

Firstly, it is passed into the *silent* state, after a connection is established between the gateway and the client. Then total amount of received (for client) or sent data (for gateway) $\boldsymbol{d_i}$ is checked during a time slot $\boldsymbol{t_i}$. Each time slot $\boldsymbol{t_i}$ is defined as 100 milliseconds (ms). If data amount $\boldsymbol{d_i}$ is less than or equal to threshold value $\boldsymbol{D_s}$, it is remained in the *silent* state. $\boldsymbol{D_s}$ is the maximum threshold value of negligible data amount caused by communication in the silent state. In case data amount $\boldsymbol{d_i}$ is greater than the threshold value $\boldsymbol{D_s}$, then it is switched to the *head* state. It indicates that the data flow begins between gateway and client.

After passing to the *head* state, a timer $\boldsymbol{t_h}$ is started that shows total waiting time in the head state, and also average data amount $\boldsymbol{D_{avg}}$ is calculated at each time slot $\boldsymbol{t_i}$. $\boldsymbol{D_{avg}}$ indicates the cumulative average amount of data transferred (or received) after it is passed to the head state. In the head state, the amount of data $\boldsymbol{d_i}$ is compared with the amount of data $\boldsymbol{d_h}$ at each time slot. $\boldsymbol{d_h}$ is assumed as the minimum amount of data to be received/sent while in the head state. If $\boldsymbol{d_i}$ is less than a certain data amount $\boldsymbol{d_h}$, the counter $\boldsymbol{c_h}$ is increased, otherwise the counter $\boldsymbol{c_h}$ is set to zero**.** The counter $\boldsymbol{c_h}$ is used to decide whether to switch to the silent state. If the counter $\boldsymbol{c_h}$ exceeds the threshold value $\boldsymbol{T_{h1}}$, then it is returned to the silent state. Besides, if cumulative average data amount $\boldsymbol{D_{avg}}$ is greater than the threshold value $\boldsymbol{D_h}$ which is maximum average data amount to be achieved in the head state, it means that the *head* state ends and the *main*

state starts, hence it switches to the *main* state. If none of these happens, it is waited in the head state until the timer $\boldsymbol{t_h}$ exceeds the threshold value $\boldsymbol{T_{h2}}$ that represents maximum waiting time in the head state. Then, it switches to the *final* state in which data transfer protocol ends.

In the *main* state, data amount $\boldsymbol{d_i}$ is compared with a certain value $\boldsymbol{d_m}$ which is considered as the minimum amount of data transferred during the *main* state. If $\boldsymbol{d_i}$ is lower than $\boldsymbol{d_m}$, the counter $\boldsymbol{c_m}$ is increased and otherwise the counter $\boldsymbol{c_m}$ is set to zero. Here it is aimed to find the starting point of the tail state. In case the counter $\boldsymbol{c_m}$ exceeds the threshold value $\boldsymbol{T_m}$ which is a fixed time value, it is passed to the tail state. Also, the interval in which $\boldsymbol{c_m}$ ranges from 1 to $\boldsymbol{T_m}$ value is accepted as the tail part and hence consumed energy and achieved data amounts during the interval are included in the tail state.

After passing to the *tail* state, a timer $\boldsymbol{t_t}$ begins. It is kept in the *tail* state until data amount $\boldsymbol{d_i}$ is greater than the threshold value $\boldsymbol{D_t}$ and the timer $\boldsymbol{t_t}$ exceeds the threshold value $\boldsymbol{T_t}$ which is the maximum waiting time in the tail state**.** If data amount $\boldsymbol{d_i}$ is greater than the threshold value $\boldsymbol{D_t}$, then the timer is reset and it switches to the head state. Otherwise, after the time $\boldsymbol{t_t}$ has elapsed, then it switches back to the silent state.

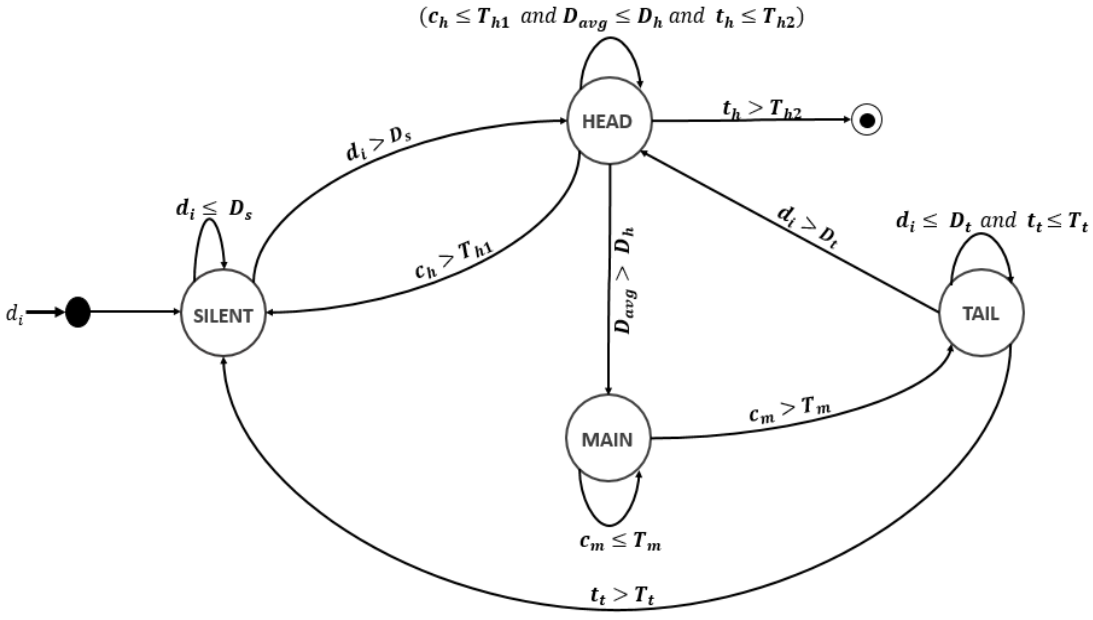


Figure 2.4 Data transfer process state diagram

During the data transfer for all the states, instant signal strength and other signal parameters (RSRQ, SINR and etc.), received/transmitted data amount, parameters related to energy consumption (voltage, current, battery level) are also measured. Actual consumed energy and provided/ achieved mean data rate are calculated at certain time intervals. Then, actual utility and actual cost values are computed by using these calculated and measured values.

Virtual Currency Transfer. A virtual currency system was presented in [14] that allows secure transfer of virtual currencies between the nodes. In this system, the network operator may give some rewards to the participating nodes before bargaining since they will spend energy while communicating with each other. After the bargaining, the virtual currency transfer from client to gateway is done according to achieved data rate for each period. This encourages participant nodes to accurately specify their parameters during bargaining.

In our study, transferring virtual currency is based on the states seen in Figure 2.4. If it is in the main state, we expect that the experienced mean data rate is within the feasible region and so the transferred amount of virtual currency is determined according to the achieved data rate by using the Eq. (2.6) or Eq. (2.7) depending on which node offers first. However, if the achieved data rate is out of the feasible region while in the main state, then it is sent to the *Evaluation* unit after the virtual currency transfer according to the achieved data rate. On the other hand, if it is in the head state or tail state, the amount of virtual currency transferred from client i to gateway j is determined according to Eq. (2.8). Because, in these states, the transferred data amount is low thus client data rate is low, but energy consumption of gateway is high (namely, energy per data is high), so if the amount of virtual currency to be transferred is determined according to the achieved data rate, client may loses. If the amount of virtual currency to be transferred is determined according to only the client utility function, than gateway may suffer in term of energy cost. Therefore, we used different virtual currency policy for the benefit of both parties. Thus, client pays virtual currency to gateway for her received data amount and network affiliation (energy) cost. Otherwise, if it is stayed in the head state for a long time during data transfer then it goes to the final state where the data transfer protocol ends.

$$V_{ij} = C_{con} + \rho D_{ji} \qquad (2.8)$$

where $\rho \in (0,1]$ is normalization parameters. C_{con} represents energy cost of gateway caused by the client being connected to the gateway. D_{ji} is transferred amount of data from gateway j to client i during one round. If it is in the silent state, then the amount of virtual currency to be transferred is computed according to energy cost of the gateway, i.e. $V_{ij} = C_{con}$.

Evaluation. Participants' sources and needs can change over time thereby utility and cost functions can also change. Therefore, participants share actual experienced utility and cost functions at the end of each period and they are compared with the agreed functions (which are based on estimated values). The expectation is that actual utility and cost values overlap the utility and cost functions that have already been agreed upon if it is in the main state. In other words, it is anticipated that the net profit for both parties remains within the feasible region shown in Figure 2.3.b. Participants decide whether to continue the protocol operation according to the net profits they receive. We assume that gateway and client are in the same state during the data transfer. If the profits of the participants are within the agreed area when client node is in the main state, virtual currency transfer is done according to Eq. (2.6) or Eq. (2.7). There is no change in agreement conditions and data transfer will continue under the same agreement conditions. However, if the client is in the main state but the profits are out of feasible region which is determined by the agreed cost and utility functions at the beginning, transfer of the virtual currency is done according to the same equations, but then the participants may make a new bargain or they terminate the data transfer operation. Also, in such cases, a new bargaining process may be initiated with another node which does not exist in the previous session. Apart from that, if the client is in one of head or tail states, the agreement conditions do not account and the virtual currency transfer is done according to Eq. (2.8).

2.2.4. Protocol operation

The protocol operation between client and gateway nodes is based on [14]. The protocol starts with a request message (REQ) sent by the client to the gateway. The request message includes the client utility function $U(R)$ and WiFi link speed L_{gc}. Then the gateway responds with a message (OFFER) that includes her cost function $C(R)$, the

function of offered virtual currency $V(R)$, and the received data rate R_g^{oc} from the operator while relaying the client which is estimated by using standalone data rate and WiFi link speed L_{gc}.

The client checks whether the received data rate R_g^{oc} of the gateway is greater than or equal to the agreed data rate R_{agr} if there is any agreement point. Then the client sends a message (OK) with her agreed virtual currency function and the agreed minimum data rate R_{agr} if she accepts the offer. When the gateway receives the accept message of the client, the gateway shares her connection and the client accesses the Internet via the gateway's hotspot (it can be used in any other technology, i.e. LTE-Direct), thus data flow starts. On the other hand, the client may not accept the gateway's offer and she can send another offer. However, as we mentioned before, delays are costly and every delay reduces the profits to be achieved due to the discount factor so that participants will not tend to reject an offer. Client calculates transferred data amount D_A and achieved data rate R_A for each time period T_s. Then she sends a transfer message (TRANS) to the gateway. The message consists of transferred data amount D_A, achieved data rate R_A (if it is in the main state) and amount of virtual currency (calculated according to the states). The gateway checks the correctness of the amounts in the received message and sends the message to the network operator. After verifying the amounts in the message by the operator, the virtual money transfer is completed and the operator informs both parties in this regard. Possible protocol operation is illustrated in Figure 2.5.

At the end of each period, gateway and client evaluate their expectations and their experience. They can start a new bargaining session if the obtained ones do not meet the expectations or utility and cost functions have changed.

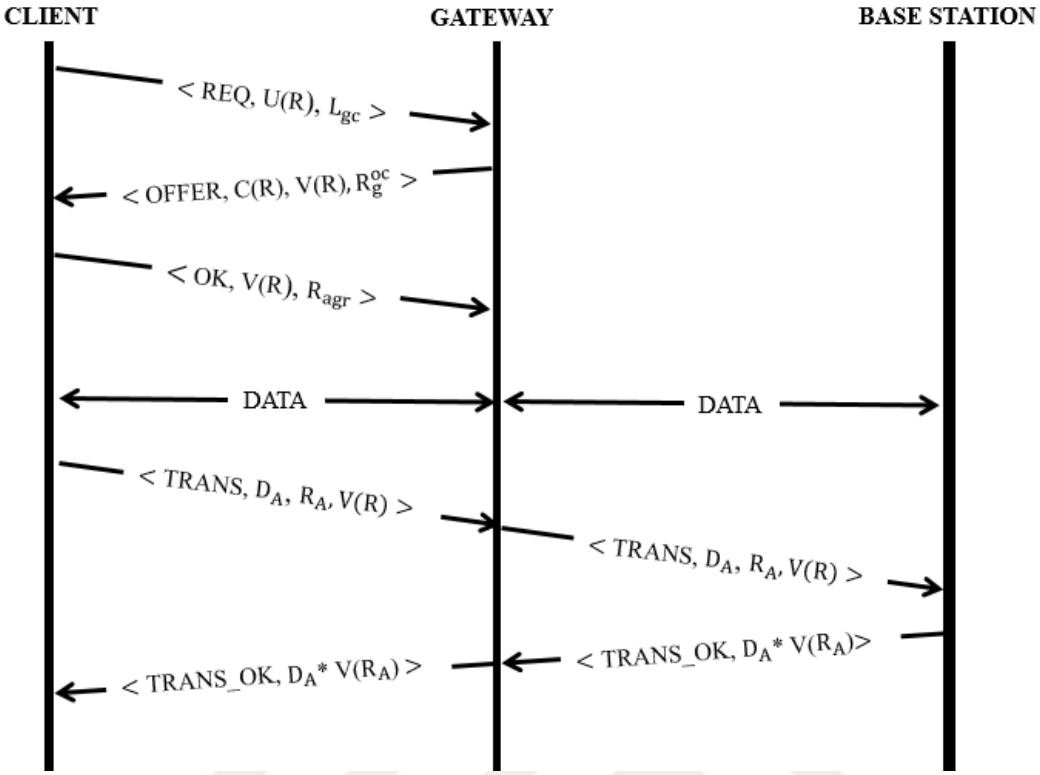


Figure 2.5 Protocol operation

3. ENERGY AND DATA RATE MODELING

In this section, we describe our performed experimental studies to examine the energy and data rate of a node acting as a standalone, gateway or a client node. Then we present several energy and data rate models for standalone and gateway nodes based on regression techniques.

3.1. Experimental Study

In this section, we introduce our developed application and describe test environment. Also, we describe our performed experiments which are idle running and data download tests (standalone, standalone vs. gateway and standalone vs. client).

3.1.1. Tool and environment

We have developed an Android application that enables users to communicate with each other via WiFi-Direct and to share the Internet connection with others via the WiFi hotspot or to request the Internet connection. The application measures voltage, current, 3G/LTE signal parameters (RSCP, RSRP, RSRQ, and SINR) and amount of downloaded data. It also calculates power spent, net consumed energy and data rate on a smartphone during the data downloading. The application consists of three different measurement units: battery meter, signal meter, and traffic meter.

In battery meter unit, the application measures the phone's current, voltage, and other battery information (e.g. battery percentage, charging status, temperature and so on). To make these measurements we have used *android.os.BatteryManager* class that provides information about current value (bundle.getInt ("current_now")), voltage value (bundle.getInt ("voltage")), remaining battery percentage (BatteryManager.EXTRA_LEVEL), charging status (BatteryManager.EXTRA_PLUGGED), and temperature (bundle.getInt ("temperature")). On some smartphones, current value drawn by the phone cannot be measured directly through the current function provided by the Android API. This is because the file path where current information is stored varies according to phone manufacturer. In this case, the file path in which the current value is stored must be

defined for each phone brand (even the model). Therefore, we have defined some current paths belonging to different phone brands (such as Samsung, LG, Asus, HTC and so on) in the application. The amount of power consumed by a smartphone varies depending on whether the phone screen is on or off and on the screen brightness level. For this reason, we have used isScreenOn () method in android.os.PowerManager class to control whether the phone screen is open or not. In addition, we have used some methods in the android.provider.Settings.System class to get screen brightness level on a smartphone.

Signal meter collects statistics about connected network types (EDGE, HSPA+, LTE or WiFi), and it measures 2G/3G/4G signal strength and other signal parameters according to network types. For example, it measures RSCP signal strength for HSPA+ network and RSRP, RSRQ, and SINR for LTE network. We have also tried to measure other signal metrics related to 2G/3G networks, but the Android API does not report any 2G/3G signal metrics except LTE signal metrics. Besides, we get information about connected WiFi network related to network name, signal strength and link speed via getSSID() , getRssi() ve getLinkSpeed() functions in android.net.wifi.WifiManager class. While developing the measurement unit, we use android.telephony.CellInfo, android.telephony.SignalStrength, android.telephony.CellSignalStrengthLte ,android.telephony.CellInfoLte, android.telephony.PhoneStateListener, and android.telephony.TelephonyManager classes and their methods.

In traffic meter unit, we use the android.net.TrafficStats class for getting data usage statistics. This class provides network traffic statistics about bytes transmitted/received and network packets transmitted/received over all interfaces (WiFi or Mobile). In this class, the GetTotalRxBytes () method is used for the number of bytes received over all interfaces since device boot, while the getTotalTxBytes () method is used for the number of bytes transmitted. The getMobileRxBytes () and getMobileTxBytes () methods are used for the number of bytes received and transmitted over the mobile interface. To obtain the number of bytes received and transmitted over the WiFi interface, the number of bytes received and transmitted over the mobile interface must be subtracted from the number of bytes received and transmitted over all interfaces. On our application, we also use Runnable (java.lang.Runnable) interface and Handler (android.os.Handler) class for making all measurements periodically. It makes various measurements every 100 ms. The application should be able to run in the background because it is considered a service

application. For this purpose, the Service (android.app.Service) abstract class and WakeLock (android.os.PowerManager.WakeLock) class is used thus the phone does not go into sleep mode after the screen is turned off and it can function as a service application.

We built a hardware circuit that measures the power consumption of a smartphone in order to check the accuracy of the power measurements performed by our application in [23]. As illustrated in Figure 3.1, the circuit consists of an Arduino microcontroller, a breadboard, resistor unit (includes 13 resistors), a smartphone (Samsung Galaxy J7) that has a removable battery and an apparatus for connecting the battery interconnects (pins) of the smartphone to the circuit. The working principle of the circuit is as follow. 5 Volt voltage supply is connected in series with the smartphone and resistor unit. We used the smartphone battery as a voltage supply on the circuit. Analog read pin (A0) of Arduino is connected to the resistor unit and thus the voltage difference (Vr) on the resistor unit can be read at the desired sampling time. The current flowing through the resistor unit is calculated by dividing the read voltage difference by equivalent resistance of the resistor unit. In this way, the current (I1) flowing through the smartphone is also calculated, because the current flowing through the resistor unit is equal to the current flowing through the smartphone due to the serial connection. Another analog read pin (A1) of Arduino is connected to the battery connector of the smartphone via the apparatus. By this means, the current and the voltage difference (V1) on the smartphone can be read at desired sampling time, and the power and energy values can be easily calculated.

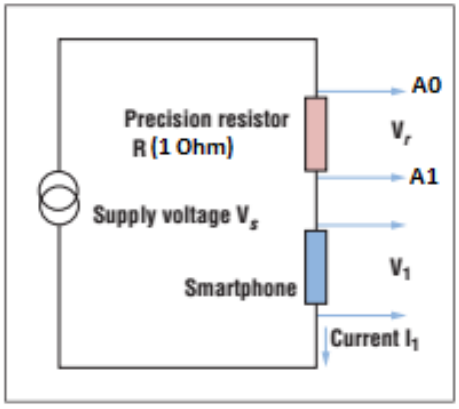


Figure 3.1 Energy meter circuit scheme

We have done various measurement (power/energy) experiments both in software and hardware. We observed that the results of the software and hardware measurements are very similar to each other and the difference between them is typically below 5% as seen in Figure 3.2. This indicates that the accuracy of the measurements made on our application is sufficiently high.

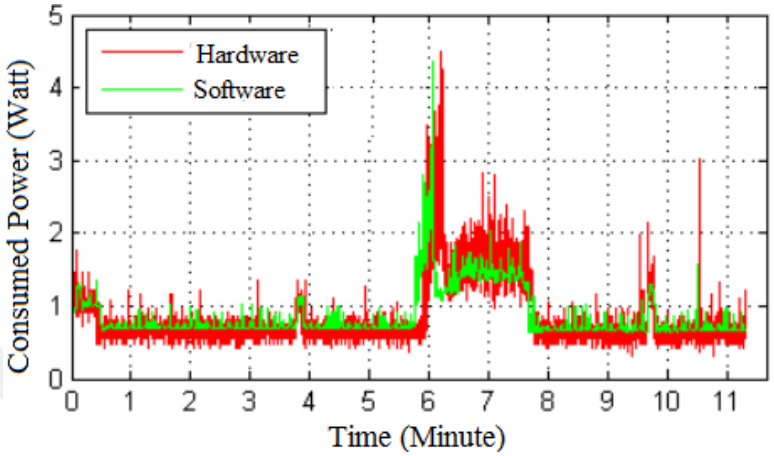


Figure 3.2 The comparison of hardware and software measurement results

We have used our developed application for gathering information about energy consumption, signal, and data rate during the standalone, standalone-gateway, and standalone-client experiments. The data download tests for the standalone experiments were conducted on both HSPA+ and LTE networks. For standalone-gateway and standalone-client experiments, data download tests are only conducted on the LTE network type. The vast majority of experiments were carried out indoors. In our experiments, we have used 5 smartphones with the properties given in Table 3.1, and we have used Turk Telekom (TTNet) as serving network operator.

Table 3.1 List of used devices

Brand	**Model Number**	**OS**	**CPU**	**Ram**	**Network**	**Battery Capacity**	**Battery Type**
Samsung Galaxy J7	SM-J700H	Android 5.1.1	1.5GHz	1.5GB	HSDA+	3000 mAh	Li-ion
	SM-J700F	Android 5.1.1	1.5GHz	1.5GB	4G	3000 mAh	Li-ion
Asus Zenfone 2	ASUS_Z00AD	Android 5.0	2.3GHz	4GB	4G	3000 mAh	Li-poly
LG G3	LG-D855	Android 4.4.2	2.5GHz	3GB	4G	3000 mAh	Li-ion

We follow some rules during the experiments. For instance, in all our experiments (including the idle state experiments), after turning the mobile phone on, we wait for 7-8 minutes before starting the measurements. We observed that this is important for obtaining more accurate results. Also, we keep a 3-4 minutes interval between any two tests. During the measurement, all other applications are closed and the phone's screen is at minimum brightness level.

3.1.2. Idle running experiments

In order to calculate the net energy consumed by a smartphone during data downloading, it is necessary to measure the energy consumed by the device in the idle state and subtract it from the total consumed energy. The idle state means that no application is running on the background or foreground. We performed some idle running tests and during these tests, WiFi, Bluetooth, hotspot and mobile data were turned off and no applications were running on the background or foreground except our measurement application. The idle running tests on smartphones are basically divided into two groups depending on the screen status: Screen On and Screen Off. If the smartphone screen is off, the brightness level is not important. However, if the screen is on, the brightness level becomes important in terms of the power consumption. Some tests were done to see the effect of the smartphone's screen state (on / off) and screen brightness level on power consumption. In these tests, for each brightness level, 6 tests were carried out, each for 10 minutes, and a 60-minute test was performed for the phone's off-screen status. The results of these tests are shown in Table 3.2. As it can be observed, each smartphone has different power consumption in the idle state. This is because each smartphone has different hardware and software features.

The power values spent in the idle state given in Table 3.2 are used to calculate the net energy consumed on a smartphone during the data download. While the calculation of energy consumption during the data download, the power spent in the idle state is subtracted from the instant total spent power. Then the net energy is calculated by using these net power values.

Table 3.2 Power consumption tests in idle state

	Samsung Galaxy SM-J700F/H		LG G3		Asus Zenfone 2	
Display	**Brightness (over 255)**	**Average Power(W)**	**Brightness (over 255)**	**Average Power(W)**	**Brightness (over 255)**	**Average Power(W)**
Off	-	0.446	-	0.092	-	0.294
On	5*	0.802	50 *	0.449	15*	0.690
On	55	0.919	80	0.481	44	0.765
On	105	1.086	110	0.509	84	0.884
On	155	1.320	170	0.586	124	0.987
On	205	1.468	200	0.688	164	1.105
On	255**	2.025	255**	1.062	204**	1.410

Min, Max***

3.1.3. Data download experiments

We have made several tests to examine how to change the energy consumption and data rate in a smartphone during downloading data from a web server [26] for itself or others. These tests are standalone tests (on both 3G and 4G networks), standalone-gateway comparison tests (only on 4G networks), and standalone-client comparison test (only on 4G networks), shown in Figure 3.3.

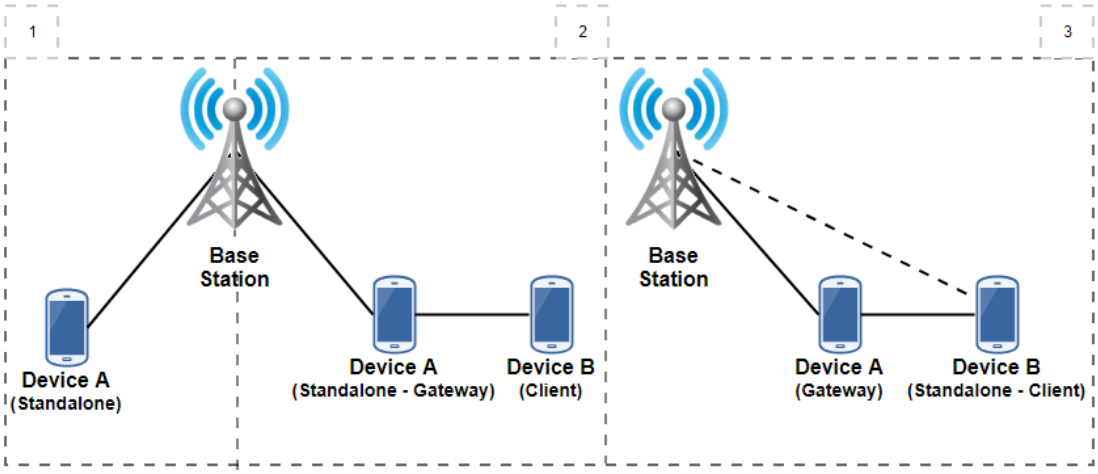


Figure 3.3 Possible data download scenarios

During the experiments, we have measured energy consumption, average data rate, several LTE parameters (i.e. RSRP, RSRQ, and SINR), WiFi signal strength, WiFi link speed and other parameters. While analyzing the tests, we apply the data separation technique based on data per time (in Section 3.2.4) and we use the main part of the experiments in the next subsections.

3.1.3.1. Measured signal metrics

In LTE networks, radio resource management is performed according to Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), Signal to Interference plus Noise Ratio (SINR) and Channel Quality Indicator (CQI). These metrics provide information about signal strength and channel quality. RSRP is a cell-specific metric indicating signal strength that is primarily used for cell selection and handover. RSRP measurement does not include noise or interference from neighboring or serving cell [27, 28]. It can be affected by multipath fading and so RSRP reading can be seen low while the actual throughput may be high [22]. The reporting range of RSRP is defined from -140 dBm to -44 dBm with 1 dB resolution. RSRQ is a cell-specific signal quality metric that provides extra information to determine when to perform a handover. It provides a ranking among different candidate cells as regards their signal quality. It can be used together with the RSRP in making cell reselection and handover decision when the RSRP measurements are inadequate to make reliable decisions. The reporting range of RSRQ is defined from -19.5 dB to -3 with 0.5 dB resolution [29]. SINR is a measure of signal quality that can be used to evaluate effect of interference upon the link. It is not defined in 3GPP specifications; it is specified by user equipment vendors. CQI is a quantized and scaled version of the experienced SINR. It is used to select the user for data transmission and to determine modulation and coding scheme [21, 30]. In the UMTS cellular communication system, Received Signal Code Power (RSCP) is an indication of signal strength. It is used for handover and calculating path loss. The range of RSCP is given from -115 dBm to -25dBm with a resolution of 1 dB [31]. However, our measurements range from -51 dBm to -113 dBm.

3.1.3.2. Standalone experiments

In this experiment, our main aim is to measure the consumed energy per data and average data rate according to 3G/LTE signal metrics during data transmission. For this purpose, we download 50 MB test files at different signal strengths. The experiments are conducted on HSPA+ and LTE networks. Our experiment scenario is shown in Figure 3.3 Part (1). A user (Device A) is a standalone node that connects to Internet via cell tower for her own internet needs and downloads data to herself. She doesn't share her connection with others, so this is called standalone.

On the evolved-HSPA (HSPA+) network, we have conducted 280 data download experiments on Samsung Galaxy SM-J700H smartphone, and measured energy consumption, data rate and RSCP signal strength during the experiments. The results of the tests are shown below. Figure 3.4 illustrates the average value and standard deviation of the power at different RSCP signal levels during data download tests on HSPA + network.

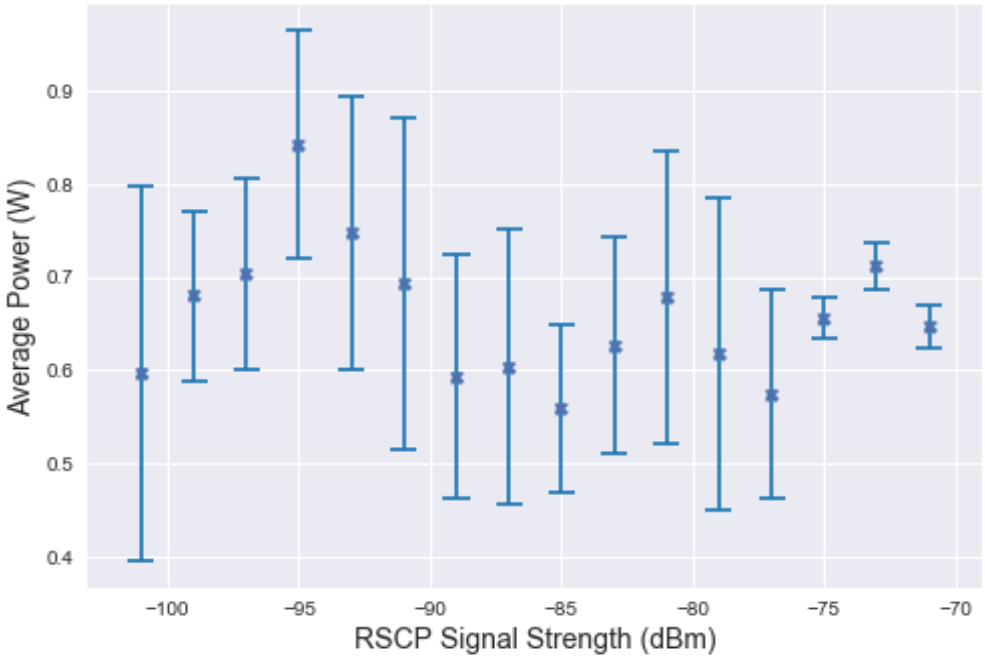


Figure 3.4 Average power and signal values while downloading 50 MB data file on HSPA+ network. Error bars represent mean and standard deviation of the power.

It is seen that the average power spent tends to decrease at the RSCP signal strengths which are stronger than -95 dBm, and the range of standard deviation also decreases as

the signal strength increases (especially at -75dBm and stronger signal strengths). Although the average consumed power is low at some points when the signal strength is high, the correlation between the consumed power and RSCP signal strength is not strong. Because, in some cases (e.g. at -81 dBm and -73 dBm), average consumed power may increase despite the signal strength is high, or vice versa.

We have also examined the relationship between energy consumption and signal strength while downloading data. Figure 3.5 illustrates the total average energy consumption and standard deviation values of the energy at different RSCP signal strengths. The average energy consumed tends to decrease as the RSCP signal strength increases (especially for the signal strengths stronger than -90 dBm). However, at some signal levels, the energy consumption is very low when the signal strength is weak (such as at -101 dBm). This is because data rate may be high at that moment due to low user density or other positive factors. In other words, even though the signal strength is low, the total amount of energy consumed is also low if the current data rate is high at that time. For this reason, the consumed energy can vary according to the data rate as well as the signal strength.

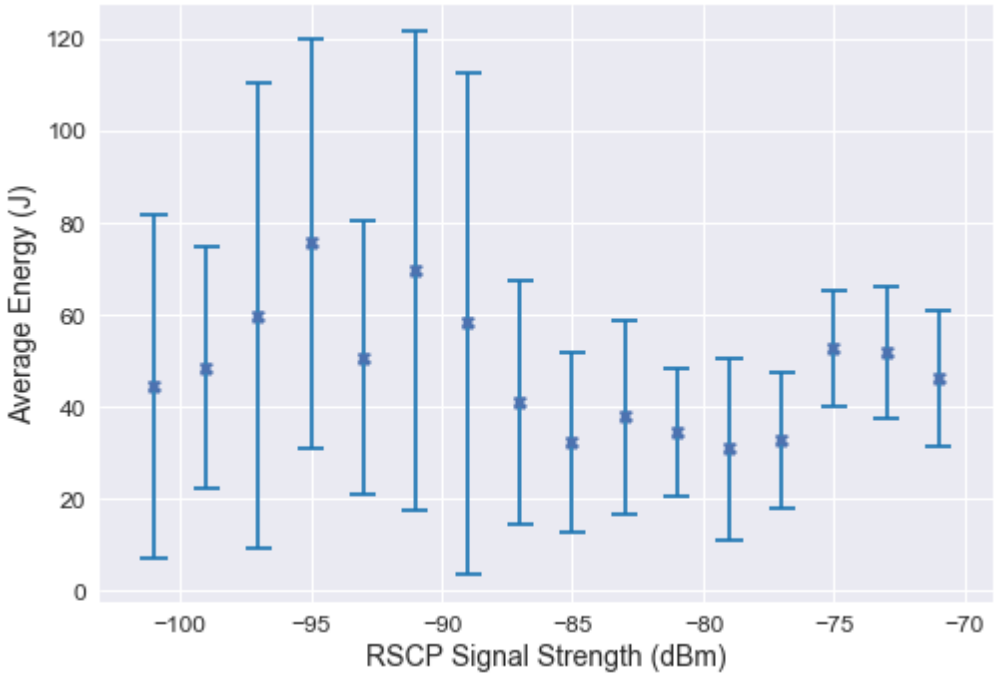


Figure 3.5 Average energy and signal values while downloading 50 MB data file on HSPA+ network. Error bars represent mean and standard deviation of the energy.

The relationship between average data rate and RSCP signal strength is shown in Figure 3.6. The average data rate increases at certain signal intervals and decreases at certain signal intervals. However, this increment and decrement in data rate are not proportional to the signal strength. For example, the data rate between -75 dBm and -71 dBm is lower than the data rate between -89 dBm and -77 dBm. We consider that data rate is affected by the factors such as weather conditions and user density on the network, as well as signal strength and other signal parameters. However, we cannot get more signal parameters on 3G networks because of software limitations, so we performed later data downloading experiments on the LTE network.

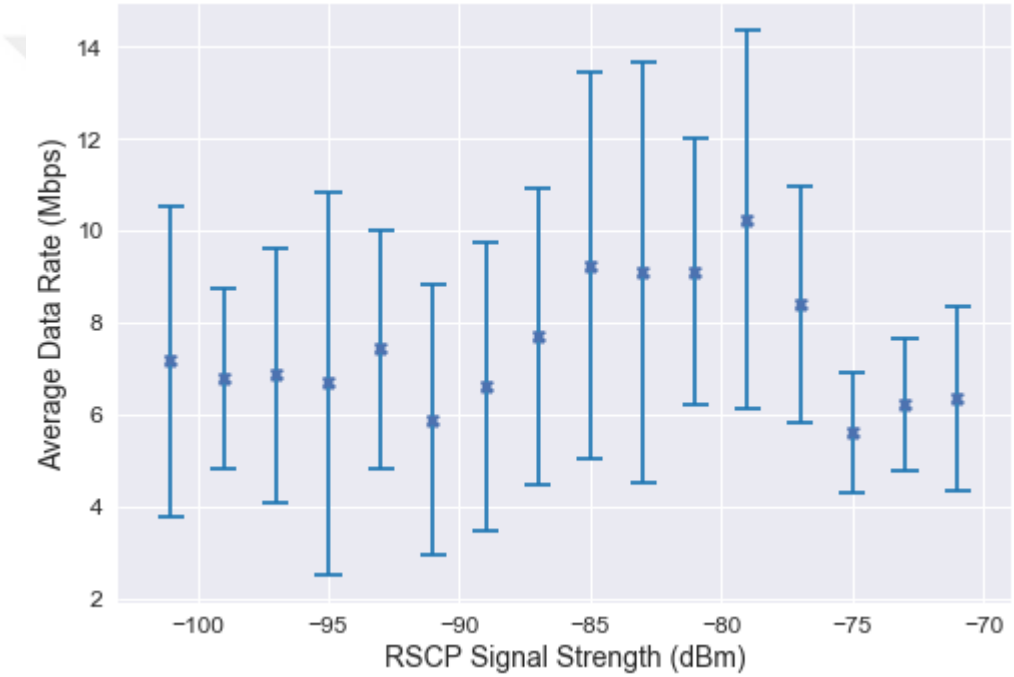


Figure 3.6 Average data rate and signal values while downloading 50 MB data file on HSPA+ network. Error bars represent mean and standard deviation of the data rate.

We have performed 360 standalone tests on LTE network by only measuring various received signal strength (RSRP). In addition, we have conducted 515 standalone tests on LTE networks by measuring some additional LTE parameters such as RSRQ, SINR, and CQI, together with signal strength (RSRP), energy consumption, power spent, and data rate. During these tests, we have used Samsung Galaxy SM-J700F smartphone. Firstly, we examined the relationship of signal strength to power, energy consumption

and data rate. Later, we have focused effects of other LTE signal metrics on power, energy and data rate. The figures 3.7, 3.8 and 3.9 illustrate the overall mean and standard deviation of the power, energy, and data rate values at different signal strengths. According to Figure 3.7, we can say that the spent power decreases as the RSRP signal strength increases (especially until -100 dBm). On the other hand, the spent power decreases very slowly at the signal strengths above -100 dBm and the mean values of the spent power are close to each other. Also, at some high signal strengths, the average spent power has increased while it is expected to decrease.

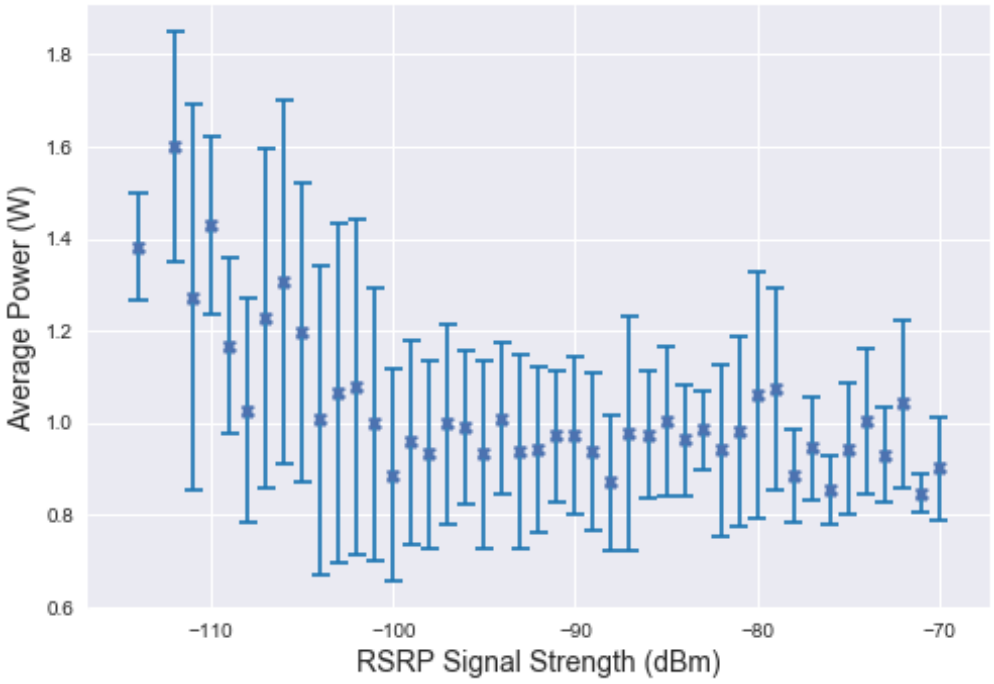


Figure 3.7 Average power and signal values while downloading 50 MB data file on LTE network. Error bars represent mean and standard deviation of the power

The relationship between average energy consumed versus RSRP signal strength is illustrated in Figure 3.8. In general, the energy consumption decreases as the signal strength increases, and a significant reduction in energy consumption was observed, especially at signal strengths greater than -98 dBm. In addition, the standard deviation of the energy also decreases as signal strength increases. At the signal strength higher than -98 dBm, the energy consumption trend is similar. However, there are some unexpected situations related to energy consumption-signal strength. For instance, the energy consumption at -110 dBm and -109 dBm is lower than the energy consumption at the

signal strengths between -108 dBm and -111 dBm. This is because energy consumption is affected by the data rate and other signal metrics along with the signal strength (RSRP), as we mentioned before.

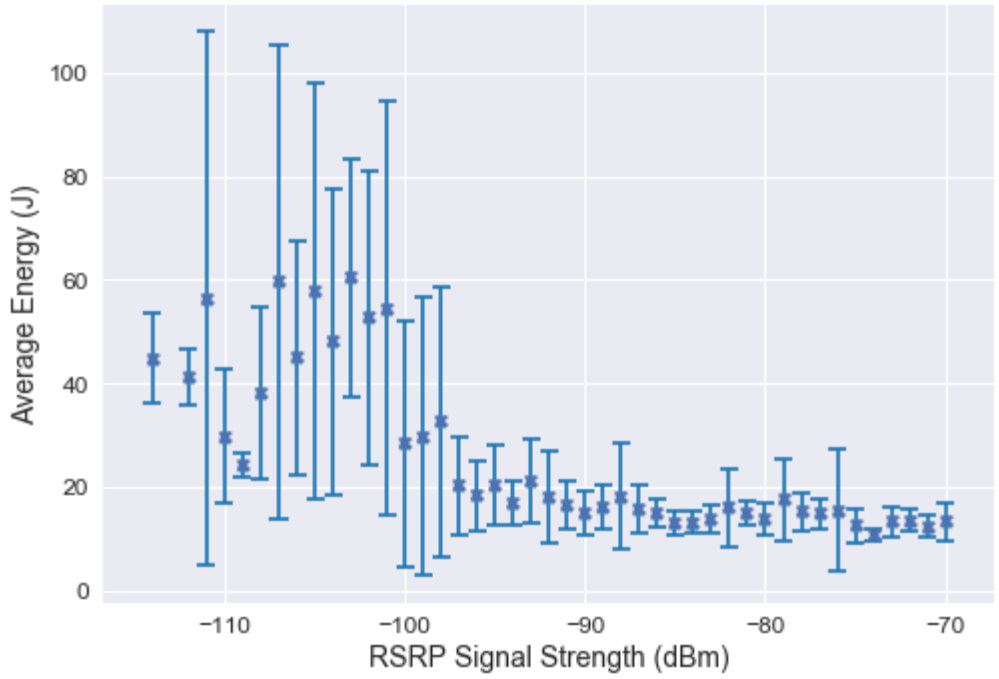


Figure 3.8 Average energy and signal values while downloading 50 MB data file on LTE network. Error bars represent mean and standard deviation of the energy

Figure 3.9 shows the change in average data rate with respect to RSRP signal strength on LTE network. In general appearance, we can say that the data rate increases as signal strength increases (especially where the signal strength is greater than -100 dBm.) However, in some cases, the increase or decrease in the data rate do not take place in an orderly manner depending on the signal strength (e.g. at -110 dBm, -93 dBm, -71 dBm and -78 dBm.

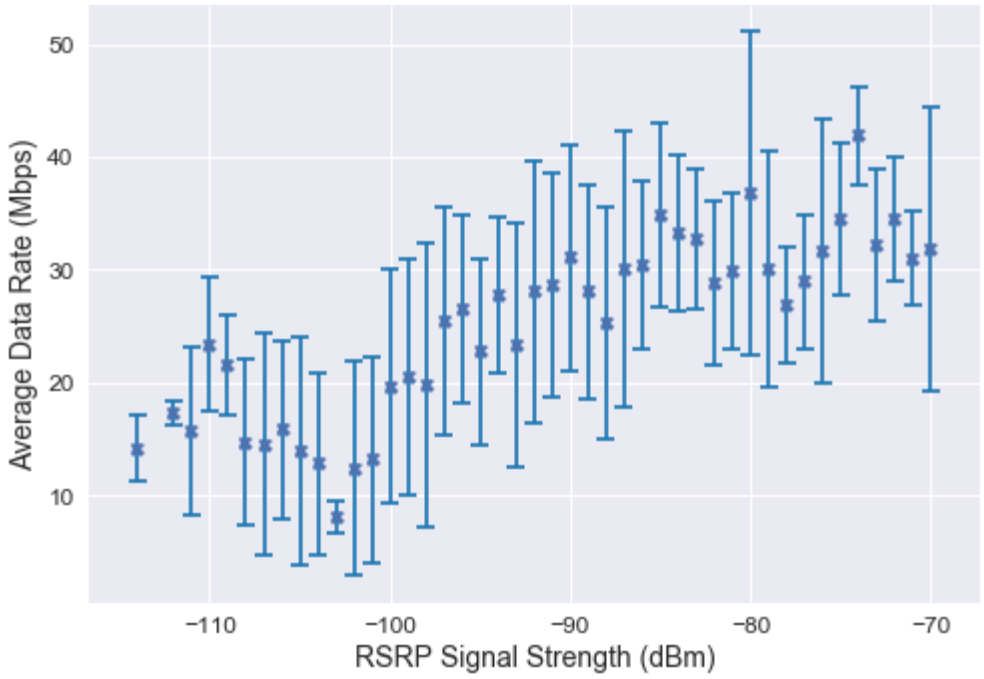


Figure 3.9 Average data rate and signal values while downloading 50 MB data file on LTE network. Error bars represent mean and standard deviation of the data rate.

The results show that the RSRP signal strength is not sufficient alone to explain the change in energy and data rate. Therefore we measured other LTE metrics along with the signal strength (RSRP). The scatter plots related to these tests are shown together with the correlation coefficient (Pearson correlation) and its significance value (p). It is illustrated the relation of average power consumption to LTE signal metrics in Figures 3.10.a, 3.10.b, 3.10.c, and 3.10.d. According to these results, there exists very weak correlation between power consumption and LTE metrics (except RSRP). Power is more correlated with RSRP than the other metrics but the correlation coefficient (that is -0.23) between the power and RSRP does not also represent a strong correlation. Roughly, we can also say that the power consumption decreases as RSRP increases, but this is not very clear due to the weak correlation.

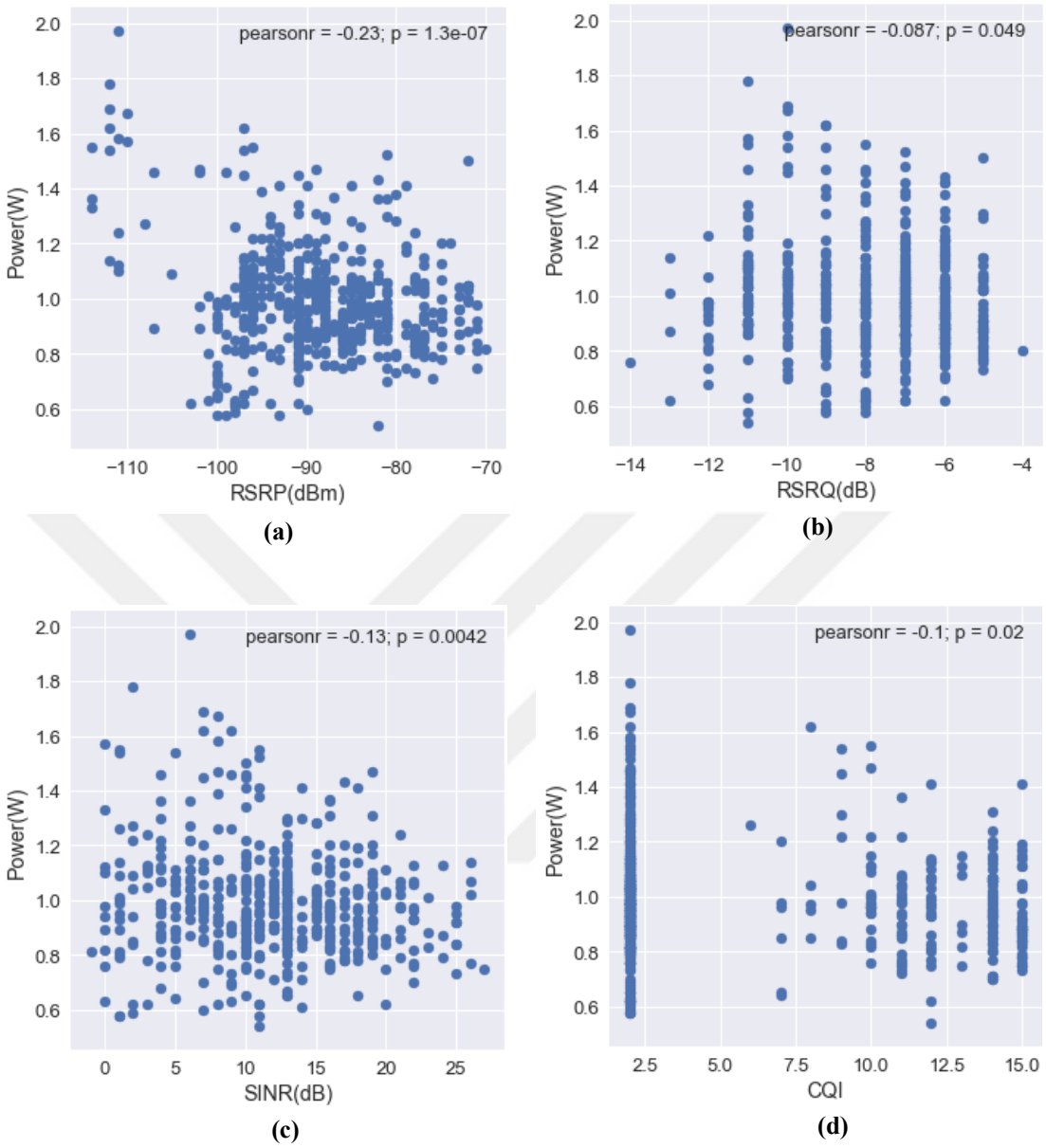


Figure 3.10 Relationship between power and LTE signal metrics

Figures 3.11.a, 3.11.b, 3.11.c, and 3.11.d show the relations between energy consumption and LTE metrics. It is seen that there exist negative correlations between the energy consumption and LTE metrics. The energy consumption is moderately correlated with RSRP, RSRQ, and SINR, and it has too low correlation with CQI. We say that the energy consumption increases as RSRP, RSRQ and SINR values become lower, but the increment on the energy consumption is not always linear. The increment on the energy consumption is faster after the certain values of the LTE metrics.

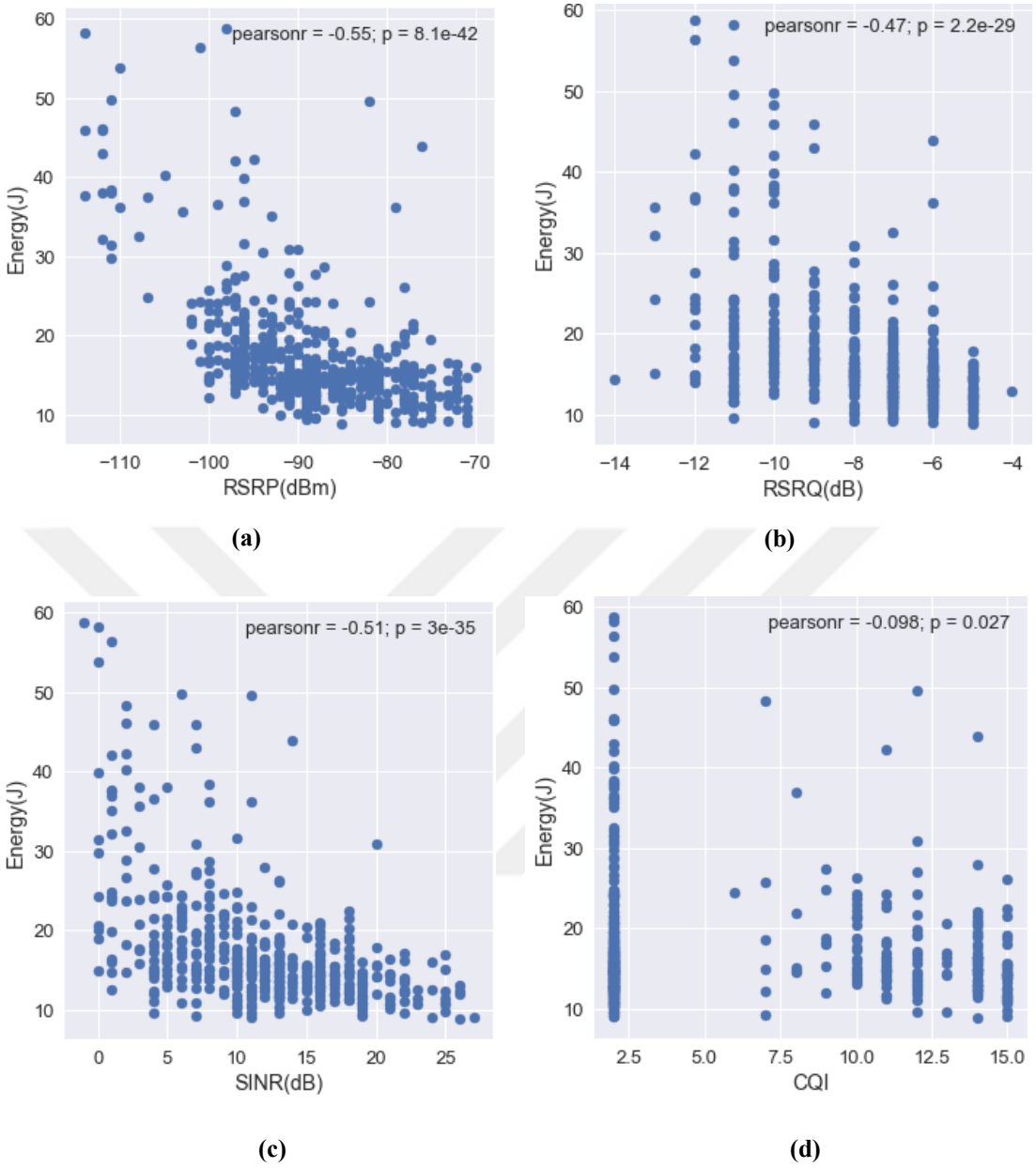


Figure 3.11 Relationship between energy and LTE signal metrics

Figures 3.12.a, 3.12.b, 3.12.c, and 3.12.d illustrate the relationship between data rate and LTE metrics. Contrary to the energy consumption, LTE metrics (excluding CQI) and data rate are positively correlated. This means that the data rate will increase as LTE metrics increase. The data rate is moderately correlated with RSRP, RSRQ, and SINR. CQI is not correlated with data rate (the correlation coefficient is close to zero). Also that the significance value of the correlation coefficient is greater than 0.05 shows that the correlation between CQI and data rate is not meaningful.

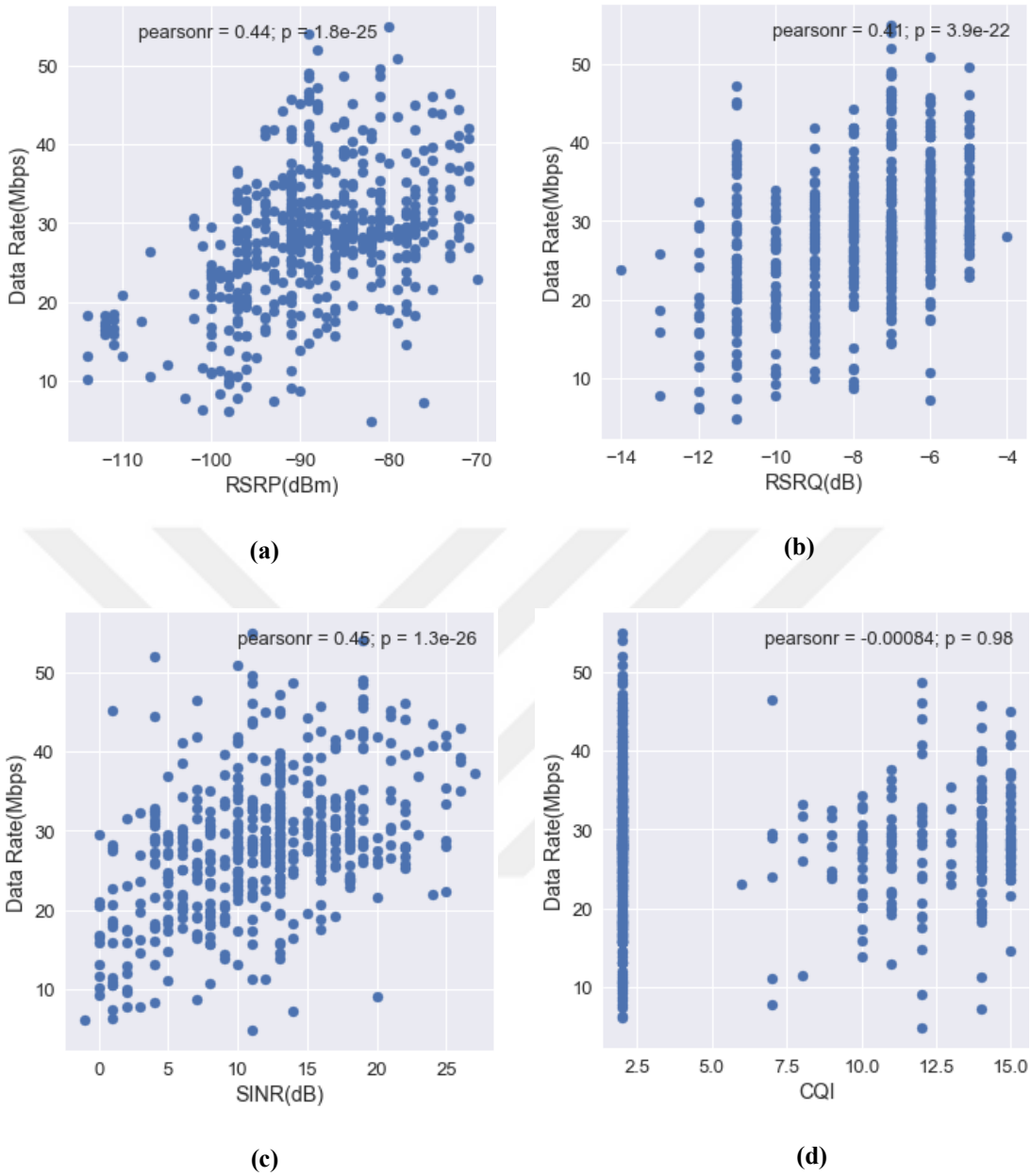
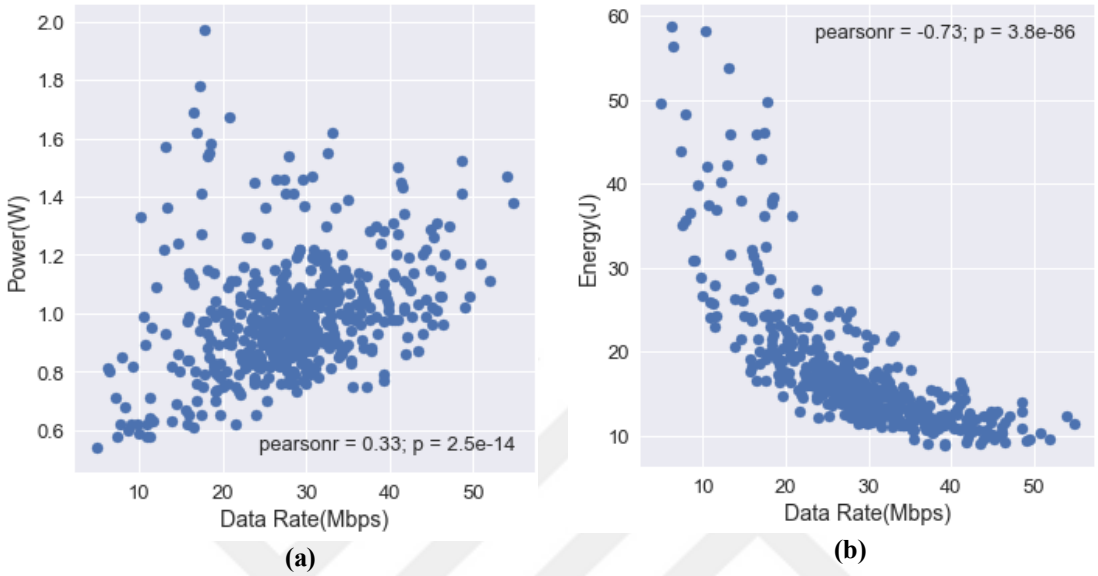


Figure 3.12 Relationship between data rate and LTE signal metrics

The relationship between data rate and power spent and between data rate and energy consumption is illustrated in Figures 3.13.a and 3.13.b. The data rate is strongly correlated with energy and there is an (negative) exponential relationship between them. Also, the data rate is positively correlated with power spent while it is negatively correlated with energy consumption. This is due to the fact that if the instant amount of data downloaded is more (that is, if the data rate is high), the power consumption increases but the total energy consumed is lower because the total downloading time is short. Also

while downloading data, power consumption initially increases rapidly because a number of signaling processes occurs to perform the requested operation. Therefore, as data rate increases, energy consumption decreases but power consumption increases.



(a) (b)

Figure 3.13 Relationship of data rate to power and energy

According to these experiments results, we say that LTE metrics are more correlated with energy consumption and data rate rather than the power spent. Also, data rate has a strong correlation with the energy consumption. By using these test results, we model energy and data rate of a standalone node in Section 3.2.5.1.

3.1.3.3. Standalone versus Gateway experiments

We examine energy consumption and data rate of a device acting as a standalone node (data download for itself) and a gateway node (data download for another device), and we also examine the effect of WiFi signal strength and link speed between gateway and client node. In our test scenario, Device A (Samsung Galaxy SM-J7000F) has used as both standalone and gateway node while the Device B (Samsung Galaxy SM-J700H) is only used as a client node. We have two different scenarios for gateway-client experiments. The first scenario is that the gateway (Device A) and client (Device B) nodes are side by side and there are no obstacles between them. Therefore, there exists a strong WiFi link between them represented by a straight line as shown in Figure 3.3 Part (2). In

the second scenario, the distance between the gateway node and the client node is too large or there can be an obstacle between them. The second scenario is not illustrated, but it can be considered as if there exists a dashed line between the gateway node and client node, which indicates the weakness of the WiFi link between them.

In order to compare the standalone and gateway nodes in term of energy and data rate, we performed three different data download tests. Firstly a node downloads 50 MB data file for herself (standalone), then the same node acts as a gateway node (Gateway-NC) downloads 50 MB data file for the near client (NC) node that has a strong WiFi connection with the gateway, and lastly the node again acts as a gateway node (Gateway-FC) but she downloads 50 MB data file for the far client (FC) node (this means that there exists weak WiFi link between them). We assume that the 4G signal parameters do not change during these triple tests (Standalone, Gateway-NC, and Gateway-FC). According to this scenario, we have made about 100 triple tests. During each triple test, LTE signal strength and other LTE signal metrics are approximately the same as seen in Figure 3.14.

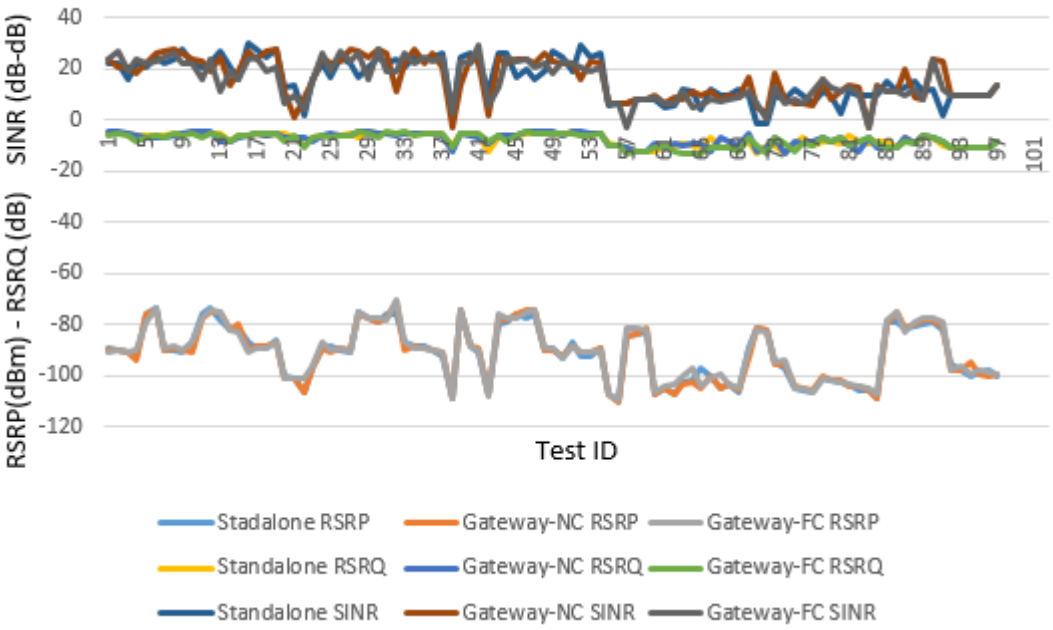


Figure 3.14 The average values of LTE signal metrics during the tests for a standalone and a gateway node

Figure 3.15 shows average WiFi signal strength and link speed between the gateway and the client nodes during these tests. For the Gateway-NC tests, the WiFi signal strength

varies between -40 dBm and -10 dBm, and the WiFi link speed between the gateway and the client is maximum. However, the WiFi signal strength weakens and the average value drops to about -60 or -70 dBm for the Gateway-FC tests. Therefore, the link speed between the gateway and the client nodes is also reduced. This also negatively affects the data rate and energy consumption of the gateway node while relaying data.

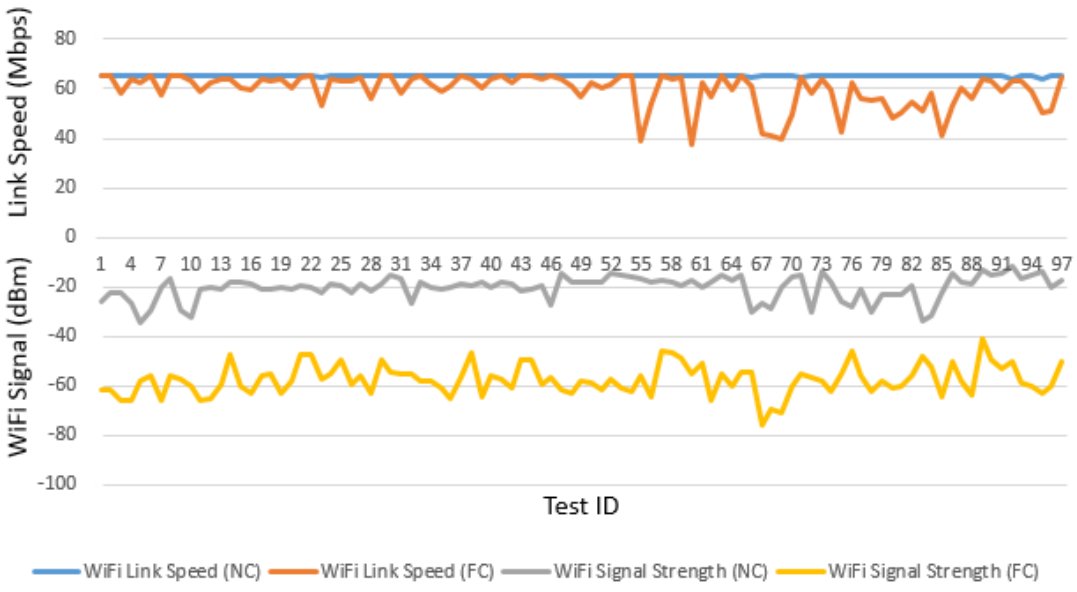


Figure 3.15 The average values of WiFi signal strength and link speed according to the near and far client node

Figures 3.16 and 3.17 indicate how data rate and energy consumption change when a node is in different roles such as a standalone node and a gateway node (relaying to near/ far-client). According to the results, the data rate of a standalone node is about 1.6 ∓ 0.49 times faster than the data rate of a Gateway-NC node that downloads data for a nearby client −this ratio may vary depending on cellular/WiFi hardware of devices−. This is because relaying takes time even if the connection between the gateway and the client nodes is very strong so this is reduced download data rate. If a gateway node (Gateway-FC) services a far client, the bandwidth between the far client and the gateway node is not fully utilized due to the weakness of the WiFi link, so the data rate is further reduced. Therefore Gateway-NC is approximately 1.3 ∓ 0.67 times faster than Gateway-FC as illustrated in Figure 3.16.

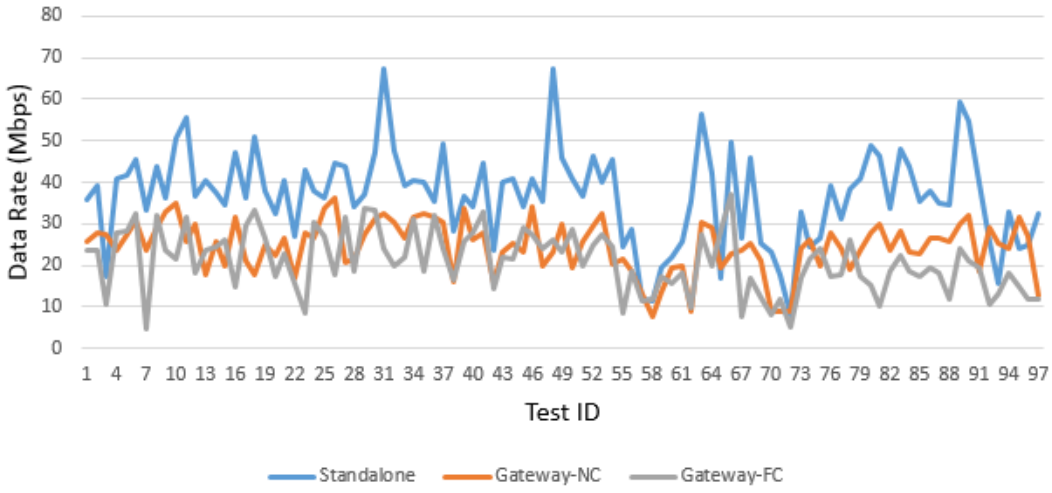


Figure 3.16 The average values of the data rate for a standalone and a gateway node

As shown in Figure 3.17, the gateway nodes consume more energy than the standalone node because they work as a hotspot. Gateway-NC node consumes about 2.5 ∓ 0.72 times more energy than the standalone node, and also Gateway-FC node consumes about 1.2 ∓ 0.4 times more energy than Gateway-NC node by reason of serving the far client node that has a weak WiFi connection.

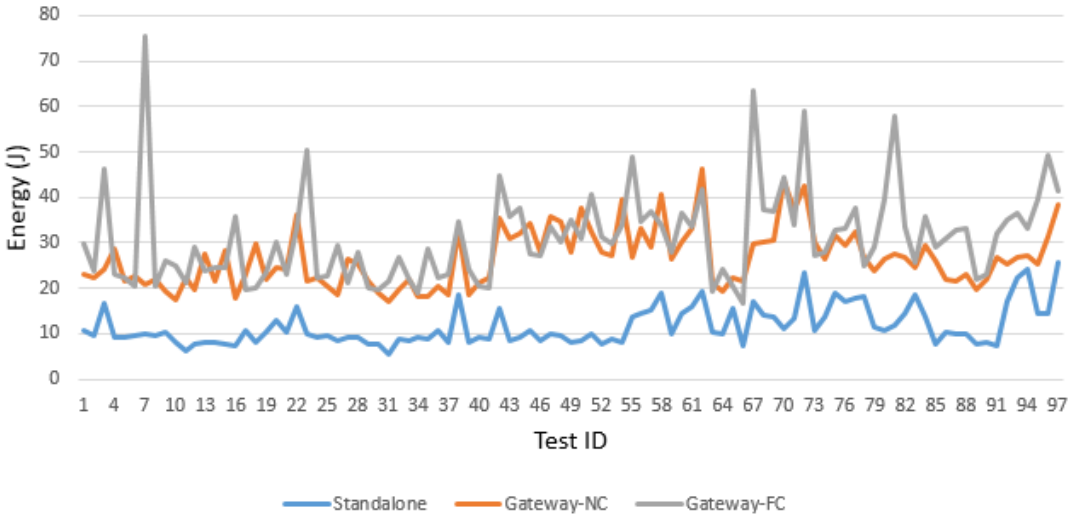


Figure 3.17 The average values of the energy consumption for a standalone and a gateway node

3.1.3.4. Standalone versus Client experiments

In this part, we have studied two different test scenarios where a node accesses the Internet directly through a base station or indirectly by a gateway node. Our purpose of doing these tests is to prove that a node, which has weak signal parameters on LTE networks, can gain more benefit in case it participates a user-provided network system and connects to the Internet via a gateway node with strong signal parameters. Our test scenarios are illustrated in Figure 3.3 Part (3). Device A represents a gateway node that has strong 4G signal parameters which are indicated by a straight line between the base station and Device A. Device B is a node (that may act as a standalone and a client node) that has a weak LTE signal strength represented by a dashed line between the base station and itself. The link between Device A and Device B represents WiFi connection and it is shown a straight line because WiFi signal strength between them is strong. According to our test scenario, firstly Device B acts as a standalone node and downloads 50 MB data file via own LTE network connection and then she acts as a client node and downloads the same data file via gateway Device A. Next, we compare two test cases in terms of data rate and energy consumption. We performed the experiments on LG G3 (Device A) and Asus Zenfone 2 (Device B) smartphones. As seen in Figure 3.18, SINR and RSRQ of the client node show similar behavior with LTE signal strength (RSRP).

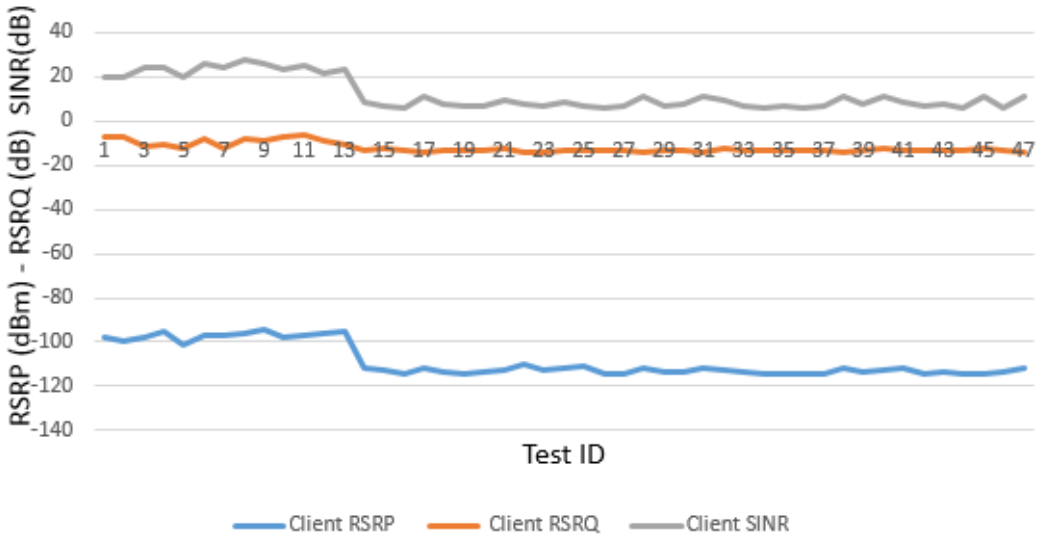


Figure 3.18 The average values of LTE signal metrics for the client node

Figure 3.19 shows averages of WiFi signal strength and WiFi link speed between the gateway and the client nodes for each test. WiFi signal strength ranges from -20 dBm to-40 dBm and WiFi link speed generally lies on 72 Mbps band.

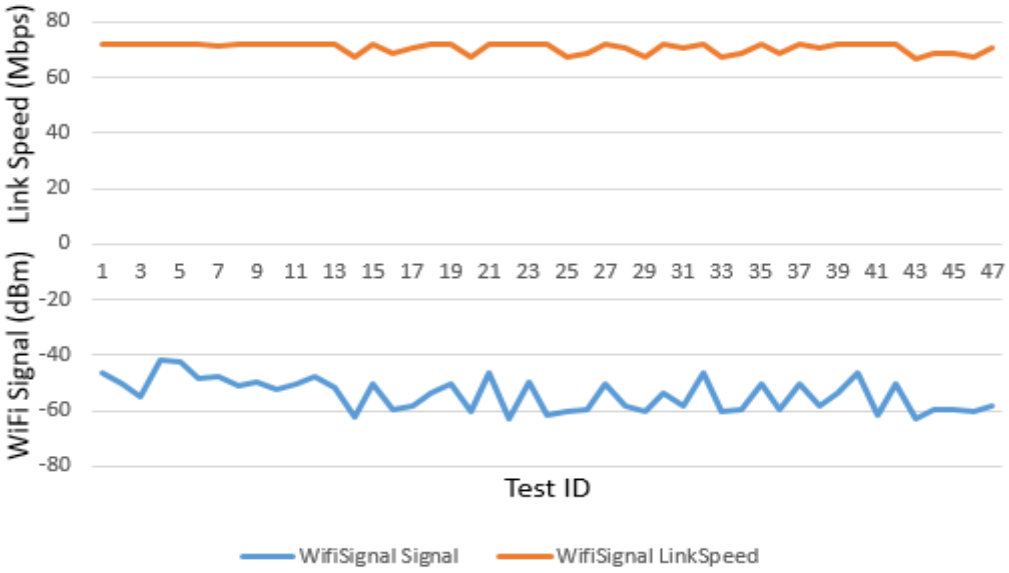


Figure 3.19 The average values of WiFi signal strength and link speed

Figure 3.20 shows the data rates of a node that acts as both a standalone node (downloading data using own internet connection) and a client node (downloading data via a gateway node). The standalone data rate of the client node is higher than the data rate that the node accesses the Internet through the gateway if the signal strength (RSRP) of the client node is higher than -100 dBm and close to the signal strength of the gateway node of whom the client uses the Internet connection. As also seen in Figure 3.20, it is more profitable for the client node to access the Internet through a gateway node instead of using its own connection providing that the signal strength of the client is less than -100 dBm and the gateway signal strength is greater than -100 dBm. In that case, the data rate of the client which uses the network connection of the gateway node is 1.8 ∓ 0.34 times as higher than the standalone data rate of the client.

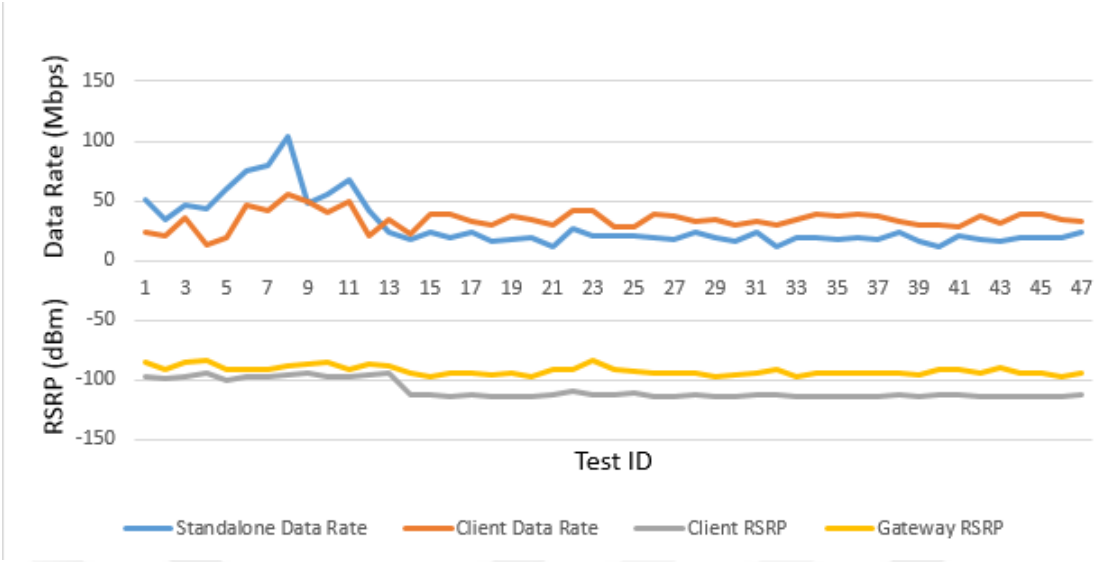


Figure 3.20 The average values of the data rate for a standalone and client node, and the average values of RSRP signal strength for the client and gateway nodes

In addition, as illustrated in Figure 3.21, the client node consumes more energy (about 2.7 ∓ 0.68 times) if it uses own cellular Internet connection when the signal strength is lower than -100 dBm. This is due to low data rates experienced at low signal strengths (below -100 dBm) and low power efficiency of LTE compared to WiFi.

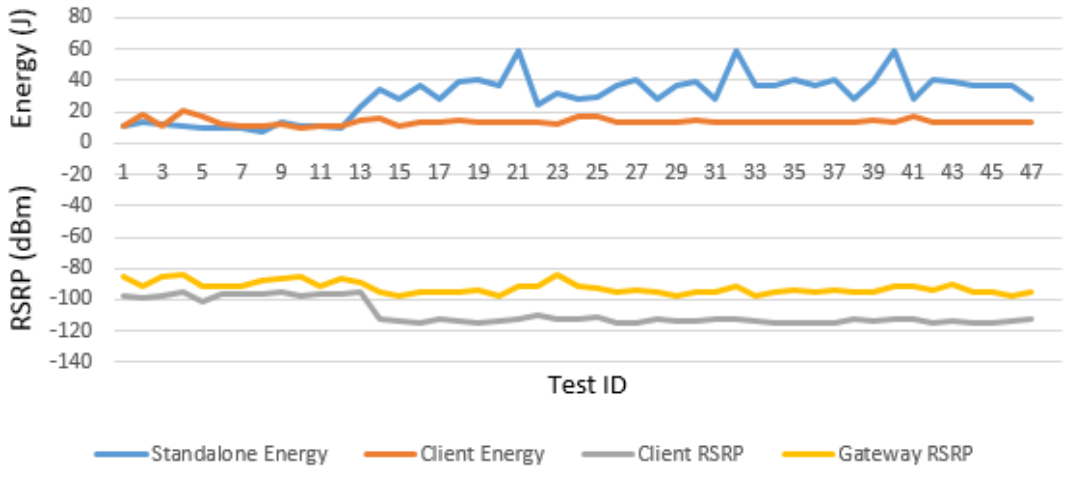


Figure 3.21 The average values of the energy consumption for a standalone and a client node, and the average values of RSRP signal strength for the client and gateway nodes

3.2. Regression Analysis

In this subsection, we briefly explain linear regression assumptions, applied regression techniques, and regression evaluation criteria. Then we present two data separation techniques and we introduce energy and data rate models for standalone and gateway nodes.

3.2.1. Linear regression assumptions

Multiple linear regression is a widely used method to estimate the relationship between a dependent variable and one or more independent variables. In other words, it provides us to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. In order to be able to apply properly the multiple linear regression analysis and correctly interpret the analysis outputs, it is necessary to satisfy some assumptions [32, 33]. These assumptions are;

Linearity. There must be a linear relationship between a dependent variable (DV) and independents variables (IVs). More specify, dependent variable must be a linear function of an independent variable(s). If the linearity assumption is violated, the estimates such as regression coefficients, standard errors and test of statistical significance may be biased. The violations in the linearity assumption can be detected with LOWESS (Locally Weighted Scatterplot Smoothing) curve [32] that is a method of creating a smooth line that represents the relationship between independent variable(s) and dependent variable in a scatterplot. The LOWESS line looks linear (or rough linear) if the relationship between the dependent variable and independent variable(s) is linear (or approximately linear).

Independence of Residuals (IoR). There should not be a meaningful relationship between the residuals. If it is present autocorrelation (serial correlation) between the residuals, the standard errors and significance tests will be affected by that and they will not be accurate. The violations of the assumption can be diagnosed by Durbin-Watson test [34]. Durbin-Watson value lines between 0 and 4. The best value is 2 which means that there is no autocorrelation between the residuals. The values under 1 or more than 3 are a definite cause for concern [35].

Constant Variance (Homoscedasticity). The residuals should be constant variance (CV) for all values of IVs. This means that the variance of the dependent variable should not change depending on values of IVs. When there exists changing variance (heteroscedasticity) problem in a regression model, the estimates of standard errors and hence significance tests and confidence intervals will be incorrect. Homoscedasticity can be visually checked via scatterplots. Ideally, if a plot of standardizes residuals by standardized predicted values is randomly scattered around zero line, it can be said that the homoscedasticity assumption is met. Another way of checking the constant variance assumption is to apply Levene's Test [36]. In the Levene's, the residuals are initially divided into two or more groups, and a value is calculated to assess the equality of variances for the groups. If the significance value (p-value) of Levene's test is greater than 0.05, equal variance assumption is satisfied.

Normality. The residuals have normal distribution. Violations of the normality assumption do not cause biased estimation of regression coefficients. However, they can affect the estimates of the significance tests and confidence intervals depending on the sample size. The normality assumption can be checked via histograms of the standardized residuals, Q-Q and P-P plots [37]. Also, the normality of the residuals can be detected statistically by Jarque Bera test [38]. According to this test, if the significance value (p-value) is greater than 0.05, the assumption is satisfied.

Multicollinearity. IVs are not highly correlated with each other. Multicollinearity is analyzed through collinearity statistics such as Condition Index (CI). CI is a statistic for diagnosing of the multicollinearity. If the CI is above 30, it indicates highly severe problems of multicollinearity. Some authors have proposed values of 15 or 20 as a threshold value for the CI [32].

In a linear regression model, independent variables can affect each other. Therefore, the interaction of independent variables should be examined and if there is an interaction between the variables (namely if the significance value of the interaction term is less than 0.05), the interaction term that is the multiplication of two independent variables should be included in the regression model.

3.2.2. Applied regression techniques

In our study, we have applied some regression analysis techniques such as Ordinary Least Squares (OLS) regression, robust regression, and some other supervised learning techniques.

OLS regression is a generalized linear modeling technique that determines the best fit to the data points [39]. OLS try to minimize the sum of squares of all the residuals (differences between the observed and predicted values). This technique is sensitive to the assumptions mentioned in Section 3.2.1. Violations of some of the assumptions can lead to biased estimates of regression coefficients, incorrect standard errors and incorrect significance tests.

Robust regression is an alternative approach to OLS regression when some assumptions are violated. Robust regression techniques are not as vulnerable as OLS regression techniques against violations of assumptions. They mitigate the effects of violations of assumptions on the estimates. Robust regression techniques [40] can be used when data contain outliers or there is a strong suspicion of heteroscedasticity. Theil-Sen [41] , RANSAC [42] and Huber [43] techniques are robust regression techniques. Theil-Sen is a median based estimator. It fits a line to data points by choosing the median of the slopes of the all lines through pairs of points. It copes with medium-size multivariate outliers, but the robustness of the estimator decreases quickly in large dimensional problems. This estimator has no any assumption about data distribution because of being a non-parametric method, so it can be used for heteroscedastic data. RANSAC (RANdom SAmple Consensus) is an iterative technique that estimates model parameters from a subset of observed data. RANSAC regressor selects randomly minimal samples from the input dataset and then fits a model to the random subset. Then it splits dataset as inliers and outliers using the fitting model. It generates candidate solutions by repeating the steps until reaching maximum number of iterations or other special criteria. Then it chooses the best model among the candidate solutions, which has contains maximum inliers. RANSAC deals better with large outliers in the y-direction. Therefore, it also can be interpreted as an outlier detection method. Huber regressor is different from Theil-Sen and RANSAC regressors. Because it does not filter completely the outliers but gives a lesser weight to them and thus it reduces the influence of the outliers and it is less affected

by very large outliers. Apart from these, we used other common machine learning methods such as Decision Tree [44], Random Forest [45], K-Neighbors [46] and Gradient Boosting [47] in order to evaluate the models performance. Decision trees are non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting. K-Neighbors regressor implements learning based on the k nearest neighbors of each query point, where k is an integer value specified by the user. Gradient Boosting regressor builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage a regression tree is fit on the negative gradient of the given loss function.

3.2.3. Evaluation criteria and standardization

In order to measure regression models performance and compare model results, we have used R square (R^2) and root mean square error (RMSE) evaluation metrics. These metrics indicate the performance of the model and reveal us about prediction error of the model.

R square ($\boldsymbol{R^2}$) [48] is a statistical measure that indicates a regression score function. The best possible score is 1.0 and it can be negative. It is the proportion of the variance in the dependent variable that is predictable from the independent variable(s) in the sample. It provides a measure of how well future samples are likely to be predicted by the model [49]. R square function is defined as

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n_{samples}-1}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{samples}-1}(y_i - \bar{y})^2} \qquad (3.1)$$

where $\bar{y} = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} y_i$. $\hat{y}_i$ is the predicted value of the i-th sample and y_i corresponds to true value.

Root mean square error (RMSE) is a model performance metric that shows how residuals spread out around a line of the best fit. It is the square root of the mean of the squares of the difference between the values predicted by a model and the values actually observed [50]. The function of RMSE is defined as

$$RMSE\ (y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{n_{samples}-1}(y_i - \hat{y}_i)^2}{n_{samples}}} \qquad (3.2)$$

Standardization is widely used in data analysis. It also is a requirement for many machine learning estimators. Raw data can contain features with different ranges and standardization helps to create features that have similar ranges to each other. Also, it reduces the risk of multicollinearity in a regression model. z-score is a standardization technique [32, 51], it is also used for extreme value analysis. The function of z-score is

$$Z = \frac{x - \mu}{\sigma} \qquad (3.3)$$

where x is a value of raw data, μ is mean of population, and σ is standard deviation of population.

3.2.4. Data preprocessing

While calculating energy and data rate, we exclude the TCP slow start phase in which data rate is low but energy per data is too high. This is important for obtaining consistent results. Also, we have to exclude the final part, where data download is almost completed but there are some negligible activities. For this purpose, we have proposed two data separation techniques in order to divide data into parts.

The first technique is based on energy consumed per data that divides the data download process into three parts as *head*, *main* and *tail* as shown in Figure 3.22.a. Head part is considered as TCP slow-start phase where data rate is too low but energy per data is too high. In the main part, data flow is high and energy per bit is lower than the head and tail part. This part gives more accurate information about energy consumption and data rate. Therefore we have used the main part of the collected data while modeling energy and data rate. The tail part is the end of the data download process but small data activities can still exist as shown in Figure 3.22.b. In the tail part, the power consumption

is close to zero so this part is neglected. Separation of head, main and tail parts is done as follows. Firstly, the tail is separated from the main part by detecting a long silent period after the completion of data download. After excluding the tail part, the head and main parts are separated from each other by the following process. Firstly, we set a threshold value as the overall energy over overall downloaded data (mJ/kb) during the whole time period. Then starting from the beginning, we check energy consumption per data at each time slot of 1 second. Typically, this value is initially high (due to warm-up) and then decreases sharply. After the first time slot where it is encountered that the energy per data is lower than or equal to the threshold, head part ends up and main part starts as shown in Figure 3.22.c.

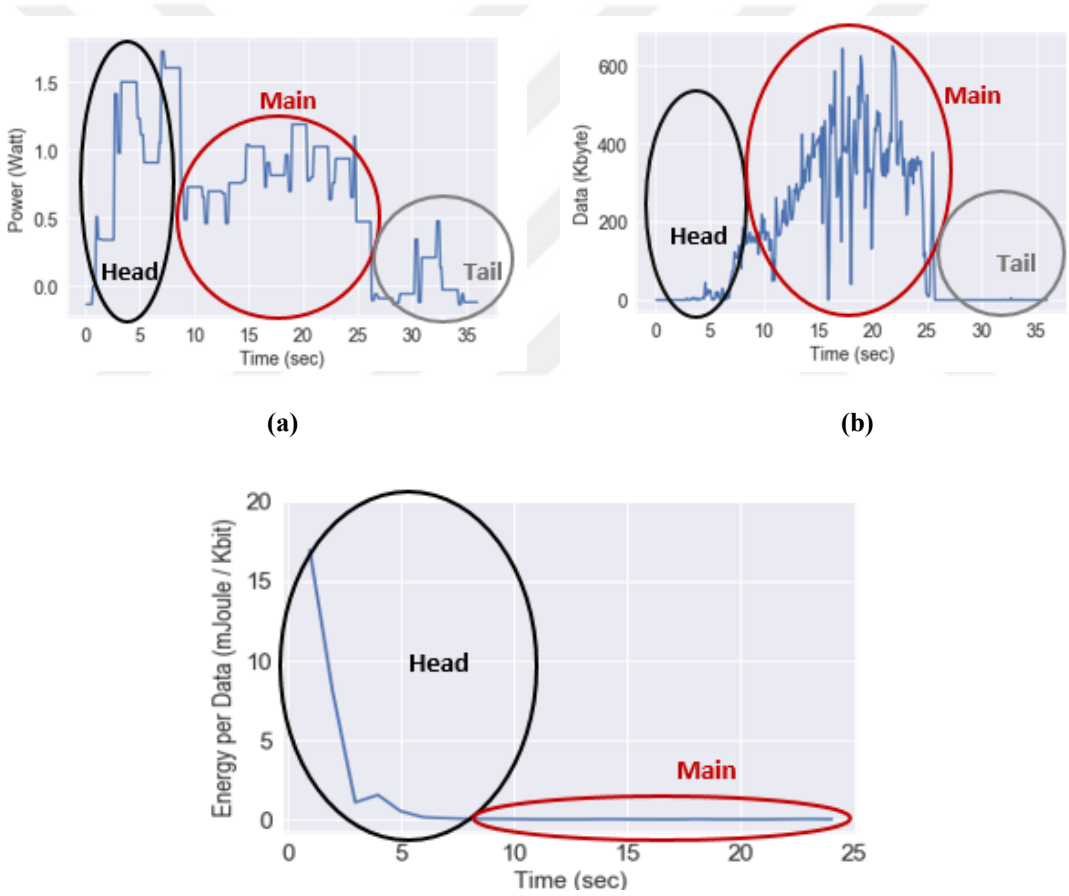


Figure 3.22 Separation of head, main and tail parts of a data transfer period

According to the first technique, we can set a specific threshold value (energy per data) for each data download process while dividing data into the parts because we know

the entire data download process. However, in order to use this technique on streaming data, we need to set a general threshold value based on energy per data, but this is not easy due to the fact that smartphones have different measurement sensitivity. Some smartphones could provide enough fine-grained measurements about power and energy while others could provide coarse measurements, so the average threshold value differs from device to device. Since the threshold value to be determined does not specify an overall average value for all devices, segmentation of data download process will not be done correctly on some devices. Therefore, we have proposed another technique that is a generalized method for all smartphones. This technique is based on received (or sent) data amount per time. It is used for separating data downloading process into states while data is flowing. The details of this technique were described in Section 2.2.3 (Data Transfer) and the parameter values used are given in Table 3.20 in Section3.3.

We have compared the results of two separation techniques. For this purpose, we have applied both data separation techniques separately on about 400 data download tests and we find that approximately 95 % of the results (for the main part) are similar. The results of main part of the energy and data rate with respect to the separation techniques are shown in Figure 3.23. We use the first (in Section 3.2.5.1) and the second (in Section 3.2.5.2) techniques for the separation of experimental data into the states before modeling.

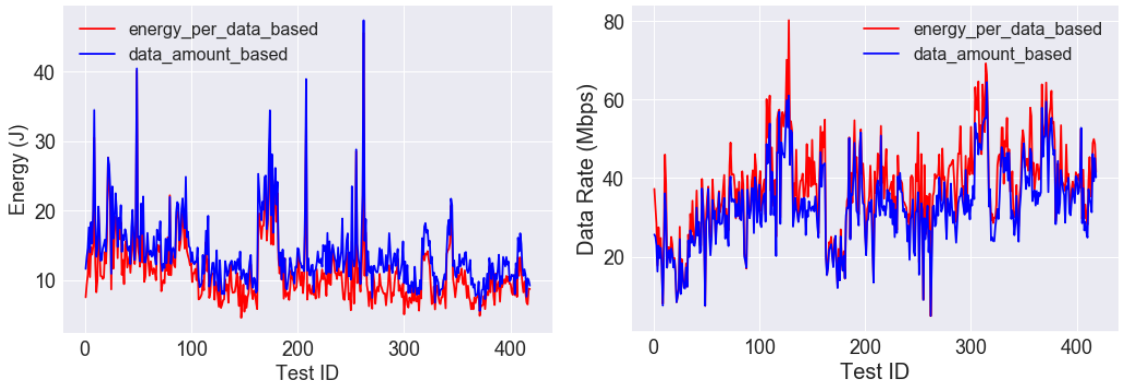


Figure 3.23 Comparison of the separation techniques

3.2.5. Modeling

We present energy consumption and data rate models for the standalone (in [24]) and the gateway nodes by using different regression techniques. We have used Python's the statsmodels library [52] for Ordinary Least Square (OLS) regression method and the linear_model module of the scikit-learn library [53] for robust regression methods (Theil-Sen, RANSAC and Huber Regressors). In the built models, all independent variables are standardized with z-score for reducing multicollinearity. Standardized variables are represented as IV^z which is standardization of independent variable. Also, we apply other supervised learning methods such as Decision Tree regressor, Random Forest regressor, K-Neighbors regressor and Gradient Boosting regressor with default parameters by using the tree [54], neighbors [55] and ensemble [56] modules in the python scikit-learn library.

3.2.5.1. Models for standalone case

We performed 430 tests (excluding outliers) with various received signal characteristics in various locations. We measured energy consumption, average data rate and several LTE parameters such as *RSRP, RSRQ,* and *SINR* during the tests. Also, we measured CQI metric but the CQI values were not included in the dataset because we could not obtain significant results about CQI metric. Some descriptive statistics about collected data are given in Table 3.3. The statistical information such as mean and standard deviation are important for the standardization of the parameters and interpretation of models. During the modeling, standalone energy consumption and standalone data rate of a node *i* will be represented as E_i^{sta}, R_i^{sta} respectively.

Table 3.3 Descriptive statistics of data for the standalone case

Descriptive Statistics					
	N	***Min.***	***Max.***	***Mean***	***Std. Dev.***
Energy (mJ/Mbit)	430	11.78	47.27	24.83	6.75
Log(Energy)	430	2.47	3.56	3.18	0.26
Data Rate (Mbps)	430	12.12	66.28	39.12	10.10
RSRP(dBm)	430	-106	-70	-86.84	7.77
RSRQ (dB)	430	-14	-4	-7.55	1.97
SINR(dB)	430	0	30	13.10	6.20

In Table 3.4, correlations among the measured values are given. According to the table, energy is moderately correlated with LTE metrics but it has quite strongly correlated with data rate. Also, the data rate is less correlated with LTE metrics according to the energy.

Table 3.4 Correlations of data for the standalone case

Correlations							
		Energy	Log (Energy)	Data Rate	RSRP	RSRQ	SINR
Energy	Pearson Corr.	1	.987**	-.799**	-.526**	-.508**	-.563**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
Log (Energy)	Pearson Corr.	.987**	1	-.807**	-.531**	-.498**	-.556**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
Data Rate	Pearson Corr.	-.799**	-.807**	1	.415**	.377**	.413**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
RSRP	Pearson Corr.	-.526**	-.531**	.415**	1	.510**	.355**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
RSRQ	Pearson Corr.	-.508**	-.498**	.377**	.510**	1	.521**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
SINR	Pearson Corr.	-.563**	-.556**	.413**	.355**	.521**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
** Correlation is significant at 0.01 level (2-tailed).							

In Figure 3.24, the distribution and relationships of the collected data are shown. Figures in the first column show the distribution of energy, the log of energy and data rate. Other columns illustrate the relationship between energy and other variables, the log of energy and other variables and the relationship between data rate and other variables. According to these figures, we can say that energy variable has a lognormal distribution, data rate has a normal distribution and there is an exponential relationship between these two variables. As described in the next section, we will use natural logarithm of energy in our modeling henceforth. It can be said that the relationship between logarithm of energy and signal variables, and the relationship between data rate and signal variables are linear or approximately linear.

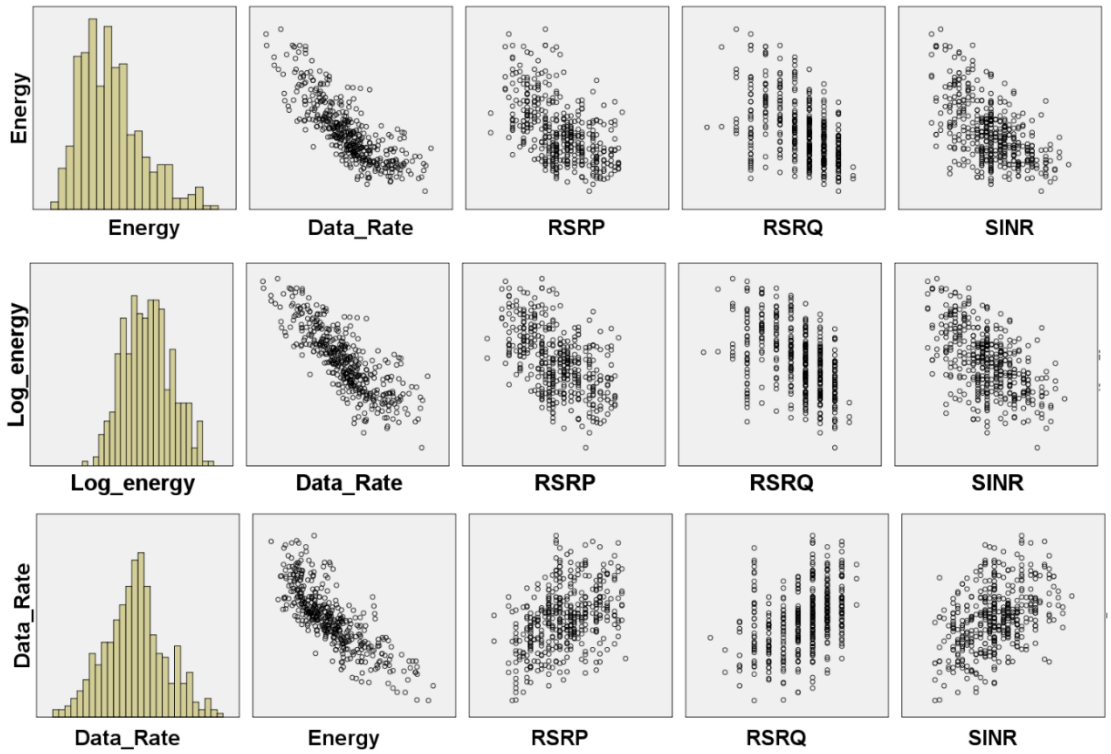


Figure 3.24 Data distribution and relationships for the standalone case

OLS regression models

Using the collected data described above, we apply OLS method for modeling the energy consumption and average data rate with respect to LTE signal parameters. The energy variable is transformed by taking natural logarithm for satisfying the linearity assumption and eliminating heteroscedasticity. Significance values (p-values) [57] of the coefficients of the variables in the models are less than the cutoff (0.05), except for the coefficient of RSRQ in the data rate model. Also, the confidence interval is 95% for all the models.

Energy consumption modeling

For energy consumption, we have basically proposed two models: Energy Consumption Modeling without Data Rate and Energy Consumption Modeling with Data Rate.

Energy consumption modeling without data rate

Firstly, the energy consumption E_i^{sta} is modeled according to *RSRP, RSRQ* and *SINR* variables without data rate variable *R*. In Table 3.5, it is given the model formula, R^2 and RMSE values for each step of the energy consumption model. We have added the variables in the model step by step according to their contributions to R^2 and finally, the interaction terms are included in the model. It has been checked whether the models satisfy the assumptions given in Section 3.2.1, and the results are shown in Table 3.6.

Table 3.5 Model summary of the standalone energy consumption without data rate

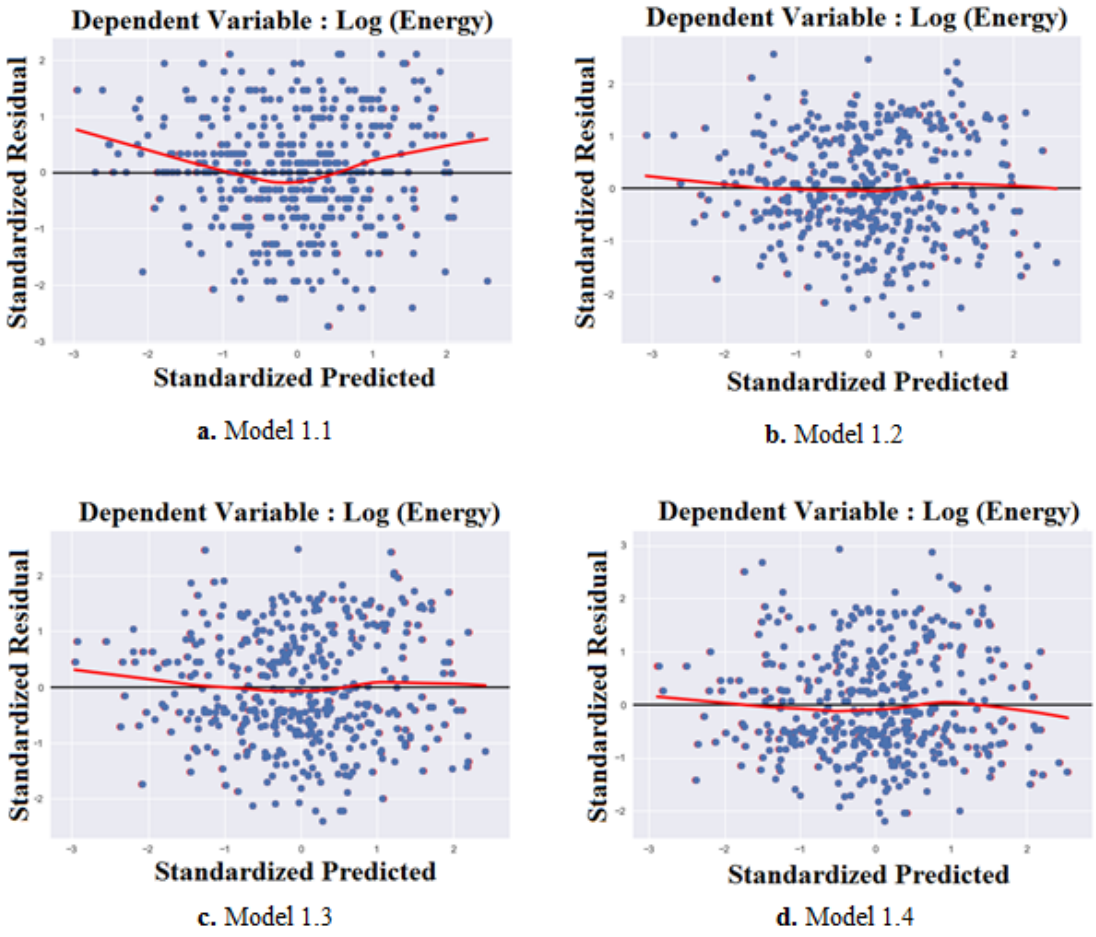
Model	**Formulas ($\log(E_i^{sta})$)**	$\boldsymbol{R^2}$	**RMSE**
1.1	$3.177 - 0.146 * \mathrm{SINR}^{Z}$	0.310	0.218
1.2	$3.177 - 0.110 * \mathrm{SINR}^{Z} - 0.100 * \mathrm{RSRP}^{Z}$	0.437	0.197
1.3	$3.177 - 0.096 * \mathrm{SINR}^{Z} - 0.087 * \mathrm{RSRP}^{Z} - 0.036 * \mathrm{RSRQ}^{Z}$	0.449	0.195
1.4	$3.162 - 0.102 * \mathrm{SINR}^{Z} - 0.085 * \mathrm{RSRP}^{Z} - 0.027 * \mathrm{RSRQ}^{Z} + 0.029 * \mathrm{RSRQ}^{Z} * \mathrm{SINR}^{Z}$	0.460	0.193

Table 3.6 Model assumptions for the standalone energy consumption without data rate

	Assumptions						
Model	**Linearity (Lowess Curve)**	**IoR (D-W Test)**	**CV (Levene's Test)**		**Normality (Jarque Bera)**		**Multi-collinearity (CI)**
			Stat	**Sig.**	**Stat**	**Sig.**	
1.1	Partially violated	1.32	0.93	0.33	3.68	0.16	1.00
1.2	Linear	1.45	0.00	1.00	0.97	0.62	1.45
1.3	Linear	1.44	0.17	0.68	1.76	0.42	2.12
1.4	Linear	1.50	0.31	0.58	1.18	0.56	2.33

It can be said that the linearity assumption is met for all other models except the Model 1.1. LOWESS lines for the models (1.2, 1.3 and 1.4) look fairly linear as seen in Figure 3.25.b, 3.25.c, and 3.25.d respectively, but for the Model 1.1, it is seen that the linearity is partially violated as shown in Figure 3.25.a. The assumption of independence

of the residuals is satisfied for all models since the Durbin-Watson value for any model is not below 1 or above 3. The assumption that the residuals have equal variance is not violated because the significance values of the Levene's test are greater than 0.05 for all models. Due to the fact that the significance values of the Jarque-Bera test for all models are greater than 0.05, we can say that the models have normal distributions and there is no risk of multicollinearity in the models because the variables are standardized, so the CI values are below the threshold value.



a. Model 1.1 **b.** Model 1.2

c. Model 1.3 **d.** Model 1.4

Figure 3.25 Scatterplots of the standalone energy consumption models without data rate

Energy consumption modeling with data rate

In this part, energy consumption E_i^{sta} is modeled with Data Rate R, RSRP, RSRQ and SINR variables. The goal of this model is to estimate how much energy is consumed if the actual data rate as well as RSRP, RSRQ, and SINR values is known. For this reason,

the actual data rate R values measured were used in the model instead of the estimated data rate. In Table 3.7, it is given the regression steps. In each step, a variable is included in the regression model according to its contribution to R^2, and for this way, we have obtained four regression models (2.1, 2.2, 2.3, and 2.4). Then, the interaction term is added in the model (2.5). Linear regression assumptions are checked and results are illustrated in Table 3.8.

Table 3.7 Model summary of the standalone energy consumption with data rate

Model	**Formulas ($\log(E_i^{sta})$)**	R^2	**RMSE**
2.1	$3.177 - 0.212 * \mathrm{R^Z}$	0.651	0.155
2.2	$3.177 - 0.183 * \mathrm{R^Z} - 0.071 * \mathrm{SINR^Z}$	0.711	0.141
2.3	$3.177 - 0.167 * \mathrm{R^Z} - 0.060 * \mathrm{SINR^Z} - 0.049 * \mathrm{RSRP^Z}$	0.739	0.134
2.4	$3.177 - 0.165 * \mathrm{R^Z} - 0.053 * \mathrm{SINR^Z}$ $-0.042 * \mathrm{RSRP^Z} - 0.020 * \mathrm{RSRQ^Z}$	0.742	0.133
2.5	$3.170 - 0.164 * \mathrm{R^Z} - 0.054 * \mathrm{SINR^Z}$ $-0.041 * \mathrm{RSRP^Z} - 0.017 * \mathrm{RSRQ^Z}$ $+ 0.017 * \mathrm{DataRate^Z} * \mathrm{SINR^Z}$	0.746	0.132

Table 3.8 Model assumptions for the standalone the energy consumption with data rate

	Assumptions						
Model	**Linearity (Lowess Curve)**	**IoR (D-W Test)**	**CV (Levene's Test)**		**Normality (Jarque Bera)**		**Multi-collinearity (CI)**
			Stat	**Sig.**	**Stat**	**Sig.**	
2.1	Approx. Linear	1.10	2.24	0.14	5.27	0.07	1.00
2.2	Approx. Linear	1.40	2.28	0.60	6.92	0.03	1.55
2.3	Approx. Linear	1.48	0.07	0.79	3.33	0.19	1.78
2.4	Approx. Linear	1.48	0.05	0.82	3.39	0.18	2.36
2.5	Approx. Linear	1.49	0.04	0.85	3.69	0.16	2.40

It is observed that linearity assumption is not violated for any model since the LOWESS fit line does not exhibit any large deviations from the 0-line (see the figures 3.26.a, 3.26.b, 3.26.c, and 3.26.d) The residuals of the models are independent because

their Durbin-Watson values are not under to 1 for any model. Also, the constant variance assumption is satisfied for all models. All models (except Model 2.2) have normal distribution and there is no multicollinearity problem for any model.

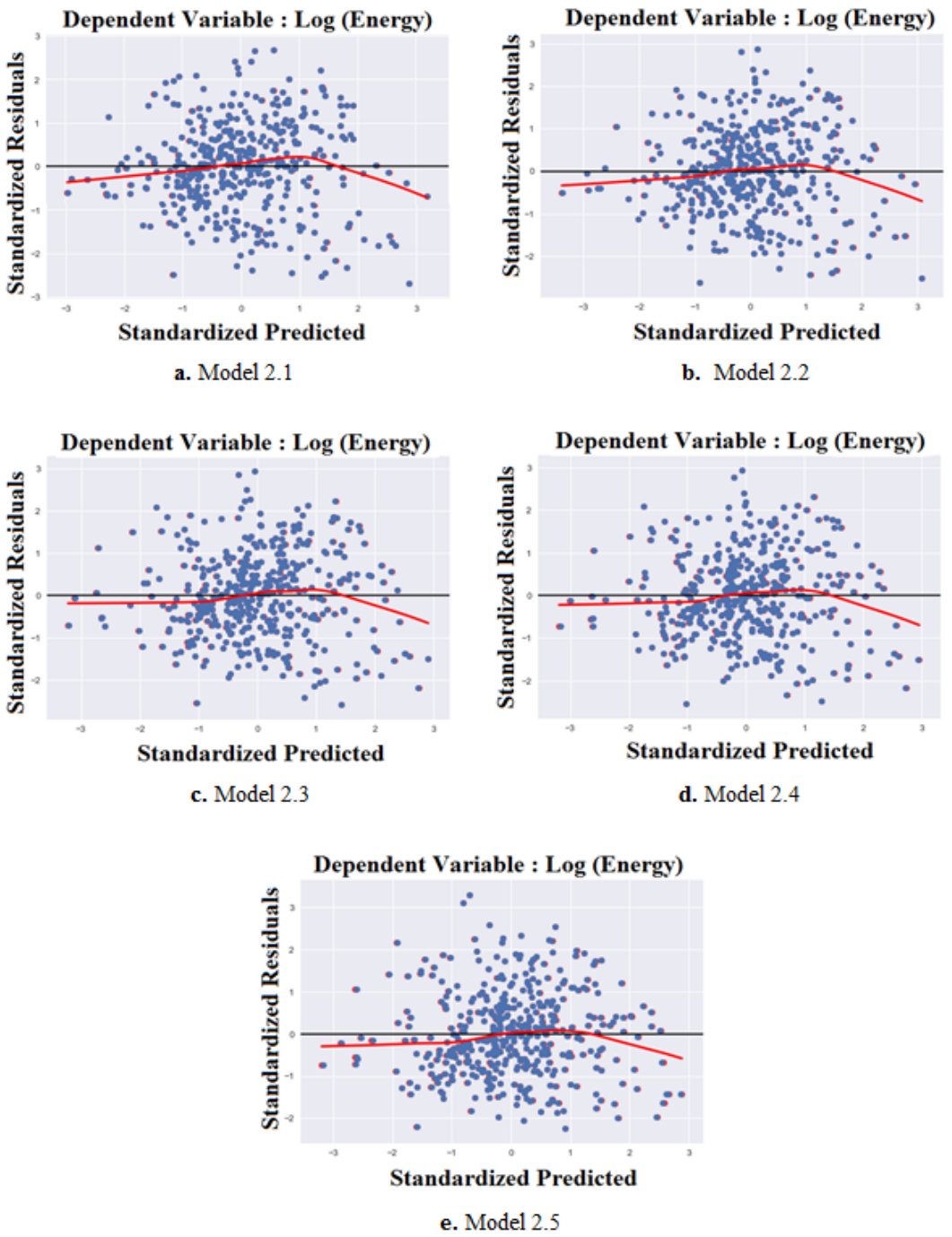


Figure 3.26 Scatterplots of the standalone energy consumption models with data rate

Data rate modeling

In this part, the data rate R_i^{sta} is modeled using *RSRP* and *SINR* variables. The RSRQ is not included in the model because p-value of its coefficient is greater than 0.05. According to the hypothesis that the coefficient of the independent variable is zero, if the p-value of the coefficient is greater than the cutoff (0.05) the hypothesis is not rejected. This means that the independent variable is not significant for the model. Therefore, we could not add RSRQ variable to our model. The regression steps are shown in Table 3.9. According to Table 3.10 and Figures 3.27.a, 3.27.b, and 3.27.c, it appears that data rate models also satisfy all the linear regression assumptions described in Section 3.2.1, except for the assumption of independence of residual terms in Model 3.1.

Table 3.9 Model summary of the standalone data rate

Model	**Formulas (R_i^{sta})**	R^2	**RMSE**
3.1	$39.124 + 4.184*\text{RSRP}^{Z}$	0.172	9.183
3.2	$39.124 + 3.093*\text{RSRP}^{Z} + 3.070*\text{SINR}^{Z}$	0.253	8.723
3.3	$39.540 + 2.969*\text{RSRP}^{Z} + 3.127*\text{SINR}^{Z} - 1.171*\text{RSRP}^{Z} * \text{SINR}^{Z}$	0.265	8.649

Table 3.10 Model assumptions for the standalone data rate

	Assumptions						
Model	**Linearity (Lowess Curve)**	**IoR (D-W Test)**	**CV (Levene's Test)**		**Normality (Jarque Bera)**		**Multi-collinearity (CI)**
			Stat	**Sig.**	**Stat**	**Sig.**	
3.1	Linear	0.98	0.98	0.32	2.44	0.30	1.0
3.2	Linear	1.09	1.82	0.18	4.13	0.13	1.45
3.3	Linear	1.13	1.46	0.23	4.05	0.13	1.53

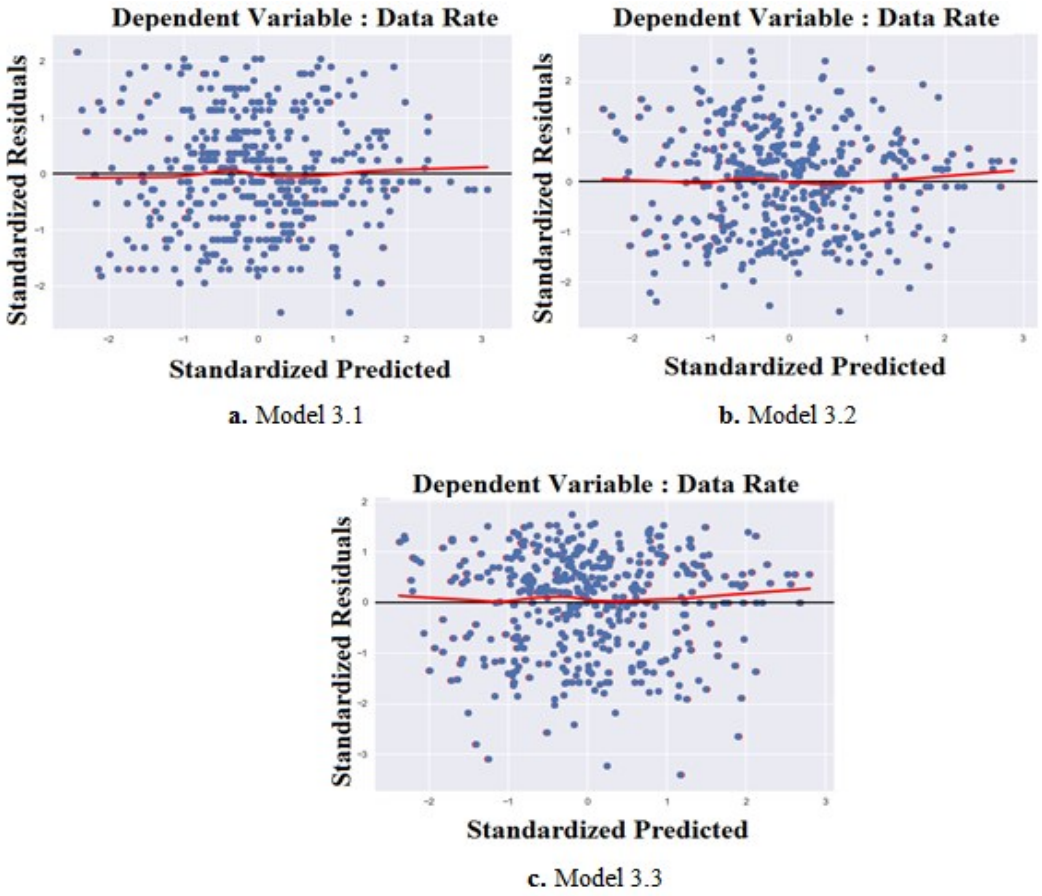


Figure 3.27 Scatterplots of the standalone data rate models

Robust regression models and other techniques

For the same dataset, we have also applied robust regression methods (Theil-Sen, RANSAC and Huber Regressors) and the obtained regression models are shown in Table 3.11. Apart from these methods, we apply other supervised learning methods such as Decision Tree regressor, Random Forest regressor, K-Neighbors regressor and Gradient Boosting regressor with default parameters. Before using these methods, firstly we shuffle data and then apply 10-fold-cross-validation [58] in order to avoid overfitting and to generalize the statistical results. According to the k-fold-cross-validation technique, the data is divided into k small sets. One of these sets is selected as a test set and the remaining k-1 sets are used as a training set. Then another set is used as a test set and the remaining part is used for training. This process continues until each of the k sets of data is used once as the test set. At the end, k pieces of results are obtained [59]. The average

of these results gives the overall success of the model ($R2$) and the average error rate (RMSE).

We also apply shuffling and cross validation for OLS and robust regression techniques and we illustrate average goodness of fit values for all aforementioned methods in Table 3.12. All results are discussed in Section 3.2.5.3.

Table 3.11 Model summary of robust regression techniques for the standalone case

Regression Technique	**Energy Model without Data Rate**	**Energy Model with Data Rate**	**Data Rate Model**
	$\log(E_i^{sta})$	**$\log(E_i^{sta})$**	**Data Rate (R_i^{sta})**
TheilSen Regressor	$3.184 - 0.081*\text{RSRP}^{Z} - 0.046*\text{RSRQ}^{Z} - 0.097*\text{SINR}^{Z}$	$3.173 - 0.174*\text{DataRate}^{Z} - 0.047*\text{RSRP}^{Z} - 0.021*\text{RSRQ}^{Z} - 0.052*\text{SINR}^{Z}$	$38.525 + 2.444*\text{RSRP}^{Z} + 1.538*\text{RSRQ}^{Z} + 2.283*\text{SINR}^{Z}$
RANSAC	$3.178 - 0.061*\text{RSRP}^{Z} - 0.007*\text{RSRQ}^{Z} - 0.147*\text{SINR}^{Z}$	$3.147 - 0.222*\text{DataRate}^{Z} - 0.039*\text{RSRP}^{Z} - 0.017*\text{RSRQ}^{Z} - 0.041*\text{SINR}^{Z}$	$36.979 + 2.357*\text{RSRP}^{Z} + 0.544*\text{RSRQ}^{Z} + 1.346*\text{SINR}^{Z}$
Huber Regressor	$3.180 - 0.087*\text{RSRP}^{Z} - 0.039*\text{RSRQ}^{Z} - 0.099*\text{SINR}^{Z}$	$3.173 - 0.176*\text{DataRate}^{Z} - 0.037*\text{RSRP}^{Z} - 0.021*\text{RSRQ}^{Z} - 0.052*\text{SINR}^{Z}$	$38.703 + 2.562*\text{RSRP}^{Z} + 1.352*\text{RSRQ}^{Z} + 2.424*\text{SINR}^{Z}$

Table 3.12 Model summary of various regression techniques for the standalone case

Models	Energy Model without Data Rate		Energy Model with Data Rate		Data Rate Model	
	Log(E_i^{sta})		Log(E_i^{sta})		Data Rate (R_i^{sta})	
Regressors	R^2	**RMSE**	R^2	**RMSE**	R^2	**RMSE**
Linear Regression (OLS)	0.430	0.196	0.722	0.134	0.236	8.708
TheilSen	0.428	0.196	0.720	0.135	0.228	8.735
RANSAC	0.379	0.205	0.704	0.141	0.161	9.144
Huber	0.428	0.197	0.720	0.135	0.234	8.716
Decision Tree	0.135	0.244	0.489	0.185	-0.049	10.104
Random Forest	0.422	0.200	0.702	0.142	0.246	8.659
KNeighbors	0.429	0.195	0.700	0.140	0.276	8.384
Gradient Boosting	0.445	0.193	0.722	0.135	0.273	8.391

3.2.5.2. Models for gateway case

We have done several standalone-gateway comparison tests (described in Section 3.1.3.3) for modeling of energy and data rate of a gateway node while serving a client. During these tests, we measured energy and data rate of a node while downloading data for herself (as a standalone node) and downloading data for a client node (as a gateway node). We also measured WiFi signal strength and link speed between gateway and client nodes. The descriptive statistics of the data is given in Table 3.13 and they are used for the standardization of the data. E_g^{oc} is energy consumption function of the gateway while receiving from the operator *o* and relaying the client *c*. E_g^{sta} is standalone energy consumption function of the gateway *g*. R_g^{oc} is received data rate of the gateway from the operator *o* while the relaying the client *c* The standalone data rate of the gateway is represented as R_g^{sta}. W_{ss} represents the WiFi signal strength while L_{gc} indicates the link speed between gateway *g* and the client *c*.

Table 3.13 Descriptive statistics of data for the gateway case

Descriptive Statistics					
	N	***Min.***	***Max.***	***Mean***	***Std. Dev.***
E_g^{oc} (mJ/Mbit)	180	38.05	133.31	64.30	16.85
Log(E_g^{oc})	180	3.64	4.89	4.13	0.25
E_g^{sta} (mJ/Mbit)	180	12.41	58.27	26.08	9.54
R_g^{oc} (Mbps)	180	5.21	37.03	23.16	6.78
R_g^{sta} (Mbps)	180	7.97	67.39	36.84	10.79
W_{ss} (dBm)	180	-66.13	-11.73	-37.91	18.96
L_{gc} (Mbps)	180	48.55	65	63.25	3.35

The correlations of the collected data are given in Table 3.14. According to the results, the gateway energy consumption E_g^{oc} is moderately correlated with her standalone energy consumption E_g^{sta}, her standalone data rate R_g^{sta} and the WiFi link speed L_{gc}. Gateway data rate R_g^{oc} is moderately correlated with her standalone data rate E_g^{sta} and it has a lower correlation with WiFi link speed L_{gc}. WiFi signal strength W_{ss} is also correlated with WiFi link speed, but energy and data rate have a low correction with WiFi signal strength. Gateway energy consumption E_g^{oc} and gateway data rate R_g^{oc} is highly

correlated with each other. However, we focus on the relationship between the gateway energy consumption E_g^{oc} and the standalone energy consumption E_g^{sta} and the relationship between gateway data rate R_g^{oc} and the standalone data rate R_g^{sta}.

Table 3.14 Correlations of data for the gateway case

Correlations								
		E_g^{oc}	Log (E_g^{oc})	E_g^{sta}	R_g^{oc}	R_g^{sta}	W_{ss}	L_{gc}
E_g^{oc}	Pearson Corr.	1	.988**	.522**	-.811**	-.448**	-.225**	-.455**
	Sig. (2-tailed)		.000	.000	.000	.000	.002	.000
Log (E_g^{oc})	Pearson Corr.	.988**	1	.520**	-.817**	-.451**	-.214**	-.426**
	Sig. (2-tailed)	.000		.000	.000	.000	.004	.000
R_g^{oc}	Pearson Corr.	-.811**	-.817**	-.517**	1	.530**	.201**	.396**
	Sig. (2-tailed)	.000	.000	.000		.000	.007	.000
** Correlation is significant at 0.01 level (2-tailed).								

In Figure 3.28, the first column shows the distribution of gateway energy, log of gateway energy and data rate. The other columns show the relationship of energy, log of energy and data rate to other variables. The gateway energy has lognormal distribution (so we will use the logarithm of the energy in order to satisfy the normality assumption) and the gateway data rate has the normal distribution. They are linearly correlated with the other variables except for WiFi signal strength (there is no an obvious relation with WiFi signal strength).

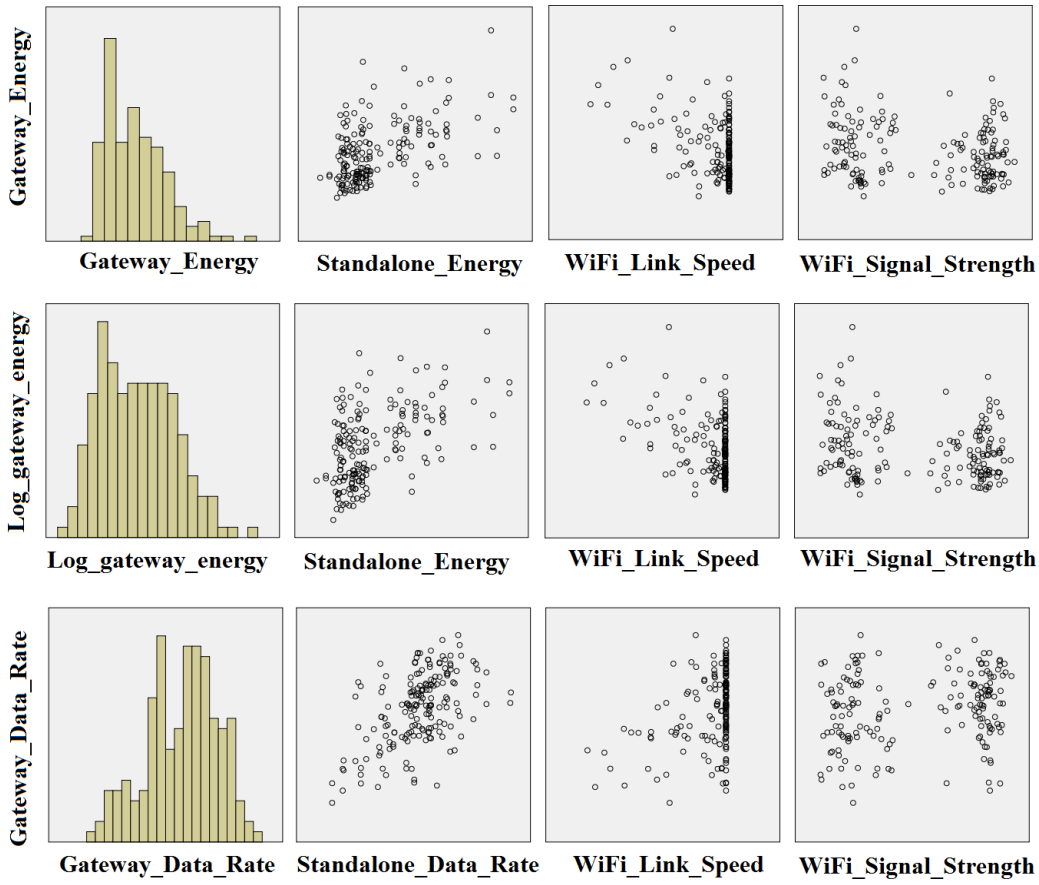


Figure 3.28 Data distribution and relationships for the gateway case

OLS regression models

We applied OLS regression for modeling energy and data rate of the gateway with respect to the standalone energy consumption and the standalone data rate of the gateway along with WiFi link speed. We do not include the WiFi signal strength in our models because it is not a significant variable for this model (its significance value is greater than 0.05). Before modeling, we clean our data and standardized the variables. We present two models that are gateway energy consumption model and gateway data rate model given in Table 3.15. Gateway energy consumption while serving a client is modeled according to standalone energy consumption E_{sta} of the gateway and link speed L_{gc}between gateway and client. Also, gateway download data rate while serving a client is modeled according to standalone data rate R_{sta} of the gateway and the link speed L_{gc}.

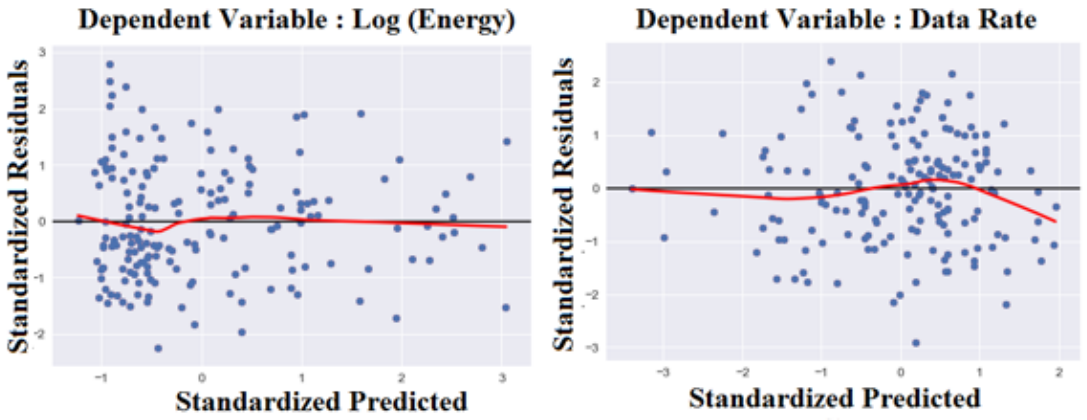
Table 3.15 Model summary of the gateway energy consumption and data rate

Model	Formulas	R^2	RMSE
Log (E_g^{oc})	$4.132 + 0.123 * E_{sta}{}^{Z} - 0.099 * L_{gc}^{Z}$	0.428	0.188
Data Rate (R_g^{oc})	$23.159 + 3.494 * R_{sta}^{Z} + 2.549 * L_{gc}^{Z}$	0.423	5.134

Table 3.16 gives information about the linear regression assumptions of the models (the results of the assumption of linearity are shown in the Figures 3.29.a and 3.29.b). The energy model satisfies all assumptions. The data rate model also satisfies all assumptions.

Table 3.16 Model assumptions for the gateway energy consumption and data rate

	Assumptions						
Model	**Linearity (Lowess Curve)**	**IoR (D-W Test)**	**CV (Levene's Test)**		**Normality (Jarque Bera)**		**Multi-collinearity (CI)**
			Stat	**Sig.**	**Stat**	**Sig.**	
Log (E_g^{oc})	Approx. Linear	1.222	0.001	0.970	5.326	0.007	1.06
Data Rate (R_g^{oc})	Approx. Linear	1.843	2.646	0.106	0.905	0.636	1.04



a. Energy Model (E_g^{oc}) **b.** Data Rate Model (R_g^{oc})

Figure 3.29 Scatterplots of the gateway energy consumption and data rate models

Robust regression models and other techniques

We also applied robust regression techniques for the same dataset. The obtained models are given in Table 3.17. In order to avoid over-fitting and to evaluate all mentioned models, we shuffle the dataset and apply 5-fold cross-validation technique for the regressors above mentioned and other machine learning techniques (with default parameters). The goodness of fit values of the models are given in Table 3.18. All results are discussed in Section 3.2.5.3.

Table 3.17 Model summary of robust regression techniques for the gateway case

Regression Technique	**Gateway Energy Model** (Log (E_g^{oc}))	**Gateway Data Rate Model** (R_g^{oc})
TheilSen Regressor	$4.056 + 0.128 * E_{sta}^{Z} - 0.018 * L_{gc}^{Z}$	$22.411 + 3.572 * R_{sta}^{Z} + 5.374 * L_{gc}^{Z}$
RANSAC	$4.064 + 0.237 * E_{sta}^{Z} - 0.070 * L_{gc}^{Z}$	$23.092 + 5.569 * R_{sta}^{Z} + 4.752 * L_{gc}^{Z}$
Huber Regressor	$4.121 + 0.135 * E_{sta}^{Z} - 0.099 * L_{gc}^{Z}$	$23.132 + 3.804 * R_{sta}^{Z} + 2.440 * L_{gc}^{Z}$

Table 3.18 Model summary of various regression techniques for the gateway case

Models	Gateway Energy Model (Log (E_g^{oc}))		Gateway Data Rate Model (R_g^{oc})	
Regressors	R^2	**RMSE**	R^2	**RMSE**
Linear Regression (OLS)	0.396	0.189	0.395	5.154
TheilSen	0.013	0.246	0.153	6.142
RANSAC	0.334	0.207	0.330	5.426
Huber	0.389	0.190	0.392	5.164
Decision Tree	0.072	0.243	0.021	6.369
Random Forest	0.390	0.194	0.305	5.826
KNeighbors	0.280	0.208	0.275	5.651
Gradient Boosting	0.413	0.186	0.406	5.102

3.2.5.3. Results and discussion

Standalone Models. For all models, the increment in SINR, RSRP, and RSRQ values raises the data rate while reducing energy consumption, as expected. Among these LTE parameters, RSRP and SINR seem to be more influential. Looking at the goodness of fit values, we can say that energy consumption has better model fit compared to data rate. For OLS regression models described in Section 3.2.5.1, energy consumption model without data rate variable explains 46% of the variability of the response data, while data rate model accounts for 27% of the variation. R^2 is not very close to 1, especially for data rate modeling. This is not surprising because there are other effects (such as weather conditions, download activities of nearby devices) that we cannot include in our modeling and also, some smartphones could not provide enough fine-grained measurement for signal metrics. RMSE values seem to be high for data rate models, and there is a large difference compared to energy models. Nevertheless, it should be noted that natural logarithm is applied for energy consumption values. While interpreting the RMSE values, it should be considered that mean data rate value is approximately 40 Mbps as given in Table 3.3, while mean log (Energy) value is approximately 3.2.

Addition of interaction terms slightly improves the models. For energy modeling, including measured data rate as a variable improves the goodness of fit (for OLS, R^2 improves from 0.46 to 0.75). Although in most cases data rate is not an input, including data rate in the model can be useful for some applications (such as [14]) where the estimated energy is a function of obtained data rate.

Next, we compare the goodness of fit values for cross-validated regression results illustrated in Table 3.12. It is observed that Gradient Boosting regressor yields best results for energy consumption and it is also good for data rate modeling. K-Neighbors regressor provides slightly better model fit for data rate. Nevertheless, superiority of these regression techniques over OLS is not so apparent. OLS gives similar results compared to Theilsen and Huber estimators. Besides, the goodness of fit values of OLS with cross-validation are close to OLS models without cross-validation. This results suggest that it is suitable to utilize OLS models provided in Section 3.2.5.1 for energy consumption and data rate estimations if there is some tolerance for estimation errors.

Gateway Models. For the OLS models, the link speed is important parameter along with standalone energy and standalone data rate. However, in the robust regression models (except Huber regression), the link speed is not as important as in OLS models. The gateway energy model explains 43% of the variability of the response data while the gateway data rate model explains 42% of the variation on the data. RMSE value of data rate model is much higher than RMSE value of the energy model. Because the mean value of the log of energy consumption E_g^{oc} is quite low than the mean value of the data rate R_g^{oc}, so the RMSE values of the energy seems lower. According to cross-validated regression results, Gradient Boosting gives best results for both gateway's energy and data rate models. However, linear regression and Huber models provide a very close result to the Gradient Boosting in term of R^2 and RMSE values. Also, the goodness of fit values of cross-validated OLS models are close to OLS models without cross-validation. Hence, by using these models, if we know standalone energy and data rate of the node we can estimate the energy consumption and data rate (with some error tolerance) of a node serving a client.

3.3. Application for bargaining module

In this part, we will examine the estimated and actual utility, cost, and profit of the nodes (gateway and client) involved in a UPN service. In our bargaining scenario, client estimates her standalone data rate by using Model 3.3 in Section 3.2.5.1 and then she estimates her utility function by using the estimated data rate. Also, the gateway node estimates her energy consumption and data rate while she serves a client node by using the models given in Section 3.2.5.2. She uses these estimates for her cost prediction. The users make an agreement after the bargaining according to the utility and cost estimations, the data transfer starts between them. The utility and cost estimations (also the estimates of energy and data rate) and bargaining procedure is described as follows.

For energy and data rate prediction, the instant LTE signal parameters of gateway and client are given in Table 3.19. In addition, the link speed L_{gc} between gateway and client is 65 Mbps. The client estimates her standalone data rate R_c^{sta} as 14.9 Mbps by using the LTE signal metrics. The gateway estimates her data rate R_g^{oc} as 21.9 Mbps by using L_{gc} and her standalone data rate. Also, the gateway estimates her energy

consumption E_g^{oc} by using the standalone energy consumption and link speed. In Figure 3.30, the estimated and actual energy functions are shown. The values of the estimated energy consumption are closer to the actual values when the download data rate of the gateway R_g^{oc} is greater than 5 Mbps. Since our models are done according to the main part of the data, our estimations get better after the head part.

Table 3.19 LTE signal parameters of the nodes

	RSRP (dBm)	RSRQ (dB)	SINR (dB)
Gateway	-100	-6	16
Client	-116	-9	2

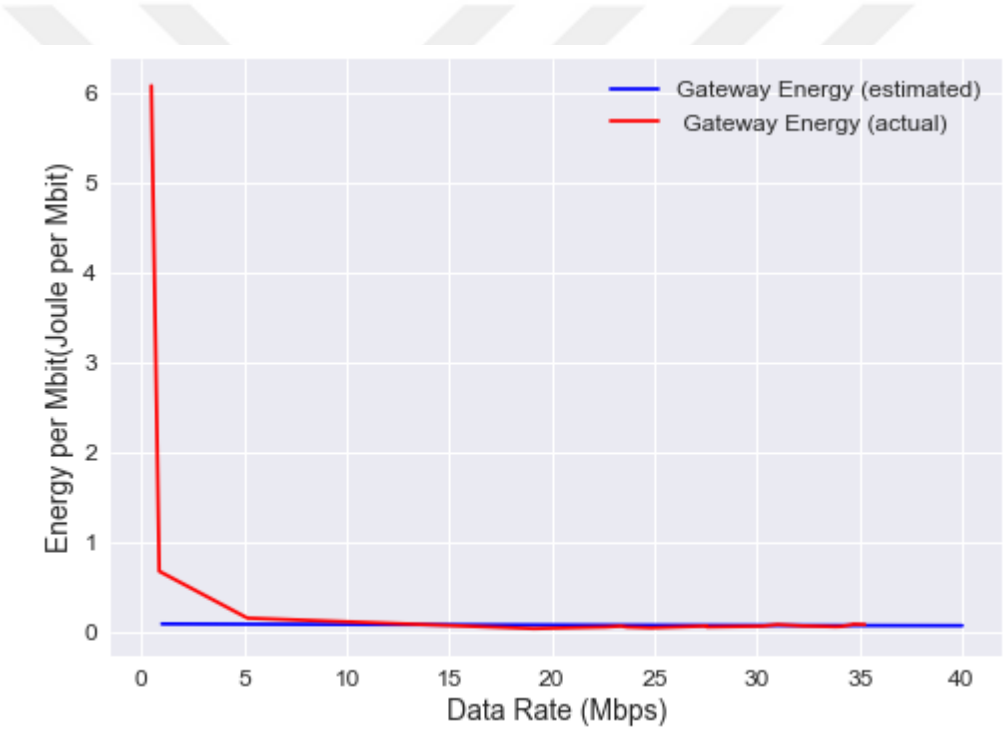


Figure 3.30 Estimated and actual energy consumption values of the gateway

Figure 3.31 shows the estimated utility and cost values and the estimated download data rate R_g^{oc} of the gateway. Also, the agreed virtual currency function is illustrated. While estimating the utility and cost functions, the parameter Q_c is set to 0.001 and α_i is set to 0.9. Also, the parameter $e_{residual}$ of the gateway is 0.9. The client and gateway agree on the virtual currency function within the cooperation region where the utility is greater than the cost, and R_g^{oc} is greater than the agreed minimum data rate R_0 (19 Mbps).

The utility and cost functions can change over time, therefore the agreed virtual currency function can be updated according to new utility and cost functions.

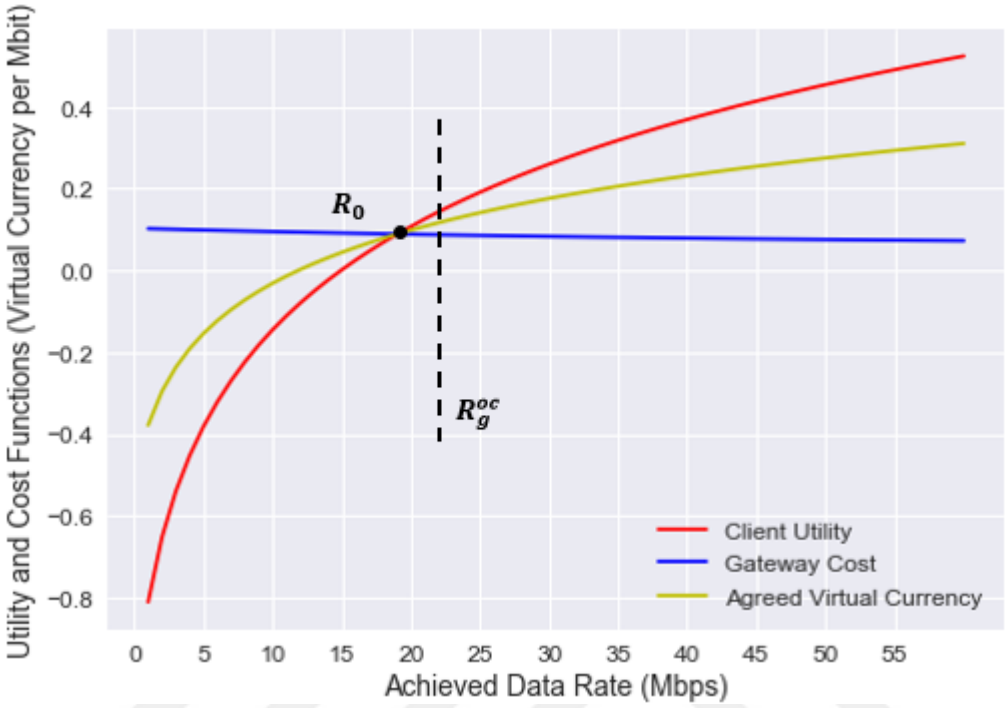
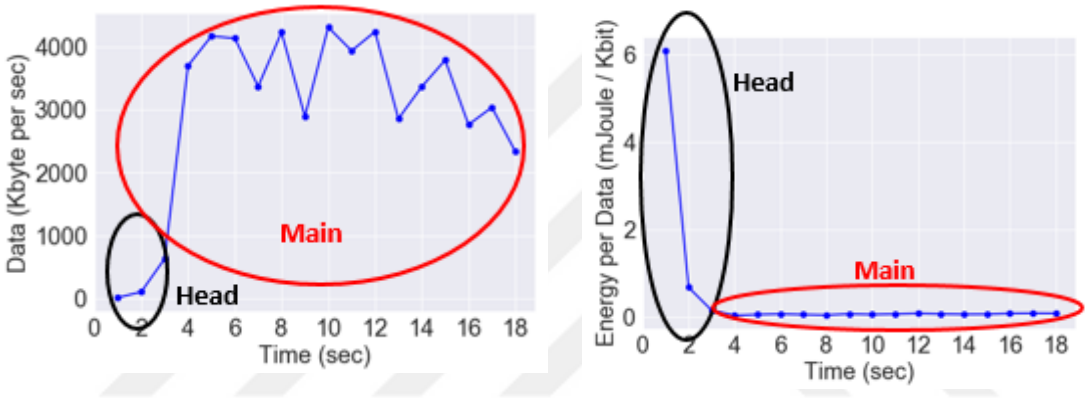


Figure 3.31 Estimated utility and cost functions with the received data rate of the gateway node while relaying the client

After the agreement, the client starts to download 50 MB data file via the gateway. When the data is flowing, the state-based separation technique is applied to divide data download process into the states for both the client and gateway nodes. The values of the parameters defined for the separation technique are given in Table 3.20. These parameters can be adjusted according to different data types such as web surfing and video streaming data. The states of the data download process are shown in Figure 3.32. It is observed that the head part takes 3 seconds and gateway consumes more energy at the head part. After the head part, the sending data rate of the gateway (to the client) increases due to the increment of the data rate received from the operator so the client data rate also increases. Thus, energy consumption per data (Kbit) of gateway decreases sharply. In this data download process, there is no any considerable data flow at the tail state so it is not illustrated in Figure 3.32.

Table 3.20 The values of the parameters used for separation of data transfer process into the states

Ds (KB)	5
D_h (KB)	25
D_t (KB)	15
T_{h1} (sec)	1
T_{h2} (sec)	15
T_m (sec)	2
T_t (sec)	2
d_h (KB)	1
d_m (KB)	15



a. The client transferred data (per sec) **b.** The Gateway energy consumption

Figure 3.32 Illustration of 50 MB data download process according to the states

Actual utility and cost values are calculated for each time period (sec) according to client's actual data rate and gateway's actual energy consumption. Figure 3.33illustrates the estimated cost and utility functions and the actual cost and utility values. The (blurred) part where the cost is greater than the utility represents the head part of the data download process. In this part, the estimated cost values are not close to actual cost values because our energy (and data rate) models were done according to the main part of the data, therefore the estimations in the head (and tail) state may not overlap the actual values. In the part where the utility is greater than the cost, the actual utility and cost values are almost the same as the values of estimated utility and cost functions (there exist some negligible deviations for the cost function), because this part is within the main state. As we expected, the achieved data rate is greater than or equal to the minimum agreed data rate during the data flow in the main state.

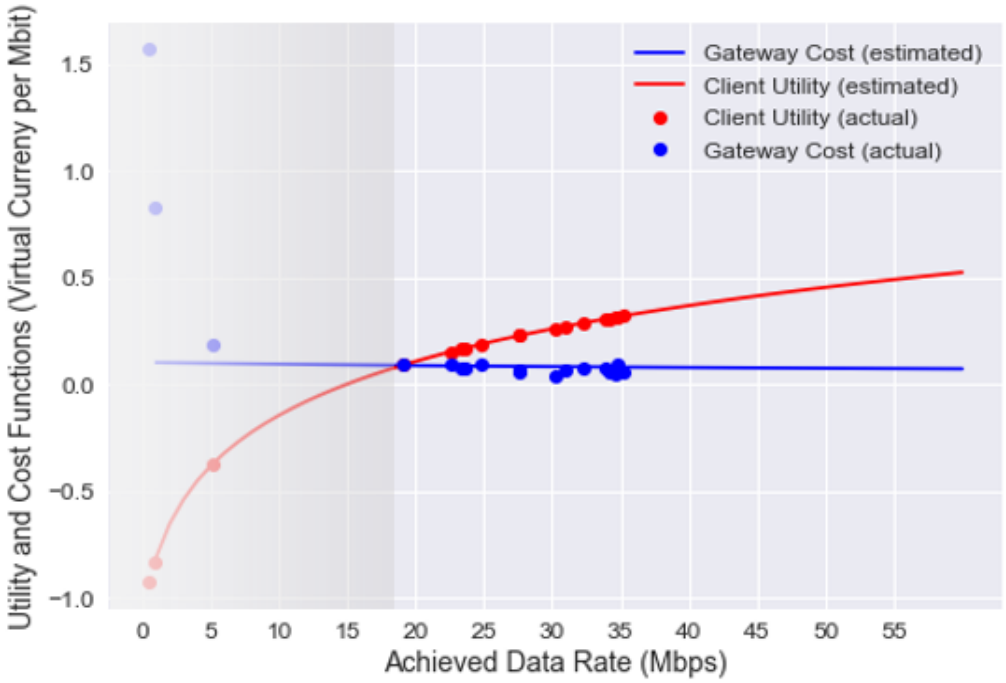


Figure 3.33 The comparison of the estimated and actual utility and cost values

The amount of virtual currency transferred is calculated according to Eq. (2.8) during the head state. In the tail part, there is no considerable data flow, so the amount of the virtual currency transferred is only calculated according to connection cost. The connection cost C_{con} is equal to 0.1 and the parameter ρ is set to 0.001. The amount of virtual currency transferred in the main state is computed according to Eq. (2.6) because the first offer is made by the client. The parameters δ_1, δ_2 are set to 0.9. Figure 3.34 shows the actual and estimated value of the virtual currency transferred in the main state. The actual amount of the transferred virtual currency is within the agreement region and the actual virtual currency values are close to the values of estimated virtual currency function. During the data transfer, the same agreed virtual function was used because there was no significant change in the functions agreed upon (the small deviations from the agreed virtual currency function can be ignored).

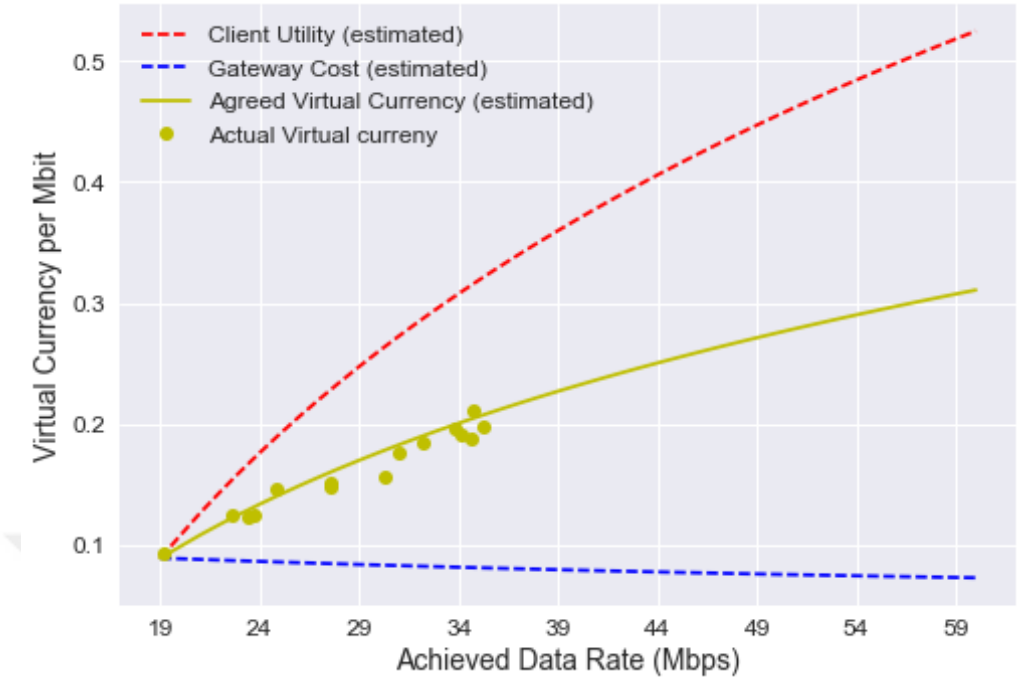


Figure 3.34 The comparison of the estimated and actual virtual currency values

The estimated net profits of the gateway and client are illustrated in Figure 3.35. It is observed that the net profit of both the gateway and the client increased as the data rate increased. Similarly, actual net profits of the gateway and client generally increased as the data rate increased but the profit decreased at some points as seen in Figure 3.36. This is because the energy consumption of the gateway can increase due to the software or hardware reasons or signaling events. This reduces the shareable pie between the gateway and client nodes, so the net profit decreases. If the nodes' profits change greatly and any of the parties can no longer profit, the agreement is renewed (according to new cost and utility functions) or terminated.

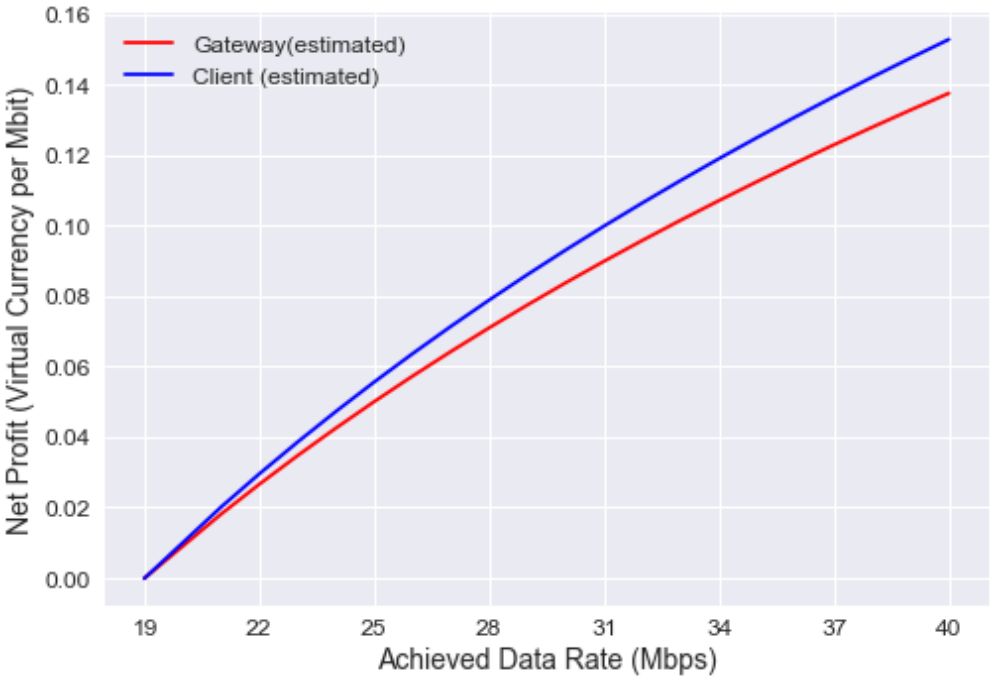


Figure 3.35 The estimated net profit of the nodes

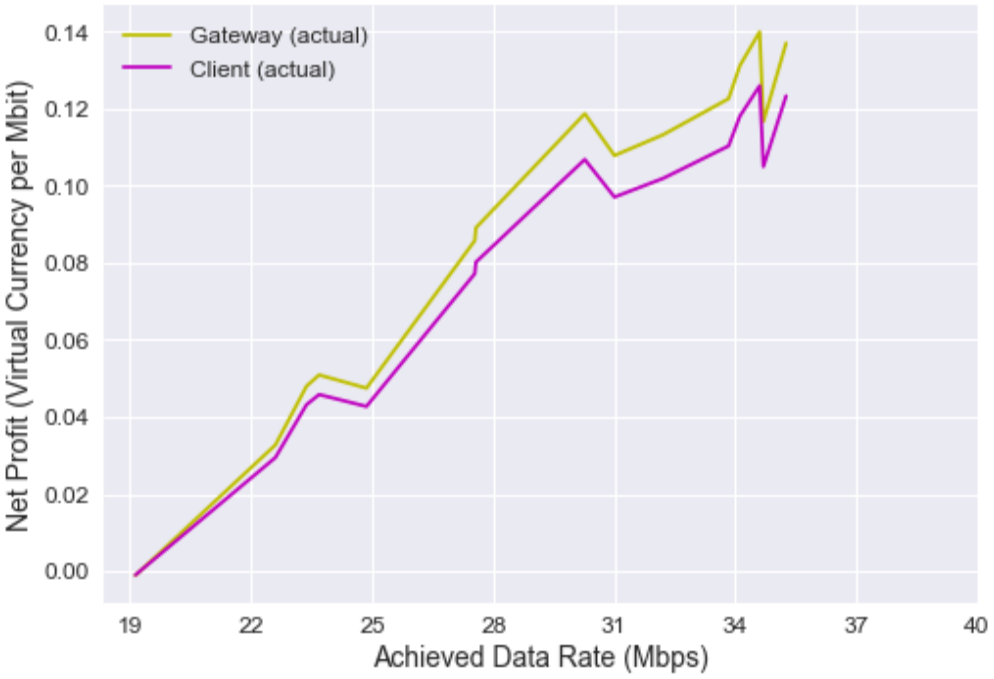


Figure 3.36 The actual net profit of the nodes

Our study shows that the utility and cost functions estimated by using the energy and data rate models overlap with the actual results (for the main state). Also, this study reveals that it is necessary to separate the data download process into the states while evaluating the agreement conditions. Because if we did not have separated the data into the states and applied different virtual currency policies accordingly, the agreement between the client and gateway nodes would be violated because the gateway could not provide the minimum agreed data rate during certain time slot due to the TCP slow start. This could negatively affect the use of UPN services. By applying a state-based evaluation policy, we contribute to encouraging users to access the Internet through UPN services.

4. CONCLUSION AND FUTURE WORK

In this thesis, we proposed a scheme for modeling energy and data rate of a node which is in the standalone case or a UPN system. In order to estimate energy consumption and throughput of the node, we performed experimental measurements in LTE networks. According to the experiment results, if a node acts as a gateway node she consumes more energy compared to her standalone case and her download data rate also is lower compared to her standalone case. This is because relaying data to a client requires some processing power and it takes time, which reduced the download data rate and increases the energy. Besides, time required for relaying data changes according to WiFi link speed between nodes. Client node consumes less energy if she connects the Internet via a gateway because WiFi networks are more power efficient than the LTE networks. Also, client can achieve higher data rates by accessing the Internet via gateway when its connection is weak enough compared to the gateway's cellular connection.

Based on the experiments, we present some energy consumption and data rate models for standalone and gateway nodes by applying OLS, robust regression and other techniques. Regarding the standalone models, we obtained better model fit for energy consumption compared to data rate modeling. Using data rate as a variable further increased the regression performance for energy modeling. Estimation error is not very low, especially for data rate modeling. This is not surprising since data rate would be affected by other factors such as weather conditions and user density. Similar results are also obtained by other regression methods such as Theil-Sen and Huber robust regression techniques, Gradient Boosting and K-Neighbors, although Gradient Boosting performs slightly better than the others. Concerning the gateway models, we can say that energy and data rate models yield close fit results. On the other hand, the estimation error of the data rate model is not as lower as the energy consumption model because the mean of the energy values is lower. Huber and Gradient Boosting regressors give similar results for energy and data rate.

We applied our presented models to estimate utility and cost functions of users on a UPN service. The utility and cost estimations are close to the actual values (especially for the main state). This shows that our presented models (energy and data rate) are

applicable for estimating the utility and cost functions. Also, we show the necessity of the separation of the data into the states.

In future work, we will enhance level-based energy consumption model that estimates the levels of energy such as high, moderate, and low instead of estimating the consumed energy value in order to generalize our energy models. Also, we will improve our data sharing application (which is a kind of UPN service) by adding the presented energy and data rate models. Besides, we plan to examine multi-node UPNs (gateway, relay, and client).

REFERENCES

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper Cisco, https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html (28.05.2018).

[2] Gao, L., Iosifidis, G., Huang, J., Tassiulas, L., & Li, D. (2014). Bargaining-Based Mobile Data Offloading. *IEEE Journal on Selected Areas in Communications*, *32*(6), 1114–1125.

[3] Sofia, R., & Mendes, P. (2008). User-provided networks: consumer as provider. *IEEE Communications Magazine*, *46*(12), 86–91.

[4] Constantinescu, M., Onur, E., Durmus, Y., Nikou, S., Reuver, M. de, Bouwman, H., Djurica, H., Glatz, P. M. (2014). Mobile tethering: overview, perspectives and challengess. *Info*, *16*(3), 40–53.

[5] Fon: The Global WiFi Network, Fontech: Leading WiFi Software Fon. https://fon.com (28.05. 2018).

[6] Open Garden - Home, https://www.opengarden.com (28.05.2018).

[7] Karma 2018 | Karma, https://yourkarma.com/wifi/2017-2 (28.05.2018).

[8] Next-Gen Mobile Networking M87, http://www.m-87.com (28.05.2018).

[9] Iosifidis, G., Gao, L., Huang, J., & Tassiulas, L. (2017). Efficient and Fair Collaborative Mobile Internet Access. *IEEE/ACM Transactions on Networking*, *25*(3), 1386–1400.

[10] G., Gao, L., Huang, J., & Tassiulas, L. (2014). Incentive mechanisms for user-provided networks. *IEEE Communications Magazine*, *52*(9), 20–27.

[11] Nash, J. F. (1950). The Bargaining Problem. *Econometrica*, *18*(2), 155–162.

[12] Gao, L., Iosifidis, G., Huang, J., & Tassiulas, L. (2014). Hybrid data pricing for network-assisted user-provided connectivity. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications* (pp. 682–690).

[13] Khalili, M. M., Gao, L., Huang, J., & Khalaj, B. H. (2015). Incentive design and market evolution of mobile user-provided networks. In *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 498–503).

[14] Taşçi, S. E., & Korçak, Ö. (2016). Incentivizing user provided connectivity for enhanced quality of service. In *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (pp. 1–7).

[15] Ananthanarayanan, G., Padmanabhan, V. N., Ravindranath, L., & Thekkath, C. A. (2007). COMBINE: Leveraging the Power of Wireless Peers Through Collaborative Downloading. In *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services* (pp. 286–298). New York, NY, USA: ACM.

[16] Sharma, A., Navda, V., Ramjee, R., Padmanabhan, V. N., & Belding, E. M. (2009). Cool-Tether: energy efficient on-the-fly wifi hot-spots using mobile phones (p. 109). ACM.

[17] Yu, T., Zhou, Z., Zhang, D., Wang, X., Liu, Y., & Lu, S. (2014). INDAPSON: An incentive data plan sharing system based on self-organizing network. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications* (pp. 1545–1553).

[18] Huang, J., Qian, F., Gerber, A., Mao, Z. M., Sen, S., & Spatscheck, O. (2012). A Close Examination of Performance and Power Characteristics of 4G LTE Networks. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services* (pp. 225–238). New York, NY, USA: ACM.

[19] Pathak, A., Hu, Y. C., Zhang, M., Bahl, P., & Wang, Y.-M. (2011). Fine-grained Power Modeling for Smartphones Using System Call Tracing. In *Proceedings of the Sixth Conference on Computer Systems* (pp. 153–168). New York, NY, USA: ACM.

[20] Schulman, A., Navda, V., Ramjee, R., Spring, N., Deshpande, P., Grunewald, C., Jain, K., Padmanabhan, V. N. (2010). Bartendr: A Practical Approach to Energy-aware Cellular Data Scheduling. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking* (pp. 85–96). New York, NY, USA: ACM.

[21] Afroz, F., Subramanian, R., Heidary, R., Sandrasegaran, K., & Ahmed, S. (2015). SINR, RSRP, RSSI and RSRQ Measurements in Long Term Evolution Networks. *International Journal of Wireless & Mobile Networks*, *7*(4), 113–123.

[22] Cainey, J., Gill, B., Johnston, S., Robinson, J., & Westwood, S. (2014). Modelling download throughput of LTE networks. In *39th Annual IEEE Conference on Local Computer Networks Workshops* (pp. 623–628).

[23] Özlü, A., Uludağ, K., & Korçak, Ö. (2017). User provided network system with incentive mechanism. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, Antalya.

[24] Uludağ, K., & Korçak, Ö. (2017). Energy and rate modeling of data download over LTE with respect to received signal characteristics. In *2017 27th International Telecommunication Networks and Applications Conference (ITNAC)*, Melbourne.

[25] Rubinstein, A. (1982). Perfect Equilibrium in A Bargaining Model. *Econometrica*, *50*, 97–109.

[26] Thinkbroadband : The UK's Largest Independent Broadband / ADSL Troubleshooting Website (Incorporating News, Reviews and Comparisons), https://www.thinkbroadband.com/download (28.05.2018)

[27] European Telecommunications Standards Institute. (2013). *TS 136 214 - V11.1.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements (3GPP TS 36.214 version 11.1.0 Release 11)*, http://archive.org/details/etsi_ts_136_214_v11.01.00 (28.05.2018)

[28] Subramanian, R., Sandrasegaran, K., & Kong, X. (2016). Benchmarking of real-time LTE network in dynamic environment. In *2016 22nd Asia-Pacific Conference on Communications (APCC)* (pp. 20–25).

[29] European Telecommunications Standards Institute. (2010). *TS 136 133 - V8.9.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management (3GPP TS 36.133 version 8.9.0 Release 8)*, http://archive.org/details/etsi_ts_136_133_v08.09.00 (28.05.2018)

[30] Holma, H., & Toskala, A. (2011). *LTE for UMTS: Evolution to LTE-Advanced*. John Wiley & Sons.

[31] European Telecommunications Standards Institute. (2000). *TS 125 215 - V3.1.1 - Universal Mobile Telecommunications System (UMTS); Physical layer - Measurements (FDD) (3G TS 25.215 version 3.1.1 Release 1999)*, http://archive.org/details/etsi_ts_125_215_v03.01.01 (28.05.2018)

[32] Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd Edition* (Third edition). Mahwah, N.J: Routledge.

[33] Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, *69*(3), 161–182.

[34] Durbin, J., & Watson, G. S. (1950). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*, *37*(3/4), 409–428.

[35] Mayers, A. (2013). *Introduction to Statistics and SPSS in Psychology*. Harlow: Pearson Education.

[36] Iachine, I., Petersen, H. C., & Kyvik, K. O. (2010). Robust tests for the equality of variances for clustered data. Journal of Statistical Computation and Simulation, 80(4), 365–377.

[37] Stevens, J. P. (2009). *Applied Multivariate Statistics for the Social Sciences, Fifth Edition* (5 edition). New York: Routledge.

[38] Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review / Revue Internationale de Statistique*, *55*(2), 163–172.

[39] Hutcheson, G., & Hutcheson, G. (2011). Ordinary Least-Squares Regression. In L. Moutinho & G. Hutcheson, *The SAGE Dictionary of Quantitative Management Research* (pp. 225–228). United Kingdom: SAGE Publications Ltd.

[40] Rousseeuw, P. J., & Leroy, A. M. (2003). *Robust Regression and Outlier Detection* (1 edition). Hoboken, NJ: Wiley-Interscience.

[41] Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389.

[42] Fischler, M. A., & Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, *24*(6), 381–395.

[43] Huber, P. J., & Ronchetti, E. M. (2009). *Robust Statistics* (2 edition). Hoboken, N.J: Wiley.

[44] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (2nd ed.). New York: Springer-Verlag.

[45] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

[46] Altman, N. S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. American Statistician - AMER STATIST, 46, 175–185.

[47] Friedman, J. H. (2000). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29, 1189–1232.

[48] Miles, J. (2014). R Squared, Adjusted R Squared. In *Wiley StatsRef: Statistics Reference Online*. American Cancer Society.

[49] Scikit-Learn, Model Evaluation: Quantifying the Quality of Predictions, http://scikit-learn.org/stable/modules/model_evaluation.html#r2-score (28.05.2018).

[50] Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? *Geoscientific Model Development Discussions*, *7*, 1525–1534.

[51] Clark-Carter, D. (2014). z Scores. In *Wiley StatsRef: Statistics Reference Online*. American Cancer Society.

[52] Statsmodels, Regression, Linear Model, OLS, http://www.statsmodels.org/dev/generated/statsmodels.regression.linear_model.OLS.html (29.05.2018).

[53] Scikit-Learn, Generalized Linear Models, http://scikit-learn.org/stable/modules/linear_model.html#robustness-regression-outliers-and-modeling-errors (29.05.2018).

[54] Scikit-Learn, Decision Trees, http://scikit-learn.org/stable/modules/tree.html (29.05.2018).

[55] Scikit-Learn, Nearest Neighbors, http://scikit-learn.org/stable/modules/neighbors.html (29.05.2018).

[56] Scikit-Learn, Ensemble methods, http://scikit-learn.org/stable/modules/ensemble.html (29.05.2018).

[57] Montgomery, D. C., & Runger, G. C. (2010). *Applied Statistics and Probability for Engineers*. John Wiley & Sons.

[58] Scikit-Learn, Cross-validation: evaluating estimator performance, http://scikit-learn.org/stable/modules/cross_validation.html (28.05.2018).

[59] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In e International Joint Conference on Artificial Intelligence (Vol. 14, pp. 1137–1145). Montreal, Canada.

RESUME

KÜBRA ULUDAĞ

Marmara University,

Computer Engineering Department, Faculty of Engineering,

Goztepe Campus, Kadikoy, Istanbul, Turkey

Phone (Cell)+90-5050098534

e-mail: k.uludag427@gmail.com

EDUCATION

M.S., Computer Science Marmara University, Faculty of Engineering, Istanbul, Turkey, Ongoing.
Thesis Topic: Energy and Data Rate Modeling in User Provided Networks with Incentive Mechanisms,
Advisor: Assist. Prof. Ömer KORÇAK.

Double Major, Electrical and Electronics Engineering Marmara University, Istanbul, Turkey, 2015
B.S., Computer Science Marmara University, Istanbul, Turkey, 2014

WORK EXPERIENCE

2015-… , Research Assistant, Wireless Networks and Systems Research Group, Marmara University, Computer Engineering Dept, Istanbul, Turkey

PUBLICATIONS

Ahmet Özlü, Kübra Uludağ and Ömer Korçak. User provided network system with incentive mechanism. In 2017 25th Signal Processing and Communications Applications Conference (SIU) (pp. 1–4), IEEE, 2017.

Kübra Uludağ and Ömer Korçak. Energy and rate modeling of data download over LTE with respect to received signal characteristics. In 2017 27th International Telecommunication Networks and Applications Conference (ITNAC) (pp. 1–6), IEEE, 2017.

RESEARCH INTERESTS

User Provided Networks, Wireless Network, Machine Learning

FOREIGN LANGUAGES

English