UNIVERSITY
OF TRENTO - Italy

DEPARTMENT OF INFORMATION ENGINEERING AND COMPUTER SCIENCE
ICT International Doctoral School

Energy Efficiency and Privacy in Device-to-Device Communication

by

Muhammad Usman

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

Advisor
Prof. Fabrizio Granelli
University of Trento, Italy

Co-Advisor
Dr. Muhammad Rizwan Asghar
The University of Auckland, New Zealand

December 2017

To my beloved family

Abstract

Mobile data traffic has increased many folds in recent years and current cellular networks are undeniably overloaded to meet the escalating user's demands for higher bandwidth and data rates. To meet such demands, Device-to-Device (D2D) communication is regarded as a potential solution to solve the capacity bottleneck problem in legacy cellular networks. Apart from offloading cellular traffic, D2D communication, due to its intrinsic property to rely on proximity, enables a broad range of proximity-based applications for both public safety and commercial users. Some potential applications, among others, include, proximity-based social interactions, exchange of information, advertisements and Vehicle-to-Vehicle (V2V) communication. The success of D2D communication depends upon the scenarios in which the users in the proximity interact with each other. Although there is a lot of work on resource allocation and interference management in D2D networks, very few works focus on the architectural aspects of D2D communication, emphasizing the benchmarking of energy efficiency for different application scenarios.

In this dissertation, we benchmark the energy consumption of D2D User Equipments (UEs) in different application scenarios. To this end, first we consider a scenario wherein different UEs, interested in sharing the same service, form a Mobile Cloud (MC). Since, some UEs can involve in multiple services/applications at a time, there is a possibility of interacting with multiple MCs. In this regard, we find that there is a threshold for the number of UEs in each MC, who can participate in multiple applications, beyond which legacy cellular communication starts performing better in terms of overall energy consumption of all UEs in the system. Thereafter, we extend the concept of MC to build a multi-hop D2D network and evaluate the energy consumption of UEs for a content distribution application across the network. In this work, we optimize the size of an MC to get the maximum energy savings.

Apart from many advantages, D2D communication poses potential challenges in terms of security and privacy. As a solution, we propose to bootstrap trust in D2D UEs before establishing any connection with unknown users. In particular, we propose Pretty Good Privacy (PGP) and reputation based mechanisms in D2D networks. Finally, to preserve user's privacy and to secure the contents, we propose to encrypt the contents cached at D2D nodes (or any other caching server). In particular, we leverage convergent encryption that can provide an extra benefit of eliminating duplicate contents from the caching server.

Keywords: Device-to-Device Communication, Mobile Cloud, Energy Efficiency, Privacy, Security, Trust, Software-Defined Networking, 5G Networks, Public Safety, Smart Cities, Cooperative Beamforming, Cognitive Radio

Acknowledgements

First and foremost, I would like to extend my sincere gratitude to my Ph.D. advisor Associate Prof. Dr. Fabrizio Granelli for his excellent guidance and unlimited support throughout my Ph.D. program. Next, I would like to sincerely thank my co-advisor Dr. Muhammad Rizwan Asghar for his guidance and constant advice throughout my research work.

I am grateful to the members of my Ph.D. assessment committee, Prof. Dr. Nelson L.S. da Fonseca, Dr. Marco Di Renzo and Assistant Prof. Dr. Claudio Sacchi. Moreover, I am thankful to Prof. Dr. Khalid Qaraqe, Dr. Imran Shafique Ansari and Assistant Prof. Qammer H. Abbasi for providing me an opportunity to visit the Texas A&M University at Qatar.

I am really thankful to all of my friends who joined me during lunch and coffee breaks, as well as all my colleagues at the University of Trento and Texas A&M University at Qatar for their assistance, support and for the collaboration work.

Last but not least, I am highly indebted to my family. My parents have given me their unequivocal support throughout my studies and they have always been with me despite the distance. I would proudly mention my caring wife for her personal support and great patience at all times. I appreciate the spiritual support of my family. Above all, I thank God for fulfilling my dreams.

Muhammad Usman
Trento, Italy
December 2017

Contents

List of Tables

List of Figures

List of Acronyms

3GPP Third Generation Partnership Project
5G Fifth Generation
AP Access Point
BDP Bandwidth Delay Product
BS Base Station
CA Certificate Authority
CDF Cumulative Distribution Function
CDN Content Distribution Network
CH Cloud Head
CM Cloud Member
CQI Channel Quality Information
CSMA/CA Carrier Sense Multiple Access with Collision Avoidance
CTWindow Client Traffic Window
D2D Device-to-Device
dB decibels
DCF Distributed Coordination Function
DHCP Dynamic Host Configuration Protocol
DIFS DCF Interframe Space
DL DownLink
DMIPS Dhrystone Million Instructions Per Second
DRX Discontinuous Reception
DTN Delay-Tolerant Networks

eNB Evolved Node B
EPC Evolved Packet Core
FDD Frequency Division Duplex
GM Group Member
GO Group Owner
HTTP Hyper Text Transfer Protocol
HTC Human Type Communication
IoT Internet of Things
IoV Internet of Vehicles
KB Kilo Bytes
LAN Local Area Network
LC Legacy Client
LTE Long Term Evolution
MANET Mobile Ad hoc NETwork
MB Mega Bytes
Mbps Mega bits per second
MBS Macro cell Base Station
MC Mobile Cloud
MCC Mobile Cloud Computing
MCPTT Mission Critical Push-To-Talk
MIMO Multiple-Input Multiple-Output
MNO Mobile Network Operator
MTC Machine Type Communication
NFV Network Functions Virtualization
NoA Notice of Absence
OFDM Orthogonal Frequency Division Multiplexing
OFDMA Orthogonal Frequency Division Multiple Access
OPS Opportunistic Power Save
P2P Peer-to-Peer

PCH Primary Cluster Head
PGP Pretty Good Privacy
PKI Public Key Infrastructure
PLS Physical Layer Security
ProSe Proximity Services
QoS Quality of Service
RAN Radio Access Network
RB Resource Block
RTT Round Trip Time
SBS Small cell Base Station
SC-FDMA Single Carrier Frequency Division Multiple Access
SCH Secondary Cluster Head
SDN Software-Defined Networking
SIFS Short Inter-frame Space
SINR Signal-to-Interference-plus-Noise Ratio
SNR Signal-to-Noise Ratio
SORI Secure and Objective Reputation-based Incentive
SSL Secure Sockets Layer
TCP Transport Control Protocol
TLS Transport Layer Security
UE User Equipment
UL UpLink
VM Virtual Machine
VNC Virtual Network Computing
WAN Wide Area Network
WPS WiFi Provisioning Setup

Chapter 1

Introduction

The exponential increase in the number of cellular devices and traffic volume in combination with the looming spectrum represents undoubtedly the primary challenge for the Fifth Generation (5G) of cellular networks. As the size of the network increases, the complexity of managing and monitoring this heterogeneous network also increases. Therefore, 5G networks intend to combine radical solutions to assure more capacity, lower latency, and higher reliability [1, 2]. Such solutions include several emerging technologies such as Network Function Virtualization (NFV), Software-Defined Networking (SDN), massive MIMO and Device-to-Device (D2D) communication. D2D communication represents one such technology that can potentially solve the capacity bottleneck problem of legacy cellular systems [3]. This new paradigm enables direct interaction between nearby Long Term Evolution (LTE) based devices, minimizing the data transmissions in the Radio Access Network (RAN). In the conventional approach, the devices communicate with each other through a common base station, while in D2D approach the devices in close proximity can directly communicate with each other by establishing a Peer-to-Peer (P2P) link (Fig. 2.1). The detailed description of D2D communication is presented in Chapter 2.

1.1 Motivation and Problem Statement

Cellular networks are mainly designed for Human Type Communication (HTC) to support higher data rates and larger data sizes, while Machine Type Communication (MTC) in Internet of Things (IoT) typically exchanges smaller data packets [4]. For example, the minimum size of a radio resource block that can be allocated to a User Equipment (UE) in LTE-Advanced (LTE-A) could be actually too big for the need of IoT applications. On the other hand, large energy consumption required by cellular communication is a major barrier in terms of its adoption as a connectivity platform for IoT applications in smart city scenarios [5]. D2D communication is considered as a viable solution to solve aforementioned problems.

D2D communication has its applications in the areas of Location-Based Services (LBS), social networking, proximity gaming, marketing, multimedia content distribution, cellular traffic offloading, healthcare, Vehicle-to-Everything (V2X) communication, and public safety. All these applications share a large portion of cellular traffic. D2D communication provides an opportunity to offload this traffic to D2D links. This practice provides certain advantages such as, high data rates, low latency and better energy efficiency.

As a consequence of aforementioned benefits of D2D communication, there has been a considerable research in recent years regarding different aspects of D2D communication. In particular, most of the works in the literature focus on resource allocation and interference mitigation in D2D communication. However, there are many scenario-dependent aspects of D2D communication that need further investigations. This includes energy efficiency and privacy in D2D communication. Moreover, there exist very few works in literature that emphasize the architecture of D2D communi-

cation, able to integrate the aforementioned applications.

In addition, despite the aforementioned benefits, D2D communication faces a serious security threat. For instance, in a smart home environment, a malicious user can pretend to be a smart terminal, to which all smart devices are connected in D2D mode and potentially take the control of these smart appliances. Similarly, a user performing proximity-based social networking can be potentially connected with a malicious user. This requires a mechanism to check the security and social status of the devices establishing D2D links. Hence, it becomes imperative to define a secure and energy efficient architecture of D2D communication, having support for preserving privacy of the users. Moreover, the UEs need to bootstrap trust before establishing any connection.

Establishing trust in untrusted environments not only motivates users to participate in D2D-based applications but also enables enterprises to leverage business models based on untrusted environments. Mckinsey estimates that the potential economic impact of IoT will reach 2.7-6.2 trillion USD until 2025 and there will be 75.4 billion connected devices around the globe by that time [6]. Given the size of the market, having a secure and energy efficient D2D communication architecture becomes a primary research challenge. Concerning this, we analyzed both the energy efficiency and the privacy aspects of D2D communication, along with the possibility of integrating different IoT applications.

1.2 Contributions of the Dissertation

Regarding energy efficiency, privacy and trust in D2D communication, we propose different solutions in this dissertation. These solutions include benchmarking energy consumption of D2D UEs in single-hop and multi-hop scenarios, building trust and preserving privacy in D2D UEs. For

most of the cases, we consider content distribution scenario, which is one of the most common applications of D2D communication. A summary of contributions is given below.

- In Chapter 3, we focus on benchmarking energy consumption of UEs in a single-hop network. The idea is built on the concept of Mobile Cloud (MC) [7]. MC is a group of UEs sharing the same service with each other. At this stage, we analyze the overall energy consumption of the observed D2D network when some UEs are participating in multiple applications with different MCs. We compare our results with traditional LTE communication.

- As a next step, we extend our work to multi-hop D2D network wherein UEs communicate with each other using WiFi Direct [8]. We find the optimal value of the group size to save energy in the UEs for a content distribution application. Since WiFi Direct has its own nomenclature, we use the term group instead of MC in Chapter 4.

- Based on D2D communication architecture proposed in Chapter 3, we extend the idea of MC to computational offloading scenarios wherein we find the optimal data size of offloading computations to save energy in the UEs. We show that the source UE can make a decision to offload computations to a local MC or to a remote cloud.

- At this stage, we decided to include the privacy aspect in our architecture and propose to bootstrap trust in D2D networks (Chapter 6). The problem of preserving privacy of the users was addressed thereafter (Chapter 7).

It is important to note that although there is a lot of work in literature on resource allocation and interference management in D2D communication but the focus of this dissertation is the system architecture and ap-

plications of D2D communication. For this reason, we validate our ideas through simulations, which are mainly carried out in NS-3 or MATLAB. As a next step, we started validating our ideas by implementing them in Software-Defined Radios (SDRs), such as ExpressMIMO2 by Open Air Interface (OAI). Since the hardware emulation of our work is in the beginning phase, we did not include it in this dissertation. However, it is worth mentioning that a part of our hardware implementation work is accepted for presentation in GLOBECOM 2017.

For brevity, we did not include some findings of our work in this dissertation. However, a complete list of publications can be found in Appendix A.

1.3 Organization of the Dissertation

The dissertation is divided into two parts. The first part summarizes our findings related to energy efficiency in D2D communication, while the second part discusses the privacy aspects of D2D communication. Each part contains following chapters.

Chapter 2 briefly elaborates the classification, standardization efforts and applications of D2D communication.

PART A: Energy Efficiency in D2D Communication

Chapter 3 analyzes the energy consumption of UEs in participating in different applications with multiple MCs and compares this energy consumption with the conventional cellular network, such as LTE.

Chapter 4 extends the proposed solution in Chapter 3 to multi-hop D2D networks for a content distribution application. This chapter analyzes the end-to-end throughput as a tradeoff to energy gain by varying the number of UEs in each WiFi Direct group.

Chapter 5 discusses the possibility of offloading computations to D2D network and finding an optimal solution in terms of energy consumption. The computations of various sizes are offloaded to MCs and remote clouds in order to compare the energy consumption of source UE.

PART B: Privacy in D2D Communication

Chapter 6 proposes to bootstrap trust in D2D networks wherein UEs need to communicate with unknown users. To this end, Pretty Good Privacy (PGP) and reputation-based models are proposed.

Chapter 7 presents a security scheme to secure the caching contents (at D2D nodes or any other caching server) with a simultaneous potential of reducing duplicate contents from the caching server by dividing a file into smaller chunks and using convergent encryption.

Chapter 8 finally concludes the dissertation by summarizing the chapters presented. It also points out some future research directions emerging from this work.

Appendix A reports a list of publications related to the work presented in this dissertation, as well as other publications during PhD time.

Chapter 2

D2D Communication

2.1 Introduction

D2D communication is a new paradigm in cellular networks, which enables direct interaction between nearby UEs, minimizing data transmissions in RAN [7]. In conventional cellular networks, the UEs communicate with each other through a common base station; whereas, in D2D, the UEs in close proximity can directly communicate with each other by establishing a P2P link between them as illustrated in Figure 2.1.

2.2 Classification of D2D Communication

Based on the used spectrum, D2D communication can be classified into two broad categories shown in Figure 2.2: inband D2D and outband D2D.

2.2.1 Inband D2D

In the inband D2D communication, same licensed spectrum is used for both cellular and D2D communication. The motivation behind choosing the same spectrum is the high control over the network in cellular spectrum. Inband communication can further be divided into two categories: underlay

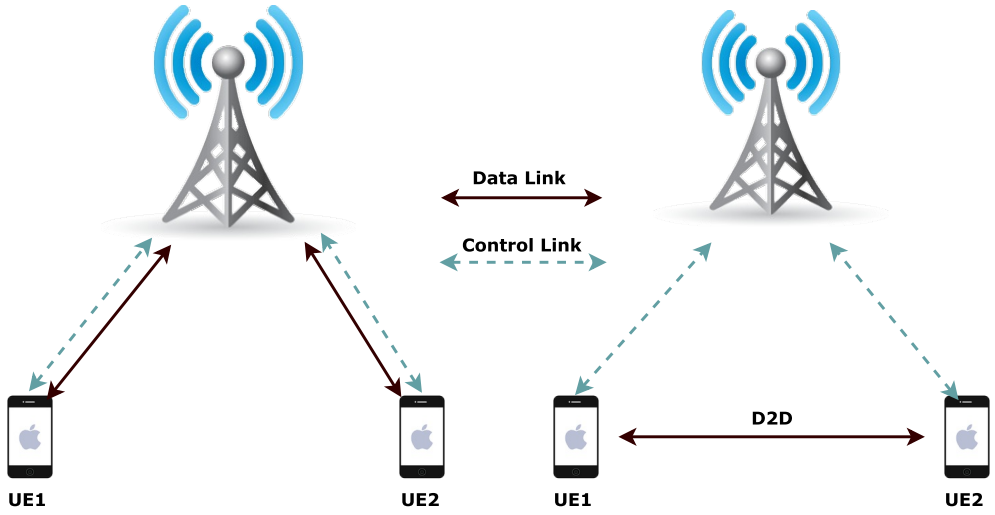


Figure 2.1: Conventional cellular communication (left side) versus direct D2D communication (right side).

and overlay D2D communication.

2.2.1.1 Underlay Inband D2D

In underlay D2D communication, D2D and cellular links share same cellular resources.

2.2.1.2 Overlay Inband D2D

In overlay D2D communication, the D2D links are given dedicated radio resources from the cellular spectrum.

The main disadvantage of inband D2D communication is the interference caused by D2D links to the cellular network.

2.2.2 Outband D2D

In outband D2D communication, D2D links use unlicensed spectrum. The motivation behind using unlicensed spectrum is to minimize interference between D2D and cellular links. This requires an extra interface and adopts

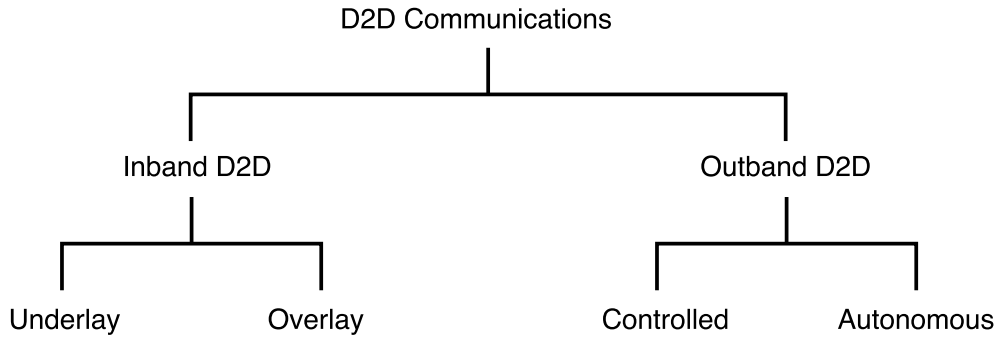


Figure 2.2: Categories of D2D communication.

other wireless technologies like Wifi Direct [9], Bluetooth [10] and ZigBee [11]. Outband communication can further be divided into two categories: controlled and autonomous D2D communication. In the controlled D2D communication, the control is given to the cellular network [12], [13], [14] and [15]. In the autonomous D2D communication, the cellular communication is kept controlled while the control of D2D communication is given to the UEs [16]. The main disadvantage of outband D2D communication is the uncontrolled nature of unlicensed spectrum.

2.3 Standardization Efforts

D2D communication has been addressed in release 12 [17] of Third Generation Partnership Project (3GPP) under the name of Proximity Services (ProSe). In particular, 3GPP RAN working group proposed two basic functions, ProSe discovery and ProSe communications, in TR 36.843, Rel. 12 [18]. However, 3GPP has initially targeted public safety applications in D2D communication. In this regard, Table 2.1 presents the supported ProSe functions (ProSe discovery and ProSe communications) for public safety and non-public safety applications in three different network scenarios.

The in-coverage scenario represents a situation when all UEs lie in the

coverage area of the cellular network. Similarly, in the out-of-coverage scenario, all UEs are located outside the coverage area of the cellular network. Partial coverage scenario represents a situation when some UEs are located outside the coverage area of the cellular network. The UEs at the edge of the coverage area relay the information of out-of-coverage UEs to the base station or core network.

Table 2.1: Supported ProSe functions in 3GPP release 12 to enable D2D communication in public safety and non-public safety applications.

Scenarios	**Within Network Coverage**	**Outside Network Coverage**	**Partial Network Coverage**
Supported Applications	*Supported ProSe Functions*		
Non-Public Safety	Discovery	-	-
Public Safety	Discovery, Communication	Communication	Communication

Regarding D2D communication, release 13 [19] of 3GPP focuses on Mission Critical Push-To-Talk (MCPTT) over LTE, which is a key enabler for many public safety features, such as person-to-person calls, group calls, group management and user management.

2.4 D2D Applications

The direct interaction between UEs improves spectral utilization, overall throughout and energy efficiency, while enabling many P2P and location-based services. Moreover, D2D communication plays a key role in enabling interoperability between critical public safety networks and ubiquitous commercial cellular networks.

Exploiting D2D communication can potentially enhance the role of IoT

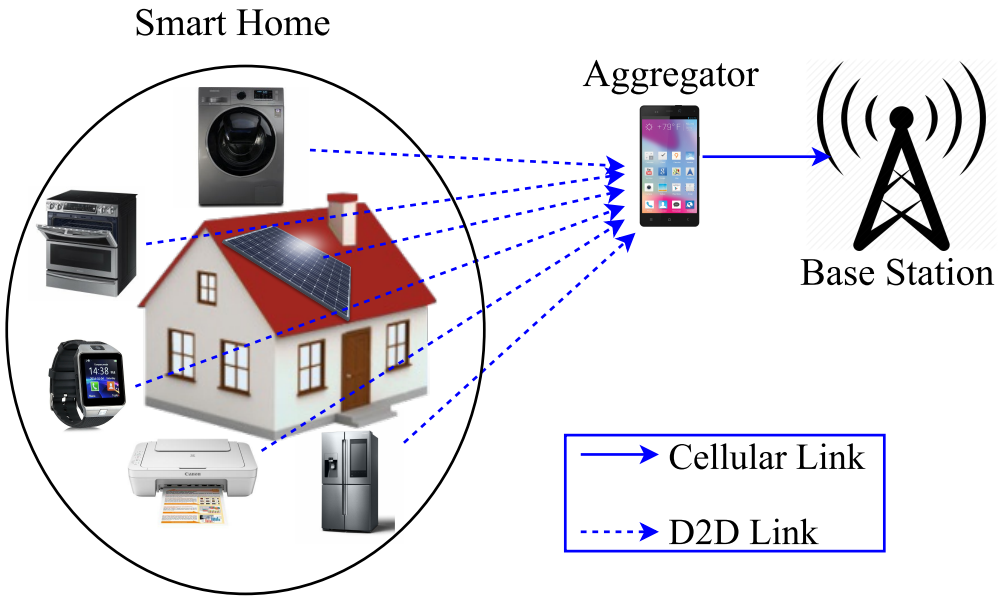


Figure 2.3: D2D communication as an aggregator for IoT traffic: Home appliances are connected with a smartphone over a D2D link. The smartphone aggregates the traffic from different sensing nodes and sends it to the base stations when it has sufficient data to be transferred.

in future smart cities. One such example is the Internet of Vehicles (IoV) wherein vehicles communicate with each other in D2D mode, without traversing any data traffic to the base station. The nearby vehicles can be automatically alerted before any change of lane. This helps vehicles to better respond to emergency situations, thus avoiding potential accidents. Moreover, the traffic on the road can be prioritized. That is, school buses and ambulances can be assigned higher priorities over normal vehicles.

Other applications of D2D communication include, but are not limited to, social networking, proximity gaming, marketing, multimedia content distribution, cellular traffic offloading, animal housing and management, healthcare, surveillance, V2X communication, and public safety. D2D communication node, such as a smartphone, can act as a data aggregator for many smart city applications. In this regard, IoT devices can be clustered together based on their proximity. A smartphone can aggregate the traffic

of the cluster to the cellular network to improve communication and energy efficiency. As an example, Figure 2.3 presents a smart home scenario wherein smart appliances are connected with a cellular network through an aggregator. Direct D2D communication is considered as a connectivity mechanism between smart appliances and the aggregator.

2.5 Chapter Summary

Due to emergence of new data-intensive applications, telecom operators are struggling to accommodate existing demand of mobile users. The current 4G cellular technologies are still lagging behind the users' data demand. The researchers are looking for new paradigms in addition with the conventional methods in cellular communication. D2D communication is one of the such paradigms.

In this chapter, we summarized the general information regarding classification, standardization and potential applications of D2D communication. In the coming chapters, we elaborate further our findings to address the research challenges concerning energy efficiency and privacy issues in D2D communication.

Part A

Energy Efficiency in D2D Communication

Chapter 3

Energy Efficiency in Single-hop D2D Communication

D2D communication enables direct communication between nearby UEs using cellular or ad hoc links thereby improving the spectrum utilization, system throughput, and energy efficiency of the network. Exploiting MC based D2D communication architecture underlying LTE cellular network has a huge importance in reducing the transmission power of the UEs, resulting an improved battery life. This chapter proposes a novel hybrid D2D communication architecture wherein a centralized SDN controller communicates with the Cloud Head (CH) in order to reduce the number of LTE communication links, thereby improving the energy consumption. In addition, UEs can participate and perform operations in multiple MCs simultaneously. The obtained simulation results confirm improved energy efficiency as compared to the legacy LTE network.

3.1 Introduction

The mobile data traffic is growing exponentially and is forecasted to surpass 24.3 Exabyte/month by 2019 [20]. Mobile operators need more capacity to meet the demands of mobile users for higher data rates and lower latency.

Legacy cellular communication systems often become overloaded [21], while D2D communication in MCs can offer solutions to improve system capacity [21].

MCs exploit D2D communication to enable a variety of services that can be used in applications such as video streaming, public safety, rich content media offloading, online gaming and energy efficient content distribution [22]. The MCs enable UEs to share their resources/services over D2D links, while preserving connectivity to the overlay network. Many aspects of MCs have been separately studied in the past, in the field of ad-hoc networks.

MC works in three different stages: *cloud formation*, *cloud operation* and *cloud maintenance* [23]. An important issue in exploring D2D communication for MC is the design of a composite architecture that accounts for dynamic characteristics of the UEs and their resources in all three stages. The architecture should be capable of establishing the rules on how resources/services are shared. In the formation of an MC, several device discovery mechanisms can be used that exist in literature, and they can be classified in two general categories: (1) centralized device discovery, where all UEs register their location for other UEs to be identified, and (2) distributed device discovery, where each UE broadcasts in a periodic time interval and listens to receive other UEs' identities in other time slots.

In this chapter, we propose a hierarchal SDN-based (hybrid) architecture for the formation and operation of MCs. We propose the idea of local and global SDN controllers that make the process of cloud formation and operation scalable, reliable and energy efficient. We divide the cloud formation into two phases. One is the training phase, where a UE initiates an MC, broadcasting cloud formation request to the UEs in the proximity over a WiFi link. Upon successful formation, the cloud is registered to the global SDN controller. In the second phase, the mature phase, the global SDN controller will have a global view of all the served MCs with the ser-

vices they offer. At that point, the global SDN controller is able to set-up the MCs upon users' requests.

The proposed architecture is analyzed using the following performance indexes: energy consumption of UEs in cloud formation and operation stage and the cloud size.

The rest of the chapter is organized as follows. System architecture and mathematical formulation are presented in Section 3.2. Section 3.3 discusses the performance evaluation and presents the results from the mathematical model. Section 3.4 reviews related work. Section 3.5 describes the chapter summary.

3.2 Software-Defined MCs

This section describes architecture and mathematical foundation of the proposed system.

3.2.1 System Architecture

We propose a hierarchal SDN architecture wherein each UE has an installed SDN application for cooperation in the MCs. This SDN application of CH is regarded as the local SDN controller. The MC is formed on demand and the SDN application uses a hybrid approach to create an MC for the demanded service. The global SDN controller, which resides in the Internet, has a global view of all MCs exist in its range (see Figure 3.1). To address the problem of scalability, we propose to have a global SDN controller for every 3 or 4 Evolved Node B (eNBs).

The global SDN controller maintains a database of all MCs, saving identity of each UE with all the services it can share with others. In case of resource sharing services, the details of resources are also stored in the SDN database. Once the database is matured, the global SDN controller,

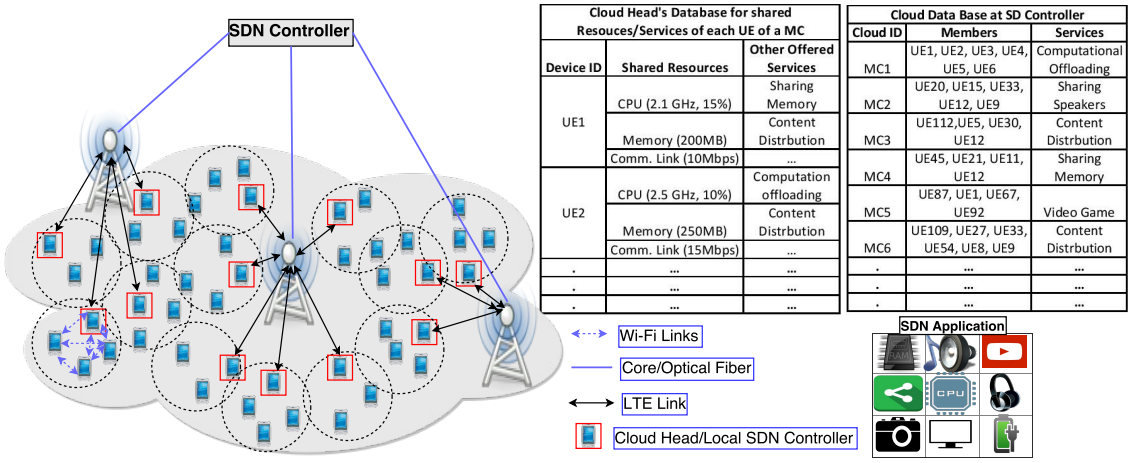


Figure 3.1: The proposed system architecture of SDN-based D2D communication.

to save energy consumption of the UEs in service discovery phase, performs the cloud formation, without involving local SDN controller. Any change in the service (such as a UE leaves the MC or a new UE joins or a UE changes its shared resources, *e.g.,* increases the size of shared memory *etc.*) will be reflected to the global SDN controller immediately through the LTE interface of CH.

The initiator broadcasts a request for the cloud formation over WiFi interface, for sharing a particular service. The UEs in the vicinity, interested in sharing that service, respond with their resources/services. SDN application in each UE maintains a database of all services and resources, a UE is willing to share. Once a request for a cloud formation for a particular service, is received from an initiating UE, all interested UEs share the complete database with the initiator. The initiator shares this database with the global SDN controller over the LTE interface. The global SDN controller registers the MC, selects a CH (mostly the initiator) and assigns an authentication key to the MC for further communication between cloud members. The key is shared with UEs of the MC to securing them from any malicious attack. The complete signaling procedure for cloud

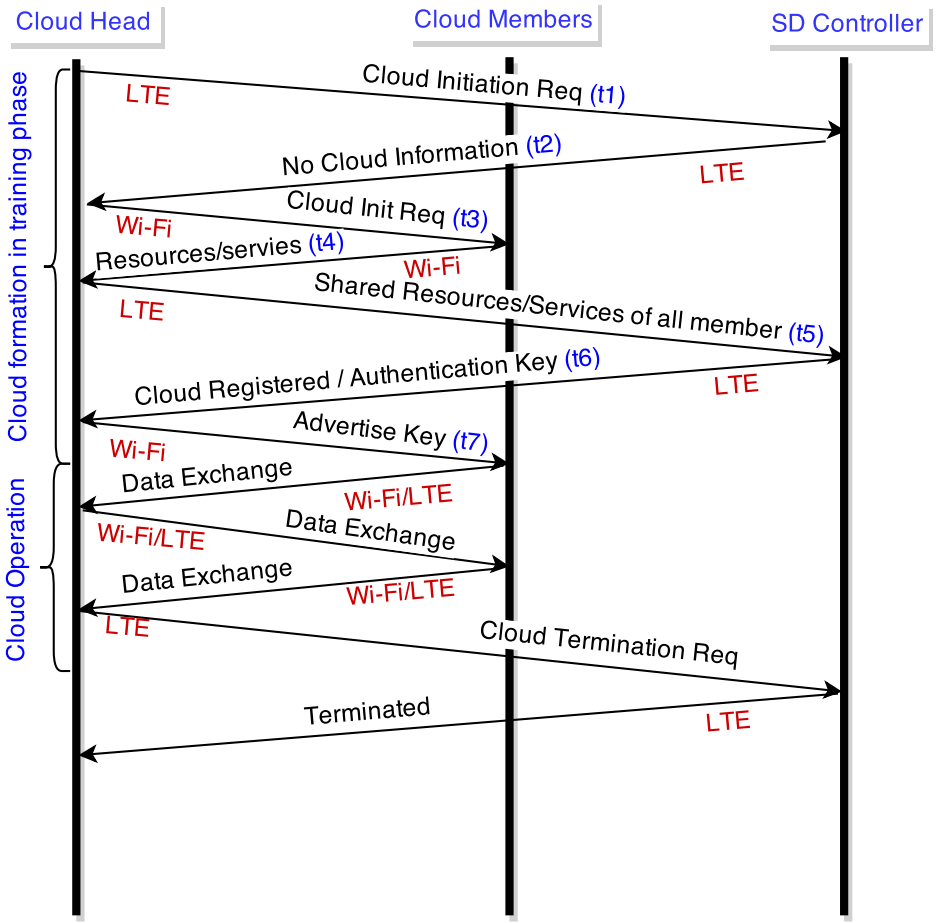


Figure 3.2: Description of signaling for cloud formation.

formation and operation is shown in Figure 3.2.

Our architecture enables a UE to participate in multiple MCs providing different resources/services. This raises two important issues that need to be considered:

- The operations belonging to different MCs should be performed in a complete isolation (one of the goals of *virtualization*), i.e., to avoid collisions between the operations.
- There should be a proper *allocation of resources* based on the Quality of Service (QoS) requirement of different services (operations). For

example, let us consider that one of the MCs provides services for file transfer and the other for video conferencing. In such situations, we need to deploy a dynamic resource allocation scheme that will take into account the service requirements with the final goal of achieving an improved network performance in terms of better spectrum utilization and/or a better network throughput.

There are several studies concerning the design and implementation of controllers (e.g., centralized, distributed, hierarchical, etc.), where each has its merits and demerits. However, the hierarchal architecture better fits our need in a way that it helps to address the problem of scalability and efficient resource utilization by lowering the communication (i.e., scarce LTE spectrum) load with the global SDN controller. The distribution of different functionalities to different levels of the controllers (i.e., local and global) helps to reduce unnecessary communication with the higher-level controllers, which use scarce radio resources (i.e., LTE spectrum). For example, the local SDN controller (initiator/cloud head) can independently make and break clouds without involving the central-controller. In addition, the hierarchical architecture is very convenient for scalability. The number of UEs participating in an MC could increase as far as the processing capacity of the CH has not been reached. Moreover, the flexibility of having local decisions carried out by local SDN controller enables each cloud to work in a distributed manner, as well.

In order to reduce the communication overhead between CH and SDN controller, the CH sends periodic updates to the SDN controller after a preset time, informing all the changes (i.e., users leaving or joining the cluster) that happened during this time interval. This significantly reduces the Ping-Pong effect of users joining and leaving the cluster, resulting in an improved performance of the network in terms of delay and overheads. The database residing in the CHs reduces the possible delay incurred in

retrieving information from the SDN controller.

3.2.2 Energy Model

Besides the improvements in bit rate and spectral efficiency, D2D communication also offers better UE battery life. LTE uses Single Carrier Frequency Division Multiple Access (SC-FDMA) for uplink instead of Orthogonal Frequency Division Multiplexing (OFDM), which suffers from poor power efficiency [24]. In addition, Discontinuous Reception (DRX) technique is employed in order to reduce the UE power consumption, as defined in the standard. Due to the boom in data services, several applications need a higher computational power in UEs leading to higher energy consumption. D2D communication offers a promising improvement in power saving by reducing the number of LTE communication links, which needs more power as a result of the longer distance between the base stations and the UEs.

This chapter proposes a mathematical formulation for energy consumption of UEs, while communicating through MC or through LTE links over cellular network. We compare the energy consumption in both cases and found that we can save a significant amount of energy if D2D links are exploited.

3.2.2.1 Energy Consumption During Cloud Operation Phase

Let we have M MCs each having N_i UEs for $i = \{1, 2, 3, \ldots M\}$. Then following relation represents the number WiFi links in the system.

$$N^{WiFi} = \sum_{i=1}^{M-1} N_i \ , \tag{3.1}$$

where M is the number of CHs that communicate with cellular network over LTE links (i.e., the communication between the CH and the global

SDN controller). Thus, the number of LTE links will be M in this case. Let n_i is the number of UEs participating in multiple MCs.

$$n_i = \rho_i \times N_i \ ; 0 \leq \rho_i \leq 1 \tag{3.2}$$

where ρ_i is the percentage of UEs that belong to multiple MCs. Based on the above model, the energy consumption of a UE for transmission on a communication link is given by the following expression:

$$E_{Tx} = P_{Tx} \times t \ , \tag{3.3}$$

where P_{Tx} is the power consumption during transmission and t is the transmission time of a UE. The average energy consumption for M LTE links can be given by the following expression:

$$E_{avg}^{LTE_M} = \sum_{i=1}^{M} E_{Tx_i}^{LTE} \ . \tag{3.4}$$

Similarly, the average energy consumption for all WiFi links is given by the following expression:

$$E_{avg}^{WiFi} = \sum_{i=1}^{M} \left[\sum_{j=1}^{N_i - n_i} E_{Tx_j}^{WiFi} + 2 \sum_{j=1}^{n_i} E_{Tx_j}^{WiFi} \right] . \tag{3.5}$$

For simplicity, we consider that a UE can participate in a maximum of two MCs. The first term in (3.5) represents the energy consumption of the UEs participating in a single cloud and the second term represents the energy consumption of the UEs participating in two clouds. Thus, the total energy consumption in D2D case will be:

$$E_{Tot}^{D2D} = E_{avg}^{LTE_M} + E_{avg}^{WiFi} \ . \tag{3.6}$$

Now, we consider the case when there is no D2D communication and all devices have to communicate through eNB only. The average energy consumption in this case will be:

$$E_{Tot}^{LTE} = \sum_{i=1}^{M} \sum_{j=1}^{N_i} E_{Tx_j}^{LTE} \ , \tag{3.7}$$

where E_{Tot}^{LTE} and E_{Tot}^{D2D} represent the average energy consumptions of UEs in legacy LTE case and in the cloud operation case, respectively.

3.2.2.2 Energy Consumption During Cloud Formation Phase

We can estimate the energy overhead due the time the UEs consume in cloud formation phase. During the training period, the time spent in the cloud formation phase is the summation of times from t_1 to t_7 (see Figure 3.2). Once the SDN controller's database is mature enough to make the clouds and assign the cloud head, the time consumed to make a cloud will be reduced. The following relation estimates the average energy consumption in the cloud formation phase.

$$E_{C.F.}^{Training} = P_{Tx_j}^{LTE} \times t_{Training}^{LTE} + P_{Tx_j}^{WiFi} \times t_{Training}^{WiFi} \ , \tag{3.8}$$

where $E_{C.F.}^{Training}$ is the average energy consumption of a UE in the training phase of cloud formation, P_{Tx}^{LTE} is the power consumption of a UE for the transmission on an LTE interface during the cloud formation, $t_{Training}^{LTE}$ is the time spent in the transmission on LTE links, P_{Tx}^{WiFi} is the power consumption of a UE for the transmission on a WiFi interface during cloud formation and $t_{Training}^{WiFi}$ is the time spent in the transmission on WiFi links.

Table 3.1: Numerical Parameters of WiFi and LTE considered in the simulations.

Parameter	**Value**
Backoff time (WiFi)	0.1554 [s]
Size of Packet (WiFi)	1500 [Bytes]
Modulation and Coding Scheme (WiFi)	24.10^{-6} [s]
Minimum Data rate (LTE)	5.2 [Mbps]
Maximum Data rate (LTE)	25.5 [Mbps]
Minimum Data rate (WiFi)	7.2 [Mbps]
Maximum Data rate (WiFi)	56.0 [Mbps]

Similarly, the following relation gives the energy consumption of a UE in the mature phase of the cloud formation.

$$E_{C.F.}^{Mature} = P_{Tx_j}^{LTE} \times t_{Mature}^{LTE} + P_{Tx_j}^{WiFi} \times t_{Mature}^{WiFi} \ . \tag{3.9}$$

3.3 Performance Analysis

The model presented in previous section, estimates the energy consumption of UEs in transmitting on D2D links and LTE links in both cloud formation and operation phase. In Figure 3.3, we compare the energy consumption of UEs for different percentages of ρ. To estimate the transmission time over WiFi and LTE links, we use the model presented in [25]. These values are presented in Table 3.1. We consider a simple scenario wherein each UE has to upload a $20MB$ data to the eNB using D2D (WiFi) and LTE links. For WiFi links, we randomly generated the data rates between the range of $7.2Mbit/s$ to $56.0Mbit/s$ (maximum achievable data rate for single spatial stream). Similarly, for LTE links, we define the range from $5.2Mbit/sec$ to $25.5Mbit/sec$ [25] for UEs belonging to category 1 and 2 according to 3GPP release 8.

We find that for $\rho < 80\%$, the D2D communication always consume

less energy than LTE. In case a UE is participating in multiple MCs, the UE has to maintain multiple WiFi links and it consumes more energy. For instance, to communicate with two MCs, the energy consumption of the UE will be almost doubled. We can note in Figure 3.3 that if no UE is participating in multiple clouds then the energy saving can go up to 45.9% and we can still save 3.5% energy even if 70% of the UEs are participating in multiple clouds. Moreover, in our analysis, we did not consider the case to use the cooperation capabilities of D2D communication, where UEs can cooperate with each other and partition the data into small chunks to send it to CH. In this case, the energy consumption will be reduced further as each UE will transfer a small portion of the data.

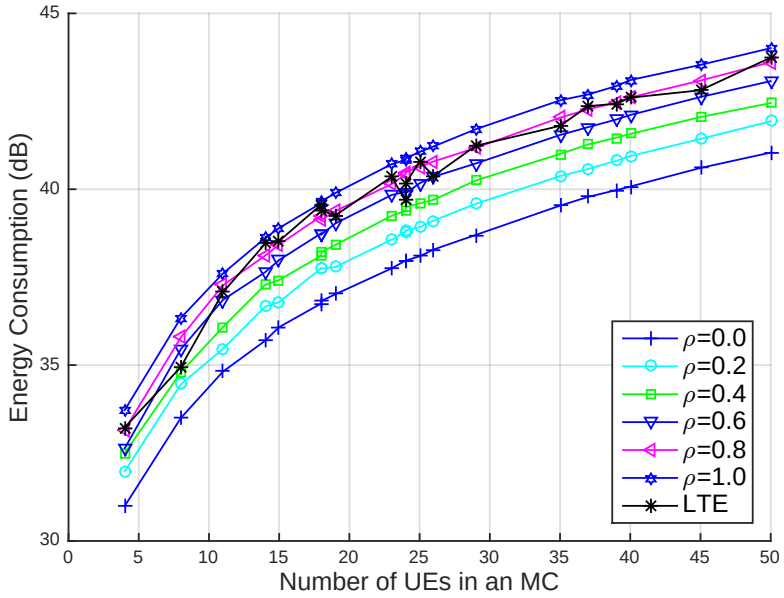


Figure 3.3: Comparison of energy consumption in operation stage of an MC.

In the proposed architecture, the average cloud formation time in mature phase is reduced to the time the cloud head or SDN Controller takes to authenticate the request and assign resources. Figure 3.4 shows the energy footprints of a single UE in training and mature phase during the

process of cloud formation. The graphs are plotted using Eqs. 3.8 and 3.9 of Section 3.2. The results show that we can save up to 96.96% energy, consumed in the training phase otherwise.

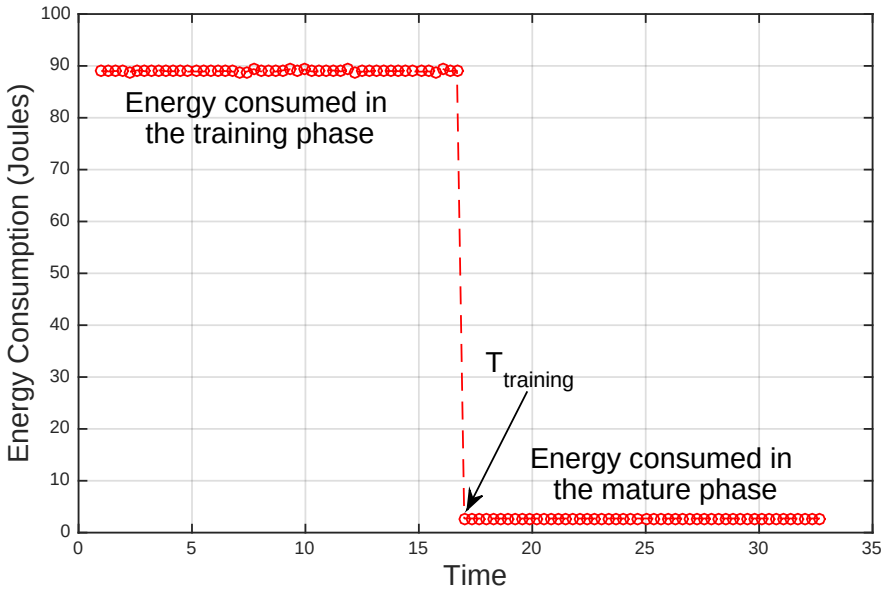


Figure 3.4: Comparison of energy consumption during training and mature phase.

In Figure 3.4, we consider the case of single cloud that can be generalized to multiple clouds with increased training period. During the training phase, the cloud formation energy fluctuates around $90J$ while during mature phase it is reduced to just $3J$. It is due to the reason that in the mature phase the number of communications with LTE and with other peers is reduced. In the mature phase, the database of services and resources of proximity users at the cloud head tends to become more mature and cloud head does not need to communicate with SDN controller to make a cloud. It can rather just inform the SDN controller about the cloud formation and uses its own database to perform the operation. In this way, we can save a significant amount of energy and almost no energy is wasted in device and service discovery.

3.4 Related Work

The MC represents the logical evolution of the concept of moving the distributed cloud more and more towards the user side. Satyanarayanan et al. [26] use the term 'cloudlet' to describe resource-rich computing environment located at the edge of the network and in the proximity of mobile users. The UEs can use this environment to offload computations and execute virtualized tasks. In [27], Hassan et al. propose a D2D-based MC architecture, where MC coverage area is divided into clusters (logical regions) of UEs and comprises a primary cluster head (PCH), a secondary cluster head (SCH) and standard UEs. PCH and SCH, which are selected based on the residual energy and Signal-to-Interference-plus-Noise Ratio (SINR) of the UEs, multicast information to the UEs of their respective clusters.

Mass et al. [28] propose an MC system that implements device discovery based on the audio data obtained from the user environment. This centrally controlled cloud system follows client-server architecture, where clients (UEs) send synchronized time series recordings to the server (Amazon Cloud) that runs a clustering algorithm on the time series in order to group them based on their audio similarity. The algorithm is not energy efficient, as clients have to be continuously synchronized with the server through cellular interface.

Doppler et al. [29] propose a distributed device beaconing scheme that exchanges small data packets and works with the assistance of cellular network. The devices transmit their beacons using Orthogonal Frequency-Division Multiple Access (OFDMA), based on the LTE beacon structure. The MC formation is not on demand rather a background network is formed based on beacon messages irrespective of the will of the devices to share resources/services.

Wu et al. [30] propose FlashLinQ, a synchronous OFDM based system, to perform device discovery, channel allocation and link scheduling in the licensed spectrum. The distributed channel allocation in licensed spectrum is claimed to give significant gain over conventional 802.11 systems.

In the proposed system, we use distributed device discovery mechanism exploiting WiFi links of the UEs with partial assistance from cellular network. The relative high bandwidth of an IEEE 802.11 cell and large coverage area of the cellular network makes the proposed cloud architecture reliable and energy efficient.

3.5 Chapter Summary

This chapter compares the energy consumption of UEs on D2D links with normal cellular links for the case of network-assisted D2D communication. In this chapter, we propose a multilayer SDN based architecture for D2D communication. The local SDN controller manages the information flow within an MC while global SDN controller has a global view of multiple MCs and manages communication among different MCs.

However, there are many scenarios when UEs need to communicate with each other while there is no connection with the cellular infrastructure. For example, in case of infrastructure damage due to disaster or hotspot traffic situation due to network overload. Concerning this, we propose a solution in the next chapter, which works on WiFi Direct based D2D communication between different MCs.

Chapter 4

Energy Efficiency in Multi-hop D2D Communication

In the previous chapter, we benchmark the energy consumption of a network-assisted D2D communication, where some UEs can participate in more than one cloud. But still the communication was restricted to one hop only. However, there are many scenarios when it becomes imperative for UEs to establish multi-hop links between them, employing other UEs as relay nodes.

In this chapter, we propose a novel power saving protocol that aims at optimizing energy consumption and throughput of UEs by controlling the WiFi Direct group size and transmit power of UEs in multi-hop D2D communication. WiFi Direct is a new technology that enables direct D2D communication. This technology has a great potential to enable various proximity-based applications such as multimedia content distribution, social networking, cellular traffic offloading, mission critical communications, and IoT. We model a content distribution scenario in NS-3 and present the performance evaluation. Our simulation results demonstrate that even a small modification in the network configuration can provide a considerable energy gain with a minor effect on throughput. The observed energy saving can be as high as 1000% for a throughput loss of 12%.

4.1 Introduction

The exponential increase in the number of cellular devices and traffic volume is undoubtedly a primary challenge in modern-day cellular networks. Current cellular networks are undeniably overloaded to meet the escalating user's demands for higher bandwidth and higher data rates. Therefore, 5G cellular networks intend to combine radical solutions to assure more capacity. D2D communication is one such solution that can potentially solve the capacity bottleneck issue of legacy cellular systems. This new paradigm enables direct interaction between nearby UEs, thereby minimizing the data transmissions in the radio access network [7]. By leveraging D2D communication, users experience various benefits, such as lower transfer delays, higher data rates, and better energy efficiency [31, 32]. Due to these potential benefits, D2D is at the forefront of standardization and research efforts. Such interests are spurred by the introduction of WiFi Direct [33] in modern-day smartphones, from Android 4.0 onwards.

WiFi Direct, formally known as WiFi P2P, is a new technology standardized by WiFi alliance [34] aimed at enabling D2D communication between nearby UEs, without requiring a wireless Access Point (AP). WiFi Direct is built on top of the IEEE 802.11 infrastructure mode, where UEs negotiate to take the roles of an AP and clients. By doing so, WiFi Direct inherits all QoS, security, and power saving mechanisms, deployed for infrastructure mode WiFi over the past years [35]. In addition, WiFi Direct defines mechanisms to save power in the UEs performing AP-like functionality. However, all these mechanisms are defined for intragroup communications only. Whereas, there are many situations that require intergroup (multi-hop) communications to route the traffic towards a destination. Multi-hop communication potentially changes the network's characteristics, thus resulting in further investigations. In the recent literature,

some solutions exist wherein the primary focus is to enable multi-hop communication using WiFi Direct [35–38]. However, no solution considers multi-hop communication as a virtue to optimize the energy efficiency and the throughput.

This chapter aims at optimizing the network's performance in terms of energy efficiency and throughput of UEs in multi-hop D2D networks. By UEs, we refer to WiFi Direct clients, APs, and the devices that do not support WiFi Direct but can join the network as Legacy Clients (LC). We explore different parameters and analyze how they change the performance of the network. In particular, we analyze the impact of group size (*i.e.,* number of UEs per group) and the transmit power of the UEs. More specifically, we propose a power saving scheme that optimizes the energy efficiency and throughput. To the best of our knowledge, this is a first work that investigates an optimal group size and transmit power of UEs to optimize the network's performance.

The rest of the chapter is organized as follows. Section 4.2 presents various motivating scenarios that exploit multi-hop D2D communication. Section 4.3 demonstrates the operations of WiFi Direct and different power saving schemes as defined in WiFi P2P standard. Section 4.4 presents the design overview and key idea addressed in this work. Section 4.5 reports on performance analysis. Related work is reviewed in Section 4.6 followed by a discussion in Section 4.7. Finally, we present the chapter summary in Section 4.8.

4.2 Motivating Scenarios

There are many situations wherein users need to communicate with each other despite intermittent or no connectivity with the Internet. Public protest is one possible scenario where people need to have mutual connec-

tions despite authorities' attempt to cut the Internet connectivity. Making a network in an airplane could be another scenario wherein a group of friends need to establish communication links among them. Similar is the case of travelling through a subway, which does not provide an Internet connection in most cases. Another situation can be a camping trip to a desert or a disaster scenario, where cellular infrastructure is not available or has been completely damaged.

There can be other scenarios where people have Internet connections but are not willing to utilize it to save cellular data. For instance, during a large concert, people want to chat and share photos with each other. A big conference could be another situation, where attendees are in the proximity of each other and want to share their thoughts and research activity with one another.

A cellular connection can be utilized in some of the aforementioned scenarios, but is expensive in terms of data cost, power consumption of UEs, cellular resources, and file transfer time. D2D communication can be a convenient option, which allows UEs to directly interact with each other.

In all the aforementioned scenarios, users may like to chat with each other and/or send a file to a specific user or broadcast a file to a certain set of users in the vicinity. As the motivation of this chapter is to benchmark the energy consumption and overall throughput of the network for different system configurations, hence, without loss of generality, we consider a content distribution scenario, where a user likes to send a file to all users in the proximity, to make sure that every UE in the network has some data to receive. However, the proposed techniques are equally applicable to situations wherein a UE needs to communicate with a specific UE, while other UEs work as relays only. The energy gain in those scenarios is left for future work.

4.3 WiFi Direct and Power Saving

In an ordinary WiFi network, clients discover and connect to APs. The functional roles of an AP and clients are predefined and UEs univocally act either as a client or as an AP. However, in WiFi Direct, these roles are not predefined but dynamic and logical that are negotiated during the group formation, allowing a UE to behave both as a client or as an AP [35]. UEs can communicate by establishing groups that are functionally identical to an ordinary WiFi infrastructure mode network. The UE that implements the functionality of an AP is generally referred to as a Group Owner (GO), while the UEs implementing the roles of a client are often termed as the Group Members (GMs). Within a group, WiFi Direct utilizes the IEEE 802.11 a/b/g/n infrastructure mode, where UEs can transmit either at 2.4 GHz or 5 GHz [36].

The UEs that support WiFi Direct go through a group formation process and negotiate the roles of the GO and the GMs. Three group formation procedures are defined in the standard [33]: standard, autonomous, and persistent. In a standard group formation procedure, UEs listen on channels 1, 6, and 11 in the 2.4 GHz band and exchange an intent value to become a GO [33, 39]. A UE with the highest intent value becomes the GO and others act as the GMs. After defining these roles, the UEs go through a WiFi Provisioning Setup (WPS) phase. Thereafter, the GO assigns IP addresses to the GMs using Dynamic Host Configuration Protocol (DHCP). In the autonomous group formation procedure, a UE declares itself as a GO and initiates the WPS process and IP assignments to create a group. During the persistent group formation procedure, the UEs exchange invitation messages to restore the roles of the group they were previously associated with. This sufficiently reduces the time for WPS process as the stored credentials of the previous group can be utilized.

4.3.1 Power Saving Modes in WiFi Direct

WiFi Direct defines two power saving modes to save power in battery-constrained devices acting as an AP: (i) Opportunistic Power Save (OPS) and (ii) Notice of Absence (NoA). WiFi Direct clients can use the legacy power saving protocols defined in the WiFi infrastructure [40].

4.3.1.1 Opportunistic Power Saving

OPS mode leverages the sleeping intervals of WiFi Direct clients using legacy power saving mode. The GO advertises a time window, referred to as Client Traffic Window (*CTWindow*), in all beacon frames and probe responses. It specifies the minimum amount of time the GO will stay awake after receiving the beacon frames. WiFi Direct clients can send their frames during this duration. After *CTWindow*, if the GO realizes that all WiFi Direct clients are in doze state, it can go to sleep mode until the next beacon is scheduled. During this interval, if one of the clients leaves the power saving mode, the GO needs to stay awake until all clients go into the power saving mode. It is important to note that in OPS mode, the decision for a GO to go to sleep mode entirely depends upon the WiFi Direct clients. To give a GO more control on its sleep intervals, WiFi Direct specifies NoA in the power saving mode.

4.3.1.2 Notice of Absence

Unlike OPS mode, in NoA, the GO advertises *absence periods*, during which WiFi Direct clients are not allowed to access the channel, regardless of whether they are in an active mode or a power save mode. *Absence periods* are also advertised in beacon frames and probe responses using signaling elements. A NoA schedule is defined via four parameters: (i) count specifies the number of *absence periods* scheduled during current NoA schedule, (ii)

start time specifies beginning of first *absence period* post current beacon frame, (iii) interval specifies the time between consecutive *absence periods*, and (iv) duration that specifies the length of each *absence period*. A GO can cancel or update NoA schedule by updating the signaling elements.

4.4 Proposed Methodology

In this chapter, we aim at optimizing energy efficiency of multi-hop D2D networks. More specifically, we present a power saving scheme for WiFi Direct clients and GOs that consider two parameters of WiFi Direct protocol: (i) group size and (ii) transmit power of UEs.

4.4.1 Group Size

Group size plays a vital role in overall performance as it impacts the throughput and energy consumption of WiFi clients in infrastructure mode. As the bandwidth is shared among different users, increasing the number of clients can significantly impact the throughput and energy consumption. In this chapter, first, we analyze the effect of WiFi Direct group size on the energy consumption and throughput of UEs. This is done by limiting the number of UEs per group. It means, we do not allow more UEs to join the group if the limit is reached. Second, we propose to tune the transmit power of the UEs in order to control the group size.

4.4.2 Transmit Power

Tuning transmit power of UEs potentially limits the transmit range and consequently the number of clients per group. To control the transmit power, we utilize the information elements defined in infrastructure mode WiFi as part of the management frames [41]. An information element

consists of three fields; *element ID* number, *length*, and a variable length component. Each *element ID* number is associated with a different attribute that are listed in [41]. The element ID number 32 is associated with the power constraint attribute that is defined in infrastructure mode WiFi to reduce the power consumption of clients. Table 4.1 demonstrates the fields of power constraint attribute. The last field (*i.e., Local Power Constraint*) in the table tells WiFi clients to reduce their transmit power by a certain amount. For instance, if the regulatory maximum power was 10dbm and the value of this attribute is 2 then the client can change its maximum transmit power to 8dbm, and so on.

Table 4.1: Power constraint information element: this attribute in information element is used to tune the transmit power of a client. The first field has a size of 1 byte and contains element ID number, which is 32 in this case. The second field is also 1 byte in size, which represents the length of upcoming bytes associated with this ID. The third field is also 1 byte long and contains the power reduction value in dB scale.

Bytes	1	1	1
	Element ID	Length	Local Power
	32	1	Constraint

A typical representation of the aforementioned scenarios is depicted in Figure 4.1. The UE, referred as an initiator in Figure 4.1, wants to share a file with other UEs in the network. It is possible that all UEs in the network do not lie in the direct range of the initiator. For such scenarios, there can be multiple groups, where some UEs act as gateways between groups (see Figure 4.1). The UEs adopt the standard procedure of group formation and the UEs with the highest intent value become GOs while others join the groups as GMs. The GMs, getting beacons from more than one groups are selected as gateways between them. It is possible that more than one GM receives beacons from two same groups. In this case, selection of gateway is based on the intent values they share with GOs of

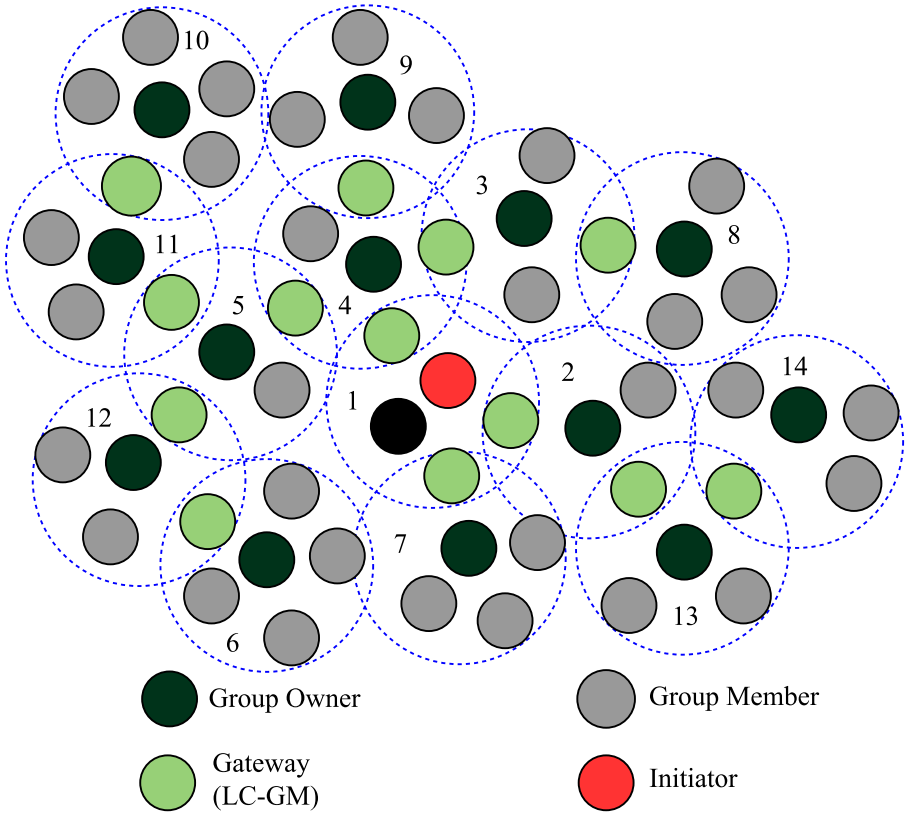


Figure 4.1: The D2D network is divided into multiple groups. WiFi Direct Group Owners (GOs) are the UEs that behave as APs in WiFi Direct, while Group Members (GMs) act as clients. Gateways act as relay nodes between groups. The GOs and gateways are chosen based upon the intent value, they share. The number of groups required to cover the entire network depends upon the size of individual groups.

both groups, to become a gateway. The selected gateway communicates its status to the other associated group to avoid selection of multiple gateways.

One possibility for a UE to act as a gateway is to act as an LC in one group and GM in the other group as found in [36, 38]. This is represented by LC-GM in the chapter. For all other possibilities described in [38], the simultaneous communication of a gateway with two groups is possible either if different wireless interfaces are utilized in two groups [37] or the interface is time-shared among the groups [36]. The UE, which wants to share a file with other UEs in the network, broadcasts the file in its group. Upon receiving this file, the gateway nodes relay it to the other groups they are associated with. In a similar way, the gateway nodes in other groups relay the file further to cover the entire network.

In this work, we consider only one configuration of gateway node that is LC-GM (elaborated in [38]). The number of hops (gateways), required to link a UE with the entire network, depends upon the size of individual groups (number of UEs in a group). It is important to note that, acting as LC-GM, a gateway node can simultaneously receive/send data from multiple groups.

4.5 Performance Analysis

This section presents the simulation results of the proposed techniques. The D2D network (presented in Figure 4.1) is implemented in NS-3 simulator. For simulation, we consider 50 UEs randomly located within a radius of 100 meters. An assumption of 50 UEs is reasonable to cover a wide range of scenarios such as an airplane or a subway or a conference, among many others. In Figure 4.1, the UE indicated by a red circle wants to share a 5 MB file (a typical image size of a 12 megapixels camera) to the other 49 UEs. For such a scenario, we present the performance of the network by

measuring two parameters; (i) total energy consumption of the network in decibels (dB) and (ii) overall throughput of the network in megabits per second (mbps).

To measure the total energy consumption of the network, we consider energy consumed by the UEs during device discovery, group formation, and data transmission phase. For this purpose, we measure energy consumption of individual UEs and add them numerically to have the total energy consumption of the network. Note that by energy consumption of a UE, we are referring to the energy consumption of its WiFi interface only.

To estimate the path loss between UEs, we use the Friis propagation loss model, given by Eq. 4.1, where G_t is transmit antenna gain, G_r is receiving antenna gain, P_t is transmitted signal power, P_r is received signal power, λ is wavelength of transmitted signal, and d is the distance between the transmitter and receiver.

$$P_r = P_t + G_t + G_r + 20\log_{10}(\frac{\lambda}{4\pi d}). \tag{4.1}$$

Table 4.2: Average group size and transmission range: Fixing received signal power, P_r, at -75dBm, the transmission range of a D2D group can be modified by altering the transmitted signal power, P_t. Subsequently, this will change the group size as well.

Sr. No.	**Maximum Range (meters)**	**Transmit Power (dBm)**	**Average Group Size**
1	5	-12	2
2	10	-6	3.1
3	20	0	3.84
4	30	4	5.09
5	40	6	7.14
6	50	8	12.5

For all simulations, G_t is set at 1dB, G_r is set at -10dB, and lambda is set at 0.125 meters (2.4GHz is the operating frequency of WiFi). Note that these values are consistent with a typical WiFi antenna utilized in

smartphones [42]. On substituting these values into Eq. 4.1, we can obtain the WiFi transmit range of a UE, provided that we are aware of P_t and P_r. P_r must be at least -75dBm for a stable wireless connection between two UEs [43]. Hence, reducing transmit power of UEs reduces the transmission range and consequently the number of UEs that can join a group with at least -75dBm received power. In this way, we observe the average group size of the network for different values of transmitted signal power (see Table 4.2). It is important to note that as NS-3 does not directly support WiFi Direct, so to estimate the time spent by UEs in device discovery and group formation phase, we rely on the work presented in [36]. Moreover, within a group, we implement IEEE 802.11g protocol, whose maximum achievable MAC layer throughput for broadcast scenario is 6 Mbps [38].

Figure 4.2 presents total energy consumption of the network for different number of UEs per group. Note that the average group size is calculated by adding individual group sizes and dividing by the total number of groups in the network. The energy consumption is presented for two different scenarios. In the first scenario, we change the group size but let the UEs transmit at normal transmit power, which is 20dBm [43]. In the second scenario, we tune the transmit power of the UEs according to the different transmission ranges. We can observe from Figure 4.2 that tuning the transmit power of UEs provides significant energy gain over the scenario where all UEs transmit with maximum power. For example, an energy gain of 1000% is observed for an average group size of 4 UEs. More importantly, in both scenarios, the energy consumption increases with increasing the group size. From this, one may realize a general perception that the smallest group size must be the best option to opt for. However, in order to choose an optimal group size, we need to examine its effect on the throughput of the network as well, which is demonstrated in Figure 4.3.

Figure 4.3 presents the overall throughput of the network for different

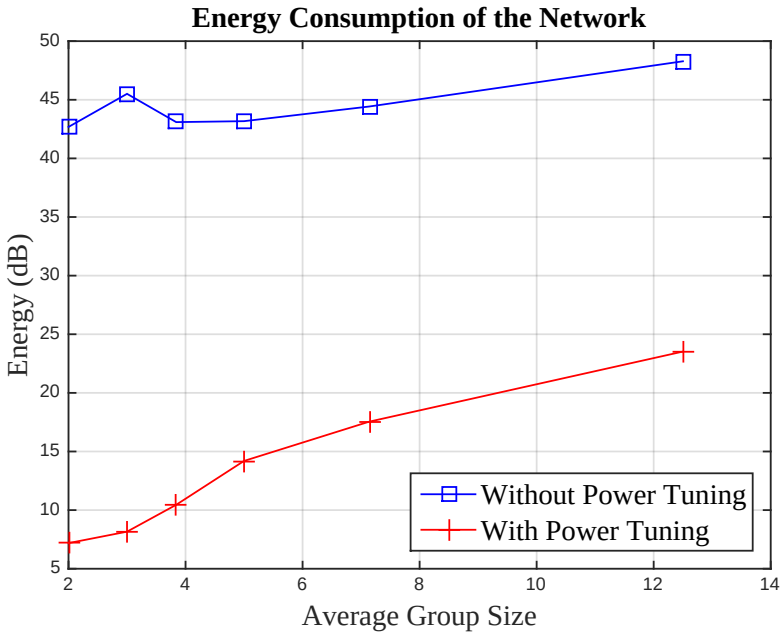


Figure 4.2: Energy consumption of the network with respect to the average group size: The energy consumption of the network increases with the group size. However, significant energy gain can be observed if the transmit power of UEs is tuned properly. For a scenario wherein we do not tune the transmit power but change the group size only, the transmit power is fixed to 20dBm. While, for power tuning scenario, the transmit power values associated with different group sizes can be found in Table 4.2.

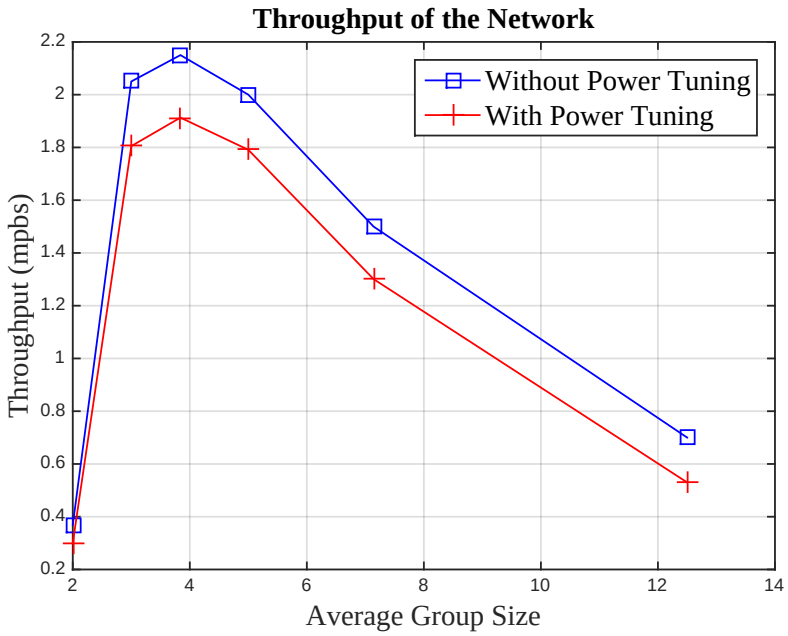


Figure 4.3: Overall throughput of the network: The throughput is generally inversely proportional to the average group size, especially when group size increases from 4 UEs. In addition, smaller group sizes (such as with 2 UEs) decrease the overall throughput of WiFi Direct based multi-hop networks. For "without power tuning" scenario, the transmit power is set to 20dBm, while for power tuning scenario, the transmit power associated with each group size can be found in Table 4.2. Moreover, reducing transmit power of UEs, reduces the overall throughput as well but is insignificant as compared to power gain.

group sizes. To measure the overall throughput, we take the average of individual throughputs of all the UEs in the network. It can be observed from Figure 4.3 that the throughput has a directly proportional relation with the group size at the beginning *i.e.,* the throughput increases with the group size. An average group size of 4 UEs yields a maximum throughput. Thereafter, the throughput begins to drop with the group size. A similar trend is observed for both the scenarios, dependent and independent of the power tuning. However, tuning transmit power potentially decreases the observed throughput. For an average group size of 4 UEs, a throughput reduction of 12% is observed with 1000% gain in energy savings. This is due to the fact that the UEs at the edge of such a group (operating under tuning transmit power) do not observe similar signal strength as those UEs that are in the immediate proximity of a GO. For the same reason, the throughput drop is not significant for smaller group sizes (such as with 2 UEs), where the maximum range is limited to only 5 meters.

It is important to note that the low throughput for group sizes with 2 UEs is due to the fact that simultaneous connections with more than 2 groups are not possible. Therefore, some UEs, after receiving the data, terminate the current group and set-up new groups for further transmission of the information. This termination and setting-up of groups require time that we considerably include for calculating the individual throughputs, as the source has already started delivering the data. On the other hand, larger groups (*i.e.,* with more than 5 UEs) yield low throughput because of the inherently bad performance of WiFi for large number clients connected to a single AP.

Since WiFi protocol works in half-duplex mode and uses Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) to access the shared channel, an increase in the group size increases the time a UE has to wait to access the channel thereby resulting in an increase in energy consumption of

participating UEs and reduction in overall throughput. Moreover, choosing an optimal group size can significantly increase the overall energy efficiency and throughput of such a system.

To summarize, an optimal group size can provide considerable throughput gain. For example, a throughput of 369 kbps, for the case of 2 UEs per group, increases to 2.15 Mbps when network configuration is changed to 4 UEs per group. However, as far as reducing the transmit power of UEs is concerned, there is a trade-off between energy saving and throughput degradation. Reducing transmit power of UEs makes the network more energy efficient on the expense of throughput degradation (Figures 4.2 and 4.3).

4.6 Related Work

Content dissemination and data sharing in Mobile Ad hoc NETworks (MANETs) have actively been investigated in literature. A number of solutions have been proposed, for instance [44–46]. However, very few works have specifically focused on WiFi Direct based networks. One such work is presented in [47] wherein the authors demonstrate the feasibility of WiFi Direct based LTE cooperative video streaming. Specifically, they evaluate the performance of WiFi Direct based group data sharing, including latency, throughput, and power efficiency. However, their work mainly focuses on intra-group data sharing without any emphasis on inter-group data communication. In contrast, the work in [38], instead, analyzes the performance of WiFi Direct in content-centric routing among members of multi-group networks. However, their analysis is restricted to two groups only.

As for power management in WiFi Direct networks, very few solutions have been proposed in the literature. One such investigation appeared

in [39], where the authors primarily consider a single-group WiFi Direct network, sharing access to 3G network. Basically, the authors propose two protocols and demonstrate a comparative analysis with respect to the power saving protocols those are defined in the WiFi Direct standard, as a baseline for energy saving. Another such solution is proposed in [48] that dynamically adjusts the duty cycle of WiFi Direct UEs with respect to the running applications.

Some recent works in the literature investigate the performance of group formations in WiFi Direct based networks. One such work is presented in [49], where the authors present the WiFi Direct group formation methodology for opportunistic networks. The authors, additionally, propose a concept for nominating a backup GO that can potentially replace the original GO in case the group terminates. There are some related works that analyze the standard group formation procedures and their performances [35, 50].

More recently, some works investigate the energy efficiency of multi-hop networks. Concerning this, Ansari *et al.* [51] propose a relay selection scheme, where relay nodes cooperate in form of clusters to increase the networks performance. In particular, they propose random relay selection and Signal-to-Noise Ratio (SNR)-based relay selection schemes to achieve given QoS. Similarly, in [52], the authors propose an analog network coding technique to investigate the multi-hop D2D communication. The authors consider a scenario when a UE acts as a relay between two communicating UEs. They inspect the energy efficiency of this multi-hop D2D communication and compare it with traditional cellular networks and the case when UEs communicate directly without a relay node. The authors in [53] propose D2D relaying as an energy efficient solution for communication recovery in natural disasters. Similarly, the authors in [54, 55] propose energy efficient solutions for content dissemination in D2D communication.

However, to the best of our knowledge, the work in this chapter is a first work that investigates the effects of group size and transmit power of the UEs on energy consumption and throughput of the network.

4.7 Discussion

4.7.1 Potential Applications

The decision, whether to decrease the transmit power of UEs or not, entirely depends upon the application. If the throughput is more important for some applications, the UEs may use the nominal transmit power (20dBm) value to communicate with each other. This includes, but is not limited to, live video streaming applications, where a UE is getting the content from cellular network and is distributing it to others. Similarly, if the throughput is not of much importance in some scenarios, such as in Delay-Tolerant Networks (DTNs) or when a person is sharing a photo of past trip to others, the power tuning scheme may be adopted to save energy.

4.7.2 Security

Apart from power saving, the proposed power-tuning scheme in this work inherently provides security against various attacks. This is due to the following two reasons. Firstly, limiting the transmit range of UEs potentially limits the chances of an eavesdropper to overhear the information if it is not in the range. On the contrary, if the transmission takes place with maximum transmit power, *i.e.,* 20 dBm then the chances of an eavesdropper overhearing the information are much higher. Secondly, limiting the number of UEs per group, *i.e.,* once the group size limit is reached, the GO does not allow more UEs to join, thereby allowing to obtain maximum

energy and throughput gains. This consequently reduces the risk of any fraudulent user joining the group.

We further apply the power-tuning scheme to the PGP based trust mechanism recently proposed in [56], to potentially prevent any dishonest UE from joining the network. Without loss of generality, bootstrapping trust in D2D communication [56] adds an extra level of security to our proposed scheme.

4.7.3 Extending Coverage Area of Cellular Network

It is important to note that herein we consider a use case of content distribution application for delivering contents to the UEs in the proximity. However, the proposed scheme can easily be utilized in many other use cases, such as extending the coverage area of a cellular network, among others. In this particular case, a UE located at the edge can obtain the contents from the cellular network and subsequently distribute those contents to the UEs in its proximity. It is worth mentioning that reputation-based trust bootstrapping mechanism (proposed in [56]) can be employed in such a scenario as the UE at the edge has access to the profiling server.

4.8 Chapter Summary

In this chapter, we explored the possibility of tuning different parameters in WiFi Direct enabled multi-hop D2D networks. In particular, we proposed a power saving scheme that works on choosing optimal group size and transmit power of the UEs to optimize energy efficiency and throughput. Simulation results demonstrate that medium-sized groups (such as with 4 UEs) perform better in multi-hop scenarios. Moreover, transmitting with optimal power provides inherent security against various attacks.

In the next chapter, we analyze a use case of computational offloading

to mobile cloud and see how they can be used to save energy on resource-constrained UEs and in what scenarios, it is better to locally offload computations to peer UEs in the same mobile cloud.

Chapter 5

Computational Offloading to Mobile Clouds

In the previous chapter, we discussed that how can we save energy by adjusting the group size in multi-hop D2D communication. We already analyzed the energy consumption of UEs for single hop D2D communication in Chapter 3 and enlist different use cases and applications of D2D communication, including computational offloading to near by end devices.

In this chapter, we discuss different scenarios of computational offloading for a UE and find the optimal option in terms of its energy consumption. In particular, we compare energy consumption and task completion time of a mobile application for local processing, offloading to the remote cloud and exploiting the cooperation based computing in the local MC. We consider two types of applications in our work; computational intensive and communication intensive applications. We mark an offloading threshold for different offloading scenarios so the UE can decide among offloading to local mobile cloud or to remote cloud, depending upon the size of the task it is offloading.

5.1 Introduction

In recent years, mobile devices and applications are developed rapidly. In 2014, the mobile devices exceeded the Personal Computers (PCs) in terms of Internet usage [57]. Though, the hardware of mobile devices improved considerably in recent years providing higher computational power and more storage space compared to their previous generation, they still fall short to the growing demand of computational power. Additionally, battery industry is not as progressive as semiconductors and telecommunication industries. As a solution to these resource scarceness problems of mobile devices (UEs), Mobile Cloud Computing (MCC) is proposed, which Offloads the computational intensive tasks to remote clouds [58]. However, this offloading can be expensive due to higher latencies between remote cloud and the UE. In addition, offloading to cloud puts burden on the cellular access network, as it exploits radio resources to access the cloud services.

The cellular network, being widely used wireless access technology, provides near ubiquitous coverage but is likely to be overloaded due to increasing mobile traffic [59]. Offloading mobile applications to a remote cloud network can further worsen the situation by injecting more traffic in cellular access network. On the other hand, the WiFi networks provides higher data rates but their connectivity is intermittent.

The need is to have solutions that can potentially solve the resource scarceness problem of UEs without putting burden on the cellular network. To this end, Satyanarayanan *et al.* [26] proposed 'cloudlets' to describe resource-rich computing environment located at the edge of the network and in the proximity of the mobile users. This makes mobile task offloading less expensive in terms of energy and time waste [26]. The idea was then extended to offload the task to nearby UEs [60] to further reduce the

communication cost and latency. Fitzek *et al.* [61] use the term "mobile cloud (MC)" for cooperative arrangement of dynamically connected UEs sharing resources opportunistically. The cloudlets and MCs represent the logical evolution of the concept of moving the distributed cloud more and more towards the user side.

Offloading to nearby UEs uses D2D communication as an enabling technology. D2D communication enables direct interaction between nearby LTE based UEs, minimizing data transmission in the RAN [7], [62]. By doing so, it provides benefits like offloading data from treasured spectrum to D2D UEs, improving spectral efficiency. In addition, exploiting D2D communication incurs less energy costs for communicating with nearby UEs as compared to when UE has to communicate with cloud data center using LTE resources.

We exploit the concept of MCC and benchmark the energy consumption of a UE in offloading data to a local MC or a remote cloud network. We consider three scenarios for the UE to execute the desired task. One is local execution of the task on the UE. Second is offloading the task to nearby UEs acting as an MC. Lastly, the task is offloaded to a remote cloud exploiting LTE cellular network. We consider two types of applications in our work; computational intensive applications and communication intensive applications.

Computational intensive applications are those, which have very less data to be transmitted over a communication link for offloading, but incur high computational cost to UE, if processed locally. This includes, but not limited to, applications related to face recognition and language translation *etc.*. On the other hand, communication intensive applications are those, which have large data to be transferred over a communication link, in case of task offloading. The applications in this category include, but are not limited to, navigation and augmented reality etc.

In this chapter, we mark an offloading threshold to help the UE decide among one of the considered scenarios; (i) local execution, (ii) offloading to a local MC and (iii) offloading to a remote cloud. The threshold is marked based on the execution time of the task and the energy consumption of the UE. In addition, we propose an offloading model to calculate the execution time and energy consumption of these applications for the aforementioned offloading scenarios.

The rest of the chapter is organized as follows. Section 5.2 elaborates the offloading scenarios. Section 5.3 illustrates the communication model, analyzing the cost of the task in terms of the execution time and energy consumption on UE. Results and performance evaluation are discussed in Section 5.4. Section 5.5 discusses the offloading techniques presented in the literature for different offloading scenarios. Finally, Section 5.6 provides the chapter summary.

5.2 Scenario Description

Figure 5.1 explains the MCC scenarios where a UE can process a task locally, offload to nearby UEs using WiFi links or offload to a remote cloud using an Internet connection over LTE links. The source node is a Samsung Galaxy S3 smartphone, which offloads its workload using MCC. Once the offloaded task is completed, the result is sent back to the UE. The smartphone is equipped with both WiFi and LTE interfaces, which it uses for communicating with the local MC and the remote cloud respectively.

There are three different possibilities to compute a given task. The first possibility is to execute it locally on the UE. In this case, only the processing resources of the UE are utilized, as the UE does not communicate with any other UE or remote cloud. In this scenario, if the task has high computational load, it can fully deplete the battery due to very large execution

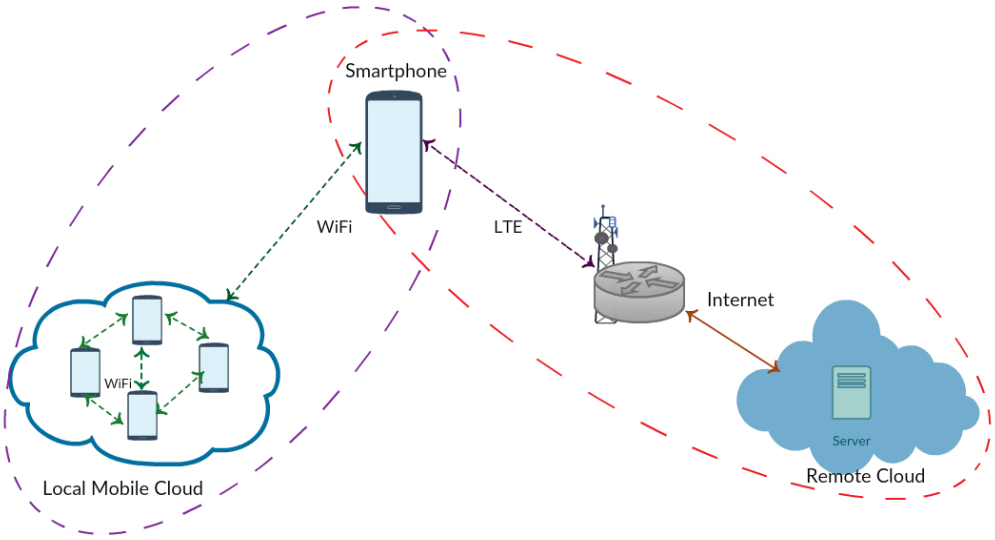


Figure 5.1: MCC scenarios: UE has two possibilities to offload its task. One possibility is to offload the task to a local MC using WiFi interface. Second is to offload it to a remote cloud over the Internet using an LTE interface.

time.

The second possibility is to offload the task to cooperation-based local MC. The task is partitioned into parts equal to the number of UEs in the local MC. For simplicity, we assume that all UEs in the MC are identical. Specifically, we perform simulations on Samsung Galaxy S3 as MC nodes. We assume a star topology where all UEs in the MC are connected to the source node on direct WiFi links. The number of nodes in an MC may vary depending on the availability. In our simulation, we consider only four nodes in the MC. There is no local processing in this case and all execution is performed at the local MC. The energy is consumed in uploading the task to the MC and downloading the computed result.

Finally, the third possibility is to offload the task to a remote cloud using LTE connection, if the local MC is more expensive in terms of execution

time and energy consumption. In this scenario also, we assume no local processing at the offloading UE.

The amount of data offloaded to different infrastructures depends upon the type of application being offloaded. We consider two applications in this work. One application is photo Translator [63], which is computational intensive application with very less data to be transferred through communication links. The other is communication intensive application, Global Mobile Map Viewing and Navigation for Online and Offline OSM Maps [64]. This application has large data to process and puts burden on the communication links of the UE with local and remote cloud.

5.3 System Model

In this section, we model the energy consumption of a UE for two offloading schemes, offloading to a local MC via a WiFi interface and offloading to a remote cloud via an LTE interface. Concerning this, we first need to model the communication cost of WiFi and LTE links. Then, we model the offloading cost of an application via each of those links.

5.3.1 Communication Cost of WiFi and LTE links

To calculate the communication cost of WiFi and LTE links, we consider the models presented in [65] and [66] respectively as the basis.

5.3.1.1 WiFi Energy Consumption

For WiFi transmission to local MC we assume IEEE 802.11g as a communication protocol between the UEs. The WiFi transmission time T_W for N packets can be represented as,

$$T_W = N(T_P + T_{ACK} + SIFS) + B + DIFS, \tag{5.1}$$

Table 5.1: WiFi setup parameters: Below values are consistent with the ones presented in [65].

Symbol	Value	Description
ρ_{idle}	3.68 ± 0.5% W	Idle energy
ρ_{tx}	0.35 ± 8.6% W	Transmission power
ρ_{rx}	0.27 ± 3.7% W	Reception power
λ_r	1000 fps	Rate of received packets
λ_g	1000 fps	Rate of generated packets
γ_{xr}	0.09 ± 8.5% mJ	Energy in the elaboration of received packets
γ_{xg}	0.11 ± 7.6% mJ	Energy in the elaboration of generated packets

where T_P represents the individual packet transmission time, T_{ACK} represents the transmission time for acknowledgments, B is the back-off time required to avoid contention if multiple nodes try to access the channel simultaneously. Short Inter-frame Space ($SIFS$) and Distributed Coordination Function (DCF) Interframe Space ($DIFS$) are inter-frame spacing specified by the IEEE 802.11 standard. As presented in [65], the power consumption P_W during WiFi transmission can be calculated by the following formula,

$$P_W = \rho_{idle} + \rho_{tx}\tau_{tx} + \rho_{rx}\tau_{rx} + \lambda_g\gamma_{xg} + \lambda_r\gamma_{xr}, \tag{5.2}$$

where, ρ_{idle} is the energy consumed by the UE in idle mode, ρ_{tx} is the power required for transmission, ρ_{rx} is the power required for the reception of the data, τ_{tx} and τ_{rx} represent the percentage of utilization of the channel during transmission and reception respectively, γ_{xg} is the energy cost required for the elaboration of a generated packet, λ_g and λ_r are the rates of generated and received packets respectively and γ_{xr} is the energy cost for the elaboration of a received packet. Table 5.1 describes the values of these parameters.

Now the energy consumption of WiFi interface can be calculated by

combining Eq. (5.1) and (5.2).

$$E_W = P_W \times T_W. \tag{5.3}$$

5.3.1.2 LTE Energy Consumption

For LTE transmission to the remote cloud, the time required for uploading and downloading data over an LTE link can be described by the following equation.

$$T_L = T_{PR} + (\frac{D \times 8}{r}), \tag{5.4}$$

where T_{PR} is the promotion time necessary to allocate resources to UE, D is the data size and r is the data rate of an LTE link.

Now, the energy consumption of the UE in using LTE interface can be calculated by the model presented in [66]. In this regard, the power consumption can be formulated by the following expression:

$$P_L = \alpha_u t_u + \alpha_d t_d + \beta, \tag{5.5}$$

where α_u is the power required for bits per second in uplink, α_u is the power required for bits per second in downlink, t_u is the uplink throughput, t_d is the downlink throughput and β is the idle power of the UE. The values of these parameters are presented in Table 5.2.

Now the Eq. (5.4) and (5.5) can be used to calculate the energy consumption of UE in sending D data bits through LTE interface.

$$E_L = \alpha_{PR} T_{PR} + P_L T_L. \tag{5.6}$$

where T_{PR} is the promotional time required to listen the status of the channel in order to transmit or receive. α_{PR} is the promotional power consumed in promotional time.

Table 5.2: LTE setup parameters: Below values are consistent with the ones presented in [66].

Symbol	Value	Description
t_u	15.6 Mbps	Uplink throughput
t_d	32.4 Mbps	Downlink throughput
T_{PR}	275*10^{-3} s	Promotion time
α_u	438.39*10^{-9} W/bps	Power for bps in uplink
α_d	51.97*10^{-9} W/bps	Power for bps in downlink
α_{PR}	1210.07 mW	Promotion Power
β	1288.04 mW	Idle Power

5.3.2 Computational Offloading using WiFi and LTE Links

To model the computational offloading cost, we consider three different cases. One is the local processing on the UE. The second case is to offload to a local MC and third is to offload to a remote cloud using the Internet (see Figure 5.1).

5.3.2.1 Local Processing in the Smartphone

The simplest way to perform the task is to execute it locally on the UE. In this case, the energy consumption depends upon the hardware resources of the UE and cannot be changed. Since the UE has very limited resources, so only the tasks with low or medium size computing can be processed locally on the UE.

5.3.2.2 Offloading to a Local MC (Offloading UE's Perspective)

In our simulations, we consider four sink nodes in local MC having same computational capabilities. To offload computational task to the MC, the data is divided into four equal parts and sent to each UE (sink node). The communication time, $T_{sourceSP}$, is a sum of the times required to send the data to each sink node, T_{sinkSP_i}.

$$T_{sourceSP} = \sum_{i=1}^{n} T_{sinkSP_i}, \tag{5.7}$$

where n is the number of sink nodes, which is 4 in our experiments. The processing time of each sink node is considered as the idle time for the source UE. As sink nodes process the tasks in parallel, the idle time can be calculated as the ratio of processing time of the complete task and the number of sink of nodes in an MC. This is given in Eq. (5.8).

$$T_{idle} = \frac{T_{pro}}{n}, \tag{5.8}$$

where T_{pro} is the total processing time of the complete task. The total energy consumed by the offloading UE can be calculated by adding the energy required in WiFi transmissions and during the idle time. This is given in Eq. (5.9)

$$E_{SP_{lc}} = \sum_{i=1}^{n} (P_{W_i} T_{W_i}) + P_{idle} T_{idle} \tag{5.9}$$

where P_{W_i} is the power consumption of the offloading UE during data transmission to each sink node over WiFi links, given by Eq. (5.2), T_{W_i} is the WiFi transmission time, given by Eq. (5.1) and P_{idle} is the power consumption of the offloading UE during idle time when sink nodes are processing the task.

5.3.2.3 Offloading to a Local MC (Sink-Node' s Perspective)

From the perspective of sink nodes, the communication and processing time of each sink node can be calculated by using following equation:

$$T_{sinkSP_i} = T_{W_i} + T_{pro_i}, \tag{5.10}$$

where T_{W_i} is WiFi communication time of each sink node that can be calculated using Eq. (5.1) and T_{pro_i} is the processing time of individual

sink nodes. Now, the energy consumption of a sink node can be calculated by the following equation:

$$E_{sinkSP_i} = P_{W_i} T_{W_i} + P_{pro_i} T_{pro_i}, \tag{5.11}$$

where P_{W_i} is the power consumption of a sink node during WiFi communication and P_{pro_i} is the power consumption of a sink node in executing the given task.

5.3.2.4 Offloading to a Remote Cloud Server

In this case, the complete task is offloaded to a remote server. There is no local processing and all communication takes place over LTE links. We do not consider the energy consumed by cloud server in executing the task. The time required to send data can be calculated by Eq. (5.4), while the energy consumption can be calculated by the following equation:

$$E_{SP_{rc}} = \alpha_{pr} T_{pr} + P_L T_L. \tag{5.12}$$

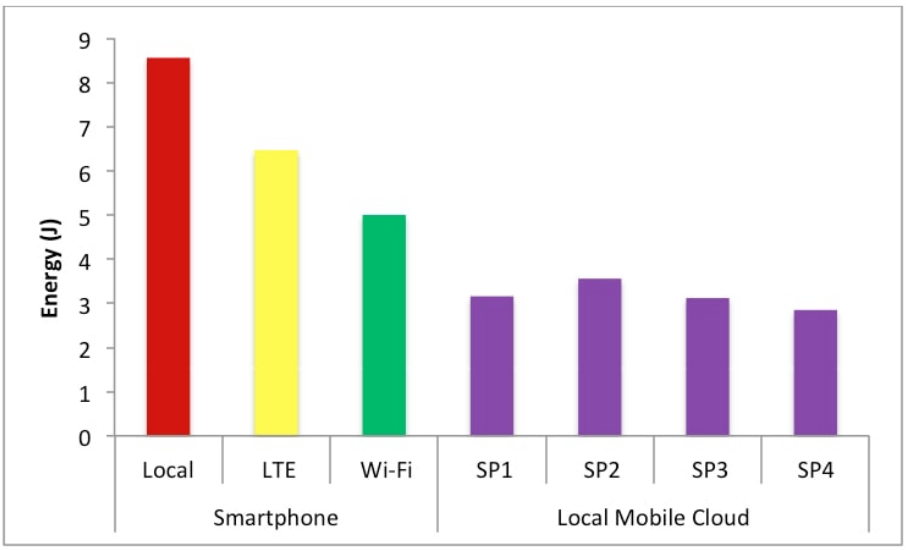
In the next section, we investigate the different offloading schemes and mark an offloading threshold to decide among local MC or remote cloud.

5.4 Performance Analysis

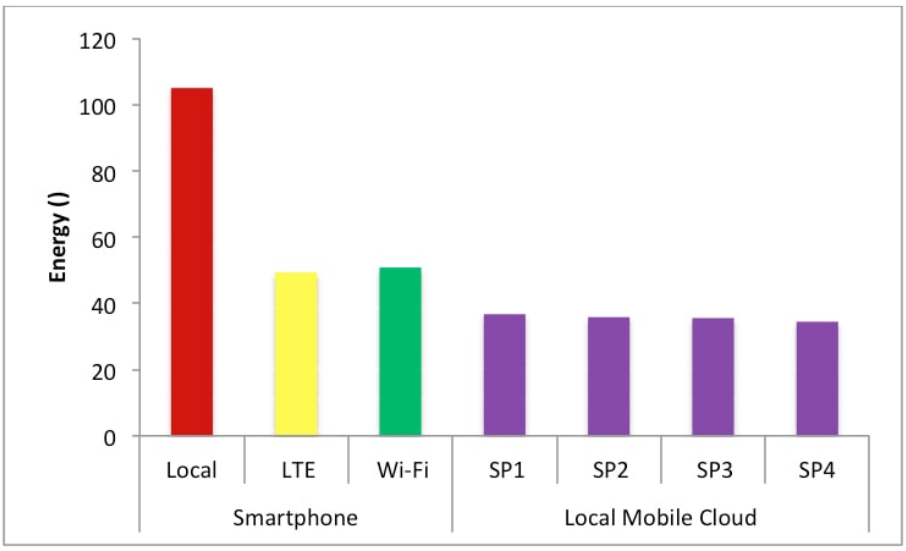
This section presents the performance evaluation of the proposed model and the simulation results.

5.4.1 Assumptions

We assume that all UEs participating in MCC have the same computational capabilities. Moreover, the task being offloaded is equally divisible to the number of participating UEs and each UE will consume same amount of energy in processing that task. It is also assumed that all UEs process their task in parallel.



(a) 2MB data size.



(b) 27MB data size.

Figure 5.2: Energy consumption of source UE: We consider two extreme cases of both applications. Image size is considered as 2MB and road map data size is considered as 27MB. offloading to a local MC is a good option for smaller data sizes while offloading to a remote cloud outperforms in the case of larger data sizes.

5.4.2 System Characteristics

We consider two types of applications in our work. One is Photo Translator [63], which is a computational intensive application that translates the text

of an image in any supported language. It requires a lot of computational resources of a UE, first to process an image and then to translate it. The image size varies from 50 kilobytes to 2 megabytes in our experiments. The second application is Global Mobile Map Viewing and Navigation for Online and Offline OSM Maps [64], which is a communication intensive app with a large amount of data to process. The data size, we consider in our experiments, is based on the road map of different provinces of Italy. For instance, we consider 3 provinces, *i.e.,* Sicily, Abruzzo and Trentino Alto Adige. Their data sizes for the road maps only are 5.4 Megabytes, 17 Megabytes and 27 Megabytes, respectively.

We have two options for offloading both kinds of applications. One is offloading to a local MC over a WiFi interface and the other is offloading to a remote cloud over an LTE interface. In case of remote cloud, the full image and the complete map is offloaded. After processing the offloaded task, the remote cloud sends the result back to the source UE. In case of offloading to local MC, the task is divided into four equal parts and sent to sink nodes of the local MC. The sink nodes process the task and send the result back to the source node. For example, The Trentino Alto Adige has an area of 13,607 km^2 [67], which results in 27MB data size for the road map. Consequently, we send 6.75MB to each sink node. Moreover, we neglect the additional energy overhead in dividing the task.

For simulations, we use NS-3 [68] network simulator extended with LTE functionality from LENA project. The UE we used in our experiments is Samsung Galaxy S3 equipped with a quad core Exynos 4412 processor with maximum clock frequency of 1.4 GHz. The UE contains Cortex-A9 architecture that is able to execute 2.5 DMIPS (Dhrystone Million Instructions Per Second)/MHz per core and has a maximum computational power of 14000 DMIPS [69]. For all simulations, the power consumption of S3 is taken as 1.5W in working mode and 666mW in idle mode. Similarly,

the power consumption over WiFi communication is set to 1264mW and over LTE communication is 1543mW. All power consumption values are consistent with the ones described in [70]. We do not consider the power consumption of the remote cloud. However, in order to calculate the processing time of the remote cloud, we consider its computational capacity as an Intel Core i7 3770K processor able to process a maximum of 106926 DMIPS at 3.9 GHz [71].

The data rate for WiFi communication ranges from 734Kbps to 24Mbps. While for LTE, the data rate ranges from 0.924Mbps to 15.6Mbps in uplink and 2.24Mbps to 32.4Mbps in downlink. The distance between UEs of local MC ranges from 0.5m to 2m. In case of remote cloud, the distance varies from 50m to 200m from LTE base station. Depending upon the image size, the transmitted data ranges from 50KB to 2MB in case of computational intensive application, while for communication intensive application, data size varies from 5MB to 27MB, depending upon the size of the map.

5.4.3 Results

Figure 5.2 demonstrates the results from the analytical model presented in Section 5.3. The energy consumption of the source smartphone in processing the task locally is compared with both offloading schemes. For this comparison, we consider an image size of 2MB for photo translator and data size of 27MB for OSM Maps applications. It can be observed from Figure 5.2a and Figure 5.2b that local processing always consumes more energy for the applications we considered. In the case of computational intensive application, the offloading to a local MC over WiFi links is more energy efficient than offloading to a remote cloud over LTE links, as shown in Figure 5.2a. On the other hand, larger data sizes like 27MB incur more energy costs on WiFi links as compared to LTE, as shown in Figure 5.2b. However, energy consumption of the UEs in the local MC is not same,

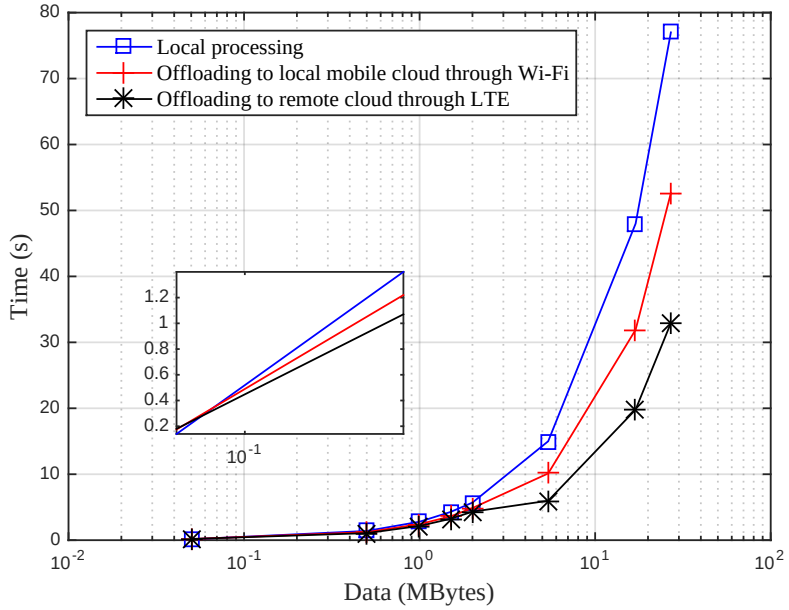
despite of the fact that same amount of data is offloaded to them. This is due to the difference in the distances of the UEs from the source node.

The simulation results are presented in Figure 5.3, which demonstrates the time and energy consumption of the source UE in the cases of local execution of the task and both offloading scenarios. The figure provides an overview of offloading threshold for process completion time and energy consumption of the source UE, which helps the source UE to decide among local processing, offloading to local MC or remote cloud. It can be observed from Figure 5.3a that for small data sizes (<50KB), the local execution time of the UE is better than any offloading scheme. For medium data sizes, 50KB to 2MB in our scenario, the task completion time remains the same independent of the offloading scenario. While for data sizes larger than 2MB, offloading to a remote cloud takes lesser time to complete the task. This is due to the limited resources of UEs in the local MC. The offloading threshold for energy consumption of the UE is around 4MB as shown in Figure 5.3b. More precisely, offloading to a local MC is more energy efficient when offloaded data size is less than 4MB. After 4MB offloading to a remote cloud begins to consume lesser energy.

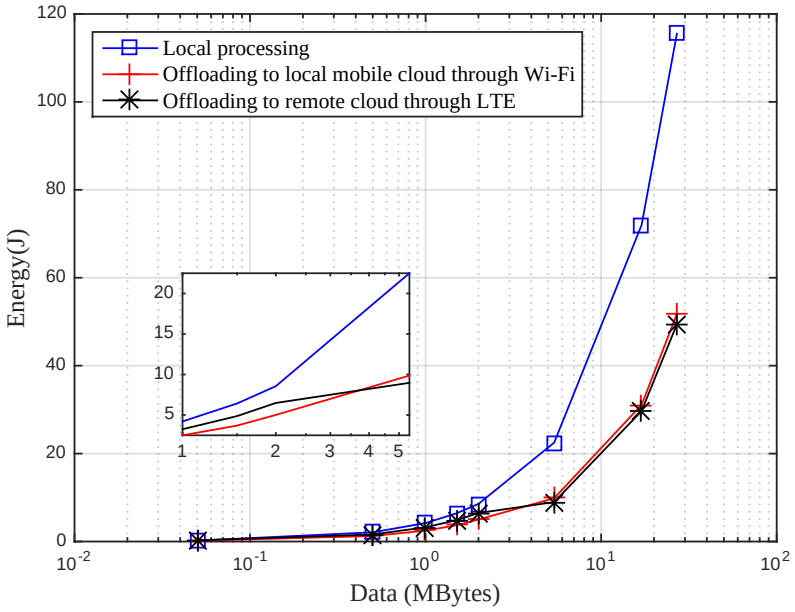
Figure 5.4 provides the percentages of offloading gain over local processing. The offloading gains are provided for task completion time and energy consumption of the UE. It can be observed from Figure 5.4a that the data sizes less than 2MB provide higher gains when they are offloaded to local MC. At around 2MB, both offloading schemes provide same gain over local processing for task completion time, which is 17%. After 2MB, offloading to a remote cloud begins performing better, providing offloading gain as high as 88%. Similar trend of offloading gain is observed for energy consumption, as shown in Figure 5.4b. The offloading threshold here is 4MB with 110% energy gain. For data sizes less than 4MB, offloading to a local MC gives higher energy gains as compared to the remote cloud. While for

data sizes greater than 4MB, offloading to a remote cloud outperforms the local MC in terms of energy consumption of the source UE. The maximum energy gain can be as high as 158%. More precisely, If data size of offloading task is greater than 4MB, it is more time and energy efficient to offload the task to a remote cloud using an LTE link.

Figure 5.5 illustrates the effect of the number of sink nodes in local MC on task completion time and energy consumption of the source UE. It can be observed from the figure that for data sizes less than 5MB, the time and energy remains almost constant, independent of the number of the sink nodes in the local MC. While for higher data sizes, an increase in the number of sink nodes in local MC makes offloading more efficient in terms of task completion time and energy consumption. However, after a certain threshold, further increase in the number of sink nodes results a slight increase in the task execution time and energy consumption of the source UE. This upper bound for the number of sink nodes in local MC is due to the fact that having too much sink nodes increases the communication and task slicing overhead, which contributes to more energy consumption and task execution time.

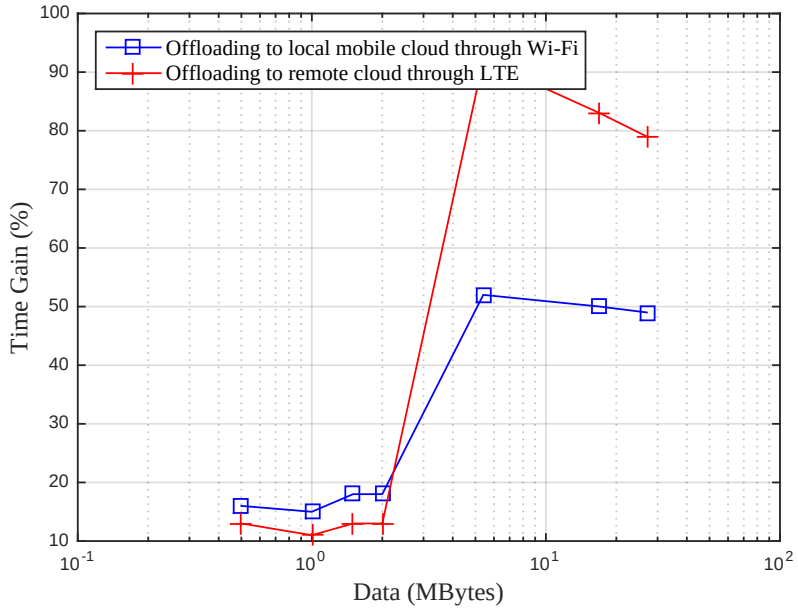


(a) Processing and communication time.

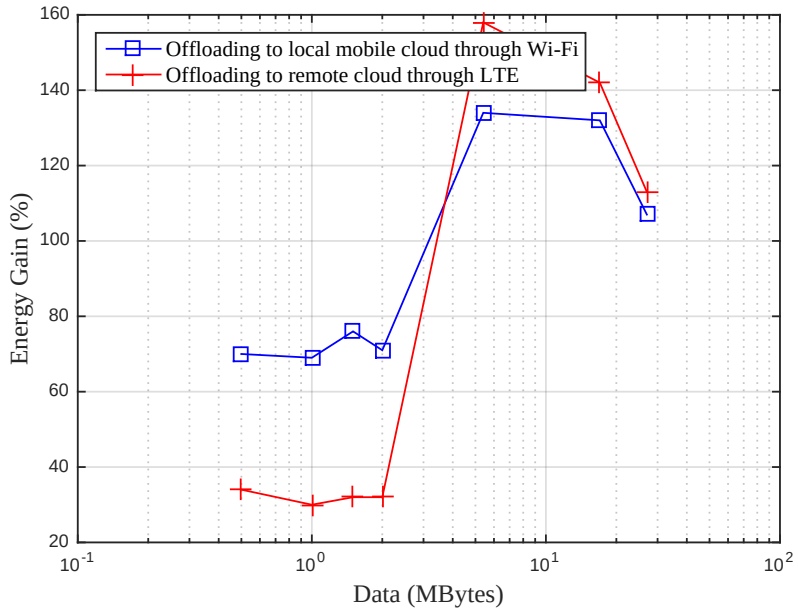


(b) Energy consumption.

Figure 5.3: Simulation results for task completion time and energy consumption of the source UE: Smaller data sizes (<50KB) are better to be processed locally on the UE, medium sized data (50KB to 4MB) can be offloaded to a local MC and larger data sizes (>4MB) can be offloaded to a remote cloud.

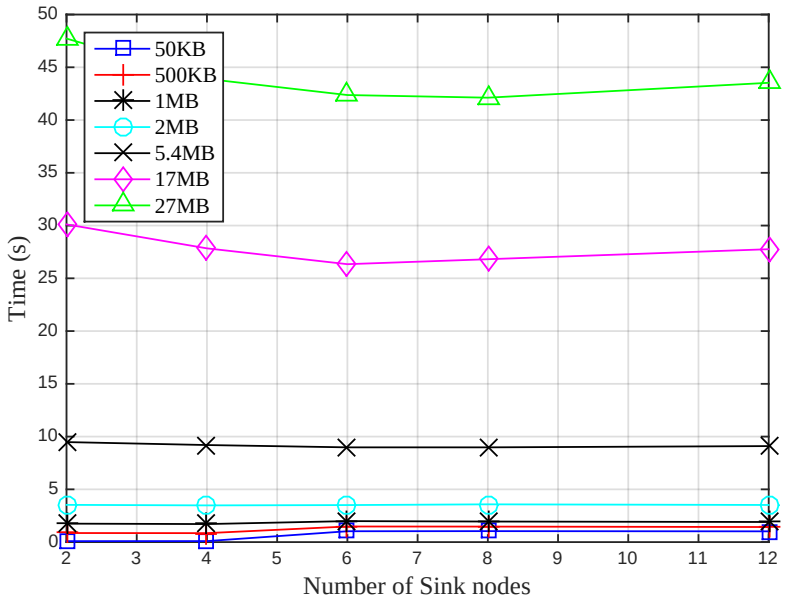


(a) Processing and communication time.

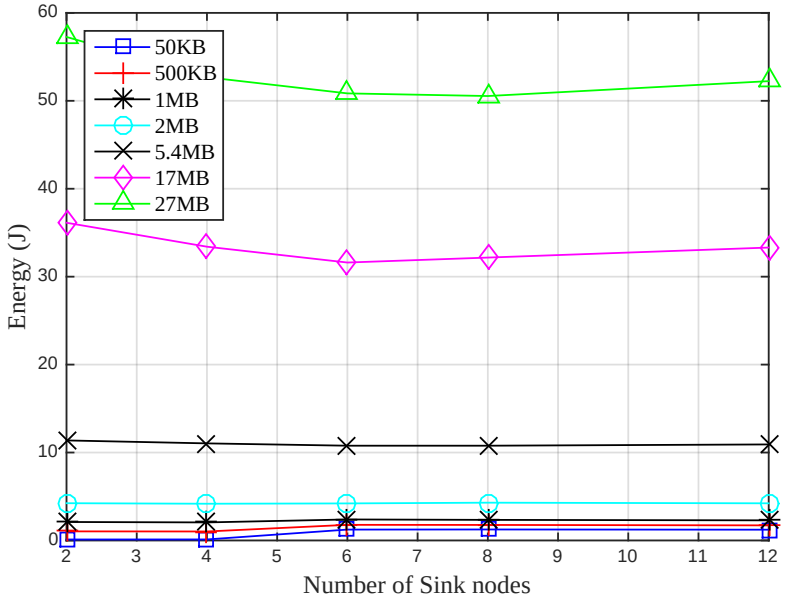


(b) Energy consumption.

Figure 5.4: Percentage gain in task completion time and energy consumption of source UE: For smaller data sizes, offloading to a local MC gives higher gains in time and energy efficiency. On the other hand, for larger data sizes, offloading to a remote cloud provides higher gains.



(a) Processing and communication time.



(b) Energy consumption.

Figure 5.5: The effect of the number of sink nodes on task completion time and energy consumption of source UE: Both time and energy are independent of the number of sink nodes for smaller data sizes. An increase in the number of sink nodes decreases the task completion time and energy consumption for larger data sizes. Further increase in the number of sink nodes can result in slightly higher energy consumption and task completion time.

5.5 Related Work

The scarceness of mobile resources (*e.g.,* processing power, storage size and battery time) and scarce bandwidth of cellular networks are the motivations to search for alternate solutions to handle the dramatic increase in mobile applications in various categories. The problem of resource scarceness of UEs is addressed through MCC that allows UEs to partition their storage demanding and computationally intensive tasks and offload them to a remote cloud with enormous computational and storage resources. However, this creates a problem of radio resource scarceness in cellular access network. In order to solve this problem, other solutions are proposed in MCC such as offloading to local MCs or cloudlets. Various offloading techniques are introduced in literature for each kind of offloading scenario. We briefly summarize these techniques in this section to compare with our approach.

5.5.1 Offloading to a Remote Cloud

Different offloading techniques based on client-server communication, Virtual-Machine (VM) migration and code partitioning are proposed in literature.

Deboosere *et al.* [72] propose a grid model for offloading task to a remote cloud. The UE is connected with the server via a thin client protocol such as Virtual Network Computing (VNC). The work focuses on the selection of an appropriate server for offloading task. Especially in case of user mobility, the task may be migrated to a nearby server to minimize the delay from server.

Kemp *et al.* [73] propose *Cuckoo* framework that uses Java stub to offload mobile task to any resource rich environment that runs Java virtual machine. To use cuckoo, the applications need to be re-written to support local and remote execution. The authors report the gain in execution time

by a factor of 60 and reduction in battery consumption by a factor of 40.

5.5.2 Offloading to Cloudlets

Cloudlets are resource rich surrogate machines, known as cyber foraging [26], near to mobile user. Solutions like CloneCloud [74] and MAUI [75] can be potentially utilized to offload task to cloudlets.

Chun *et al.* [74] propose *CloneCloud* to migrate a part of the application to a resource rich server using VM migration. This incurs an extra cost of VM migration apart from task execution at the cloud. The authors compare this cost with local execution at the UE for multiple applications of various categories. The execution time is speeded-up by a factor of 21.2 when UE is connected using WiFi connection. There is no need to rewrite the code to offload the task.

MAUI [75] uses a combination of code partitioning and VM migration to make an offload decision to different infrastructures depending on their Round Trip Time (RTT). Longer and shorter RTTs impact differently on UE's energy consumption when offloading computations [76]. The impact of shorter RTTs of cloudlets on power consumption of UE is further studied in [26].

Offloading to MCs: Several solutions are proposed in literature for utilizing peer-to-peer communication model in MCC [60, 77–83]. These works present solutions and architectures to make the MCs possible.

Serendipity [60] is among the pioneer works that focuses on task allocation among UEs. The authors emulate their scenario and test the system for possible speedups and conserved energy. However, the authors did not test their system for multiple applications of different workload. For applications with different data and computational requirements will behave differently in terms of energy consumption and time savings.

Cirrus [82] is an extension of Serendipity, where a UE can not only

offload to other UEs but also to computers installed on moving vehicles or placed in a nearby building. In this work also, the authors did not consider multiple application of varying workload in terms of data and computations.

Most the works in literature in MCC focus only on the techniques to offload data either to a remote cloud server or nearby cloudlets or UEs. On the other hand, in this chapter, we analyze different offloading scenarios for energy and execution time savings. Depending on the workload, data or computation, the UE can decide whether it is worth offloading to a remote cloud or to a local MC.

The work presented in [83] is somehow related to what we are presenting in this chapter. However, the authors focus on the task execution time only and does not focus on the energy consumption of the UE. Additionally, the primary focus of their research is to present a mathematical model to investigate the cloud size, cloud nodes' lifetime and reachable time. On the other hand, in this work, a part from presenting a mathematical model for communication of UE with remote cloud and local MC, we simulate our scenario using NS-3 simulator.

5.6 Chapter Summary

In this chapter, we characterize different offloading schemes in MCC. We present a mathematical model for these offloading schemes and mark an offloading threshold for task completion time and energy consumption of the source UE. The task completion time includes the communication time with cloud and execution time of the task at the cloud. The idea is to offload the task in such a way that both task completion time and energy consumption can be minimized. The results from analytical model are consistent with NS-3 simulations. We conclude that smaller tasks, such as

<50KB, are better to be executed locally in the smartphone. The medium sized tasks, such as 50KB to 4MB performs better if they are offloaded to local MC. For higher data sizes, such as >4MB, it is better to utilize the resource of remote cloud instead of using local MC as an offloading option.

After focusing on the energy efficiency of UEs in single and multi hop D2D networks and presenting a computational offloading scenario, we decided to focus on preserving the user's privacy. In this regard, we start by establishing trust between UEs, participating in various applications. To this end, the next part of this dissertation first proposes a solution to bootstrap trust in D2D networks and then present a solution to preserving privacy. Specifically, we consider a content distribution application wherein content can be cached at D2D UEs. Later, we extend our findings to preserve privacy in other caching possibilities in 5G networks.

Part B

Privacy in D2D Communication

Chapter 6

Bootstrapping Trust in D2D Communication

In the previous chapters, we focus on energy efficiency of UEs involved in D2D communication for various scenarios. However, security in D2D communication, which is equally essential for the success of D2D communication in future networks, is imperative for the success of D2D communication. In particular, bootstrapping trust between D2D UEs remains at the core of secure communications.

This chapter proposes a combination of the PGP and reputation-based model to bootstrap trust in D2D environments. Our proposal aims at minimizing any suspicious connection with selfish users. We show that although trust establishment between UEs adds overhead to D2D communication but offloading cellular traffic to trusted D2D links still provides significant throughput gain over the conventional cellular network. Our results show that the capacity gain can be as high as 133%.

6.1 Introduction

D2D communication is an emerging paradigm in cellular networks that enables direct interaction between nearby UEs, minimizing data transmis-

sion in the RAN. The basic idea is, first proposed by Lin and Hsu in 2000 [84], to enable multi-hop relays in cellular networks. Since then, D2D communication has been investigated for its potential applications in P2P communication, multimedia content distribution, social networking, gaming, group multicast, IoT, public safety, and cellular traffic offloading [85]. All these proximity-based applications share a lion's portion of cellular traffic. This provides an opportunity to offload this traffic to D2D links. By doing so, users get various advantages such as, lower transfer delays, higher data rates, and better energy efficiency [32]. These potential benefits along with the growing number of proximity-based applications led to the standardization of D2D communication over last few years.

Despite all potential benefits, D2D communication faces a serious security threat. For example, in a smart home environment, a malicious user can pretend to be a smart terminal, to which all smart devices are connected in D2D mode and potentially take the control of these smart appliances. Similarly, a user performing proximity-based social interactions can be potentially connected with a malicious user, who in turn can take the control of user's UE and steal personal information. This requires an efficient mechanism to potentially check the security and social status of the connecting UE before establishing any D2D connection. The problem can potentially be solved by leveraging PGP and reputation-based mechanisms.

Reputation management is an effective tool that can be utilized to facilitate decision-making in D2D communications. Reputation can be defined as the opinion of one user or UE about the other. More specifically, it can be considered as the trustworthiness of a user. In other words, reputation can be seen as the expectation that a user will behave in a particular way. For instance, if a user has a reputation for not getting jobs done or sharing inappropriate/incorrect content, then other users will avoid connecting

with such a user [86].

In this chapter, we present a mechanism that builds on top of PGP and reputation models. The main idea is to profile reputation information about D2D users. This reputation information is consulted before considering a D2D user for exchanging any content. Before exchange of content, we also have to authenticate D2D users. To this end, we propose flexible PGP policies that can authenticate users without degrading usability.

The rest of the chapter is organized as follows. Section 6.2 provides a brief overview and the trust problem in D2D. Section 6.3 presents the design overview including the system model and the key idea. Section 6.4 presents solution details. Section 6.5 reports on performance analysis. Related work is reviewed in Section 6.6. A discussion has been provided in Section 6.7. Finally, we draw some conclusions and highlight the chapter summary in Section 6.8.

6.2 Overview and Problem

D2D communication enables bundles of smart applications in future 5G networks. These applications include, but are not limited to, proximity-based social networking and gaming, local advertisements, and multimedia content distribution. Currently, in all these applications, the application's traffic takes a path through the cellular network even if the users are in the physical proximity of each other. This causes a burden on cellular access network that is already facing a resource scarceness problem [7]. This issue can be addressed by offloading cellular traffic of proximity-based applications to D2D links.

Traffic offloading is considered as a potential solution to solve the capacity bottleneck problem of cellular access network [87]. For instance, in content distribution applications, the UEs in physical proximity accessing

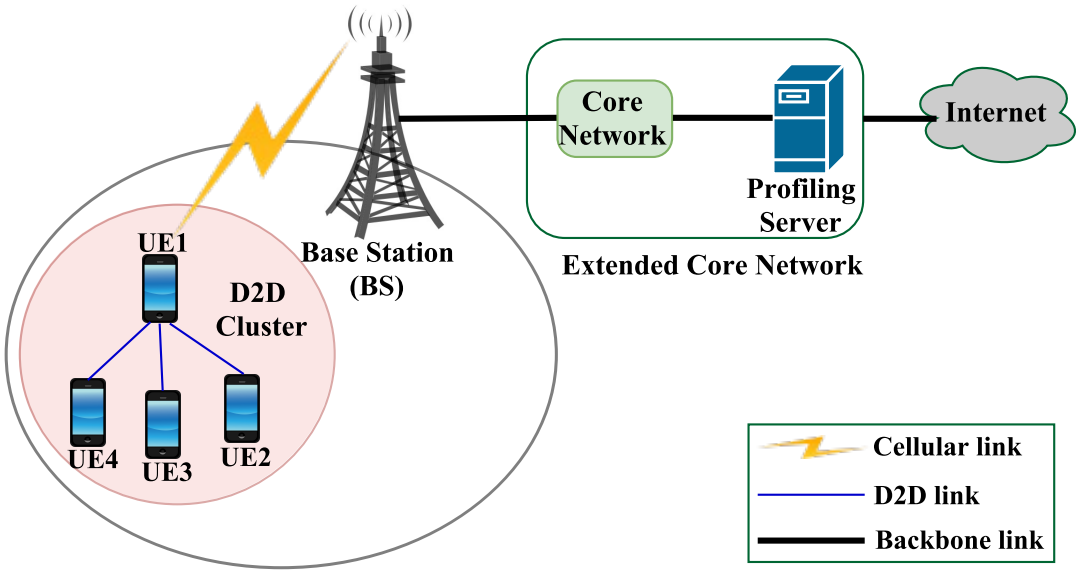


Figure 6.1: An overview of the architecture for offloading cellular traffic using D2D networks: A cluster head (*i.e.,* UE1) in D2D cluster interacts with the Base Station (BS) to download the content from the Internet through an extended core network comprised of the core network and a profiling server. Cluster members (*i.e.,* UE2, UE3, and UE4) can download the requested content, if available, from the cluster head (*i.e.,* UE1).

the same content from web form a D2D cluster. In this cluster, one UE gets content from the content server and distributes it to the other UEs in the cluster (see Figure 6.1). In this manner, a significant portion of cellular traffic could be offloaded to D2D networks, which would have traversed through access and core networks otherwise.

To offload cellular traffic, the network operator must have a prior knowledge about the UE's traffic. If UEs in physical proximity are accessing the same content, the network operator can distribute the contents by establishing D2D links between UEs to save bandwidth. From a security point of view in D2D networks, a certain level of protection can be achieved via encryption. However, in proximity-based applications, trust is a significant problem. More specifically, the question is: *How a user will trust that the UE with whom it is going to establish a D2D link is a benign user or not?* The problem here is not the secure connection with the UE but the user itself. The user can establish a social link with a malicious user. In order to solve this problem, there should be a mechanism to bootstrap trust in D2D communication. In this chapter, we propose a solution based on PGP and reputation-based mechanisms to bootstrap trust in D2D networks. The details of the proposal are provided in the next sections.

6.3 Design Overview

6.3.1 System Model

We first identify system entities, assumptions, potential adversaries, and possible attacks.

6.3.1.1 Entities

In our system, there are the following entities (interactions are shown in Figure 6.1 and explained in Section 6.3.2):

- **UEs:** UEs are the core entity of our system. As we can see in Figure 6.1, there are a number of UEs. UE1 is regarded as the cluster head, which is responsible for distributing the intended content to other UEs. Rest of the UEs (*i.e.,* UE2, UE3, and UE4) in the D2D cluster are regarded as cluster members.

- **D2D Cluster:** The group of UEs, which are interested in sharing the same service, *i.e.,* cellular traffic offloading, is named as a D2D cluster in our scenario. A D2D cluster consists of two types of UEs: a cluster head and cluster members.

- **Network Operator:** The entities in the network operator can be further partitioned into following sub-entities:

  - *Base Station:* It is an entity in the access network of the network operator that is directly connected to UEs over cellular links. In the proposed topology, the cluster head, *i.e.,* UE1 in the D2D cluster is connected to the BS via a cellular link and other UEs (*i.e.,* UE2, UE3, and UE4) in the cluster are connected to UE1 via D2D links.

  - *Core Network:* It is the typical core network of a cellular network, such as Evolved Packet Core (EPC).

  - *Profiling Server:* It is the new entity that we introduce. It can be the same global SDN controller that we introduced in Chapter 3. It is responsible for profiling D2D users. It keeps track of reputation information of D2D users based on which the trust can be bootstrapped. We combine core network and profiling manager to form an extended core network.

6.3.1.2 Adversaries and Attacks

We assume that there are adversaries in the D2D environment who would like to eavesdrop and modify the communication traffic. Second, we consider that network operators are honest-but-curious, *i.e.,* they work according to the specified protocol but they are curious to learn about the content passing through them. Network operators are trusted to manage reputation information about the users. We assume there is a secure channel between users and the content provider. To establish the trust, there could be a set of Certificate Authorities (CAs), which are trusted by D2D users.

6.3.2 Key Idea

6.3.2.1 Reputation-based D2D Communication

We propose a secure D2D communication system relying on a reputation-based mechanism wherein the UEs are able to securely connect to each other based on certain measures. These measures (*a.k.a.* reputation information) define the reputation of the UE being connected onto. To store reputation information, we introduce a new component named *Profiling Server*. The UEs in such a system have access to these reputations, which allows them to establish a reliable data connection with other UEs. There are different methods, as will be described in the upcoming section, that can be utilized to build these reputations over time. The longer the communication system will be alive, the reputation accumulated will keep on becoming more and more reliable ultimately making the D2D connectivity much more secure over the time.

After each D2D link, the UEs provide their ratings about each other. These ratings are stored in profiling server in terms of reputations. For instance, in Figure 6.1, UE2, UE3, and UE4 can rate UE1, if they get the

right content from UE1. These ratings could be provided automatically or manually. For instance, UEs can automatically send reputation information based on characteristics of the connection, such as transfer time, successful transfer, and interaction time. The manual reputation information is subject to the nature of the content. For instance, if the requested contents are provided, a user can give the UE a high rating.

It is important to note that any UE can serve in the cluster head role. That is, UEs can serve in both a cluster head and a cluster member roles simultaneously.

6.3.2.2 Certificate-based Authentication

In order to perform the authentication, communicating parties in D2D communication can share a pre-shared secret. However, we argue that this could limit the potential of D2D by excluding a number of situations wherein communicating UEs cannot set up a pre-shared secret. To dynamically bootstrap trust in D2D settings, we present a mechanism wherein users can choose a policy that assists in minimizing user intervention, which otherwise is expected in PGP-based solutions.

Before establishing any D2D connection, the UEs verify each other with the help of PGP and reputations in profiling server. After the authentication, UEs can establish a secure channel (such as SSL), which can guarantee both confidentiality and integrity.

6.4 Solution Details

In Figure 6.2, we illustrate workflow details of our proposal. Basically, we assume that the cluster head UE1 has already downloaded some content from the Internet. Next, we consider that a cluster member UE2 is looking for that content. Of course, UE2 can directly download the content

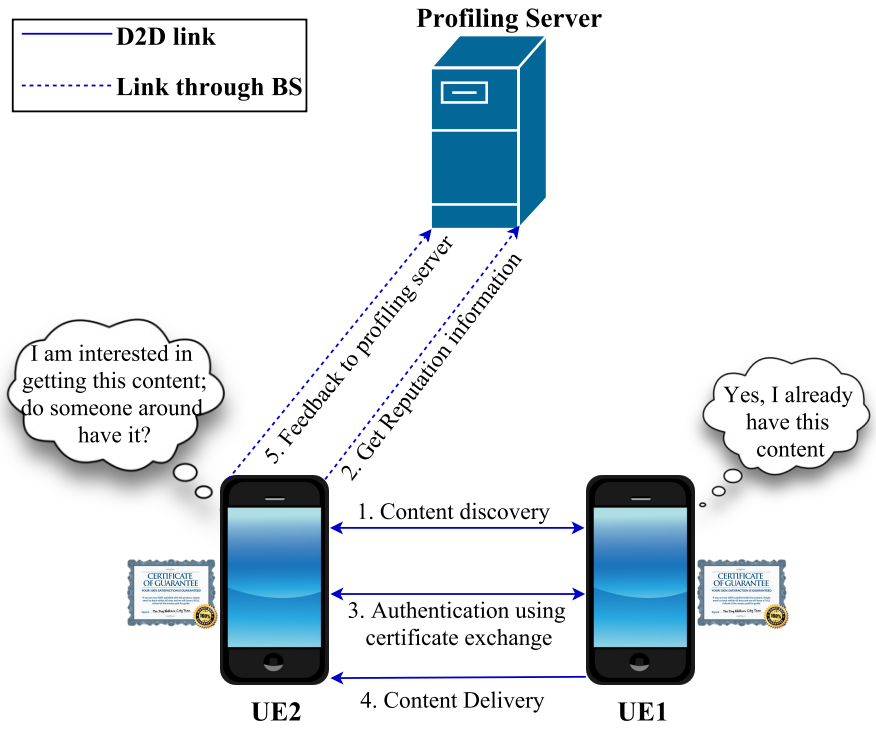


Figure 6.2: Workflow details: UE2 is looking for a specific content and disseminates the request within the D2D network (Step 1). We consider that UE1 has already downloaded that content. UE2 collects reputation information about UE1 from the profiling server (Step 2) based on which UE1 can perform authentication by exchanging digital certificates (Step 3). Next, the requested content is delivered from UE1 to UE2 (Step 4). Finally, UE2 sends some feedback about UE1 to the profiling server.

from the Internet. However, we argue that it will overload the underlying core network. In order to exploit D2D networks, UE2 will first consult its neighbors (Step 1). Since both UE1 and UE2 are part of the same D2D network, UE1 will be discovered. Before getting the content from UE1, UE2 wants to make sure that UE1 has good repute. To do so, UE2 consults with the profiling server (Step 2). The profiling server returns the reputation information (as discussed in Section 6.4.1) about UE1. Based on that reputation information, UE2 can decide whether to download the content from UE1 or not. In case the decision is *yes*, UE2 performs (Step 3) authentication using certificate exchange (as discussed in Section 6.4.2). Next, using a secure communication channel established during the authenticate phase, UE1 exchanges the requested content with UE2 (Step 4). After the exchange is completed, UE2 provides some feedback about UE1 to the profiling server (Step 5).

6.4.1 Reputation-Based D2D Communications

Reputations can be gathered based on one or combination of the following methods [86]:

- *Calculus-based Reputation:* Each user or her UE develops reputation in a calculative manner. To build-up a calculus-based reputation, each UE rationally calculates the costs and benefits of other UE's cooperating or cheating in their respective transactions.
- *Knowledge-based Reputation:* Each UE develops reputation via accumulating knowledge about other UEs either first-hand (say based on self-interaction) or second-hand, which could be based on the understanding of what, why, where, when, and how other UEs behave.
- *Institution-based Reputation:* Each UE believes the other UEs to be safe to be connected based on sociology that deals with a trustworthy

environment. For instance, one can be in a university environment, a conference or similar settings.

Reputation-based D2D communication systems can be one of the best solutions for dealing with selfish behavior. They can be very robust in curtailing insider attacks as well [86]. On the other hand, a challenge in building such a reputation-based model in a D2D system is how to avoid malicious behavior of UEs such as providing false feedback about other UEs [88].

6.4.2 Certificate-based Authentication

Before exchange of any personal information, UEs in the D2D environment might be interested in knowing each other. Technically, there must be an authentication prior to exchanging any content. In order to perform authentication, one solution is to rely on digital certificates. The question is must the UEs trust these certificates or not. Based on this trust, we can divide certificates into two main groups: the Public Key Infrastructure (PKI) model and the PGP model. In a PKI model, UEs can obtain a certificate from a CA. Each certificate includes a certificate chain, building a chain of certificates up to the root CA, which is a trust anchor in the PKI model. Typically, there are three types of digital certificates: a leaf, an intermediate, and a root, where the latter two types are part of the certificate chain. Without loss of generality, there could be a set of intermediate certificates in the certificate chain.

In the PKI model, a user can verify the certificate if s/he also trusts the same trust anchor as present in the certificate chain. More specifically, to verify a certificate, a UE must have installed the corresponding root CA. This PKI model is the one that is widely utilized between any two communicating parties over the Internet. However, the major problem

with this model is the certification cost that parties have to pay to CAs.

The PGP model is an alternative to the PKI model. It does not require any certification from CAs. Instead, public keys are shared out of band. Unlike the PKI model, there is typically no certificate chain in PGP certificates.

We can leverage both the PKI model and the PGP model to authenticate UEs in the D2D environment. After authentication, a secure channel can be established based on some session keys. These session keys can ensure both confidentiality and integrity of the exchanged data.

Applying the PKI model in D2D settings is quite straightforward. However, the challenge is how UEs can exchange their certificates out of band. To deal with this issue, a UE can manually fingerprint SSIDs of trusted UEs. One can argue that this can limit the potential benefits of the D2D network. To leverage the D2D network in a seamless manner, we propose using some flexible policies. Using these policies, a UE can indicate that s/he can trust PGP certificates based on some temporal properties. These temporal properties can include constraints based on time and location. For instance, a policy can state that *trust all PGP certificates in the next two hours* or *trust all PGP certificates at current location*. In general, we can provide UEs with a set of template policies, which could be based on temporal properties including time and location.

6.5 Performance Analysis

In this section, we present the throughput gain, where UEs establish trusted D2D links based on reputation information and PGP. Our simulations are mainly based on the system presented in [89]. We consider a Frequency Division Duplex (FDD)-based LTE system as a baseline for comparison.

6.5.1 System Parameters

In our simulations, we consider the transmission bandwidth of 9MHz, both in uplink (UL) and downlink (DL). This corresponds to 50 Resource Blocks (RB), as one RB in LTE is 180 KHz. The UL is under-utilized in FDD based cellular system, so D2D resources are shared with UL as indicated in [90]. The path loss for an LTE link can be modeled by Eq. 6.1, where d is the distance between the transmitter and the receiver and F_C is the carrier frequency [91].

$$\mathrm{PL}_{dB}(d, F_C) = 36.7 \log_{10}(d) + 22.7 + 26 \log_{10}(F_C). \tag{6.1}$$

The path loss for D2D links can be modeled by Eq. 6.2, where λ is the carrier wavelength, d_{BP} is the breakpoint distance, h_{tx} is the transmitter antenna height, and h_{rx} is the receiver antenna height [92].

$$\mathrm{PL}_{dB}(d, \lambda, D(d)) = 20 \log_{10} \left(\frac{e^{0.002d}\, 4\,\pi\, d\, D(d)}{\lambda} \right), \tag{6.2}$$

where,

$$D(d) = \left\{ \begin{array}{ll} 1 & d \leq d_{BP} \\ \frac{d}{d_{BP}} & d > d_{BP} \end{array} \right\}, d_{BP} = \frac{4\,(h_{tx} - 1)\,(h_{rx} - 1)}{\lambda}. \tag{6.3}$$

We consider a Multiple-Input Multiple-Output (MIMO) system with a configuration of 2 antennas for UEs and 4 antennas for Base Station (BS). For this system, the link throughput can be calculated using Eq. 6.4, where B is channel bandwidth, m is the number of spatial layers, λ_i is the channel matrix, ρ is 2dB penalty used for practical implementations, I_0 is the noise and interference power, and P_i is the transmit power allocated per spatial layer [93].

$$C = B \sum_{i=1}^{m} \log_2 \left(1 + \frac{P_i\, (\lambda_i)^2}{\rho\, I_0} \right). \tag{6.4}$$

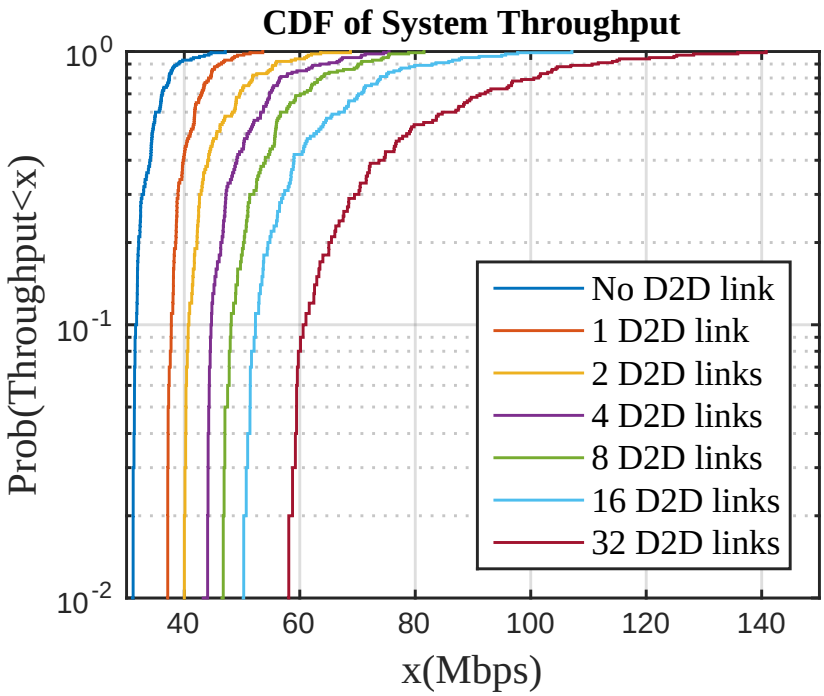


Figure 6.3: System throughput with varying number of trusted D2D links: We compare throughput with FDD-based LTE system having a transmission bandwidth of 9 MHz. The MIMO configuration of our system includes 2 antennas for UEs and 4 antennas for the BS. The D2D links are established based on the authentication mechanism presented in Figure 6.2.

6.5.2 Simulation Results

We consider 33 UEs in our system for different simulation scenarios. Figure 6.3 presents the Cumulative Distribution Function (CDF) of the system throughput. Before establishing any D2D connection, the UEs are examined for both PGP and reputation-based models, as discussed in Section 6.4. The first case is when there is no trustworthy UE to establish a D2D link and all UEs are accessing the content through cellular infrastructure. In this case, 40.05Mbps is the maximum achievable throughput between UEs and BS. In all other cases, we gradually increase the number of trustworthy UEs in our system from 1 to all 33. In the last case, only one UE is connected to the BS through an LTE link, rest all 32 UEs are getting

the contents from that particular UE. In this case, a throughput gain of 140Mbps is achieved, as compared to 45Mbps when no UE is utilizing the D2D link.

Table 6.1: A summary of percentage throughput gain at different percentiles of CDF: The throughput gain is directly proportional to the number of trusted D2D links. We get maximum throughput gain of 168% when all 32 UEs are getting the content from UE1.

Percentile	**1 D2D link**	**2 D2D links**	**4 D2D links**	**8 D2D links**	**16 D2D links**	**32 D2D links**
10	19.58%	31.56%	39.96%	56.96%	66.97%	98.73%
50	17.83%	30.71%	49.56%	63.26%	82.97%	131.05%
90	17.14%	38.90%	64.52%	87.20%	114.07%	168.45%
Average	18.18%	33.72%	51.35%	69.14%	88.01%	132.74%

Table 6.1 summarizes the percentage of throughput improvements at various percentiles of the CDF. The substantial gain in the throughput can be observed for different number of trusted D2D links, ranging from 1 to 32. There is a direct relation between the number of D2D links and the percentage throughput gain. The mean throughput gain increases to 133% using 32 D2D links.

It is important to note that these throughput gains are observed only when UEs are trustworthy to establish D2D connections. If the reputation of a UE is not appropriate and also PGP does not allow a UE to establish a D2D connection, there will be no benefit of a UE being in proximity and accessing the same content, as the system will not allow establishing D2D link between them.

6.6 Related Work

6.6.1 Authentication using Certificates

In a PGP system [94], users bootstrap the trust by manually exchanging their public keys. PGP follows a model known as *web of trust*. In [95], Capkun *et al.* proposed a solution, which is based on PGP [94], for MANETs. Basically, they presented a self-organized public key management system that allows users to generate their public-private key pairs, issues certificates, and performs authentication without requiring any centralized server. However, the major issue with their idea and all PGP-based solutions is to require user involvement in the certificate management operations including issuance and revocation of certificates.

For bootstrapping trust in MANETs, there are also some approaches based on a trusted dealer, such as [96]. In [97], Rachedi and Benslimane presented a similar model, where each cluster is supervised by a cluster head. This is the cluster head that serves as a CA. The major issue with these mechanisms is that they delegate trust from a centralized CA to each local D2D cluster.

There are schemes based on threshold-based cryptography, such as [98, 99]. The trust in these schemes is bootstrapped by contacting neighbors. The main limitation of these schemes is involvement of a certain number of neighbors in the trust bootstrapping.

6.6.2 Reputation-based D2D Communications

Reputation-based systems have made a significant presence over the past couple of decades. For instance, Srinivasan *et al.* in [86] have discussed various methods for implementation of reputation-based ad hoc network systems. In [88], Xiong and Liu have utilized community-based reputations to assist in estimating the trustworthiness between various peers. They

have further discussed various parameters, describing how each peer will influence ultimately in order to establish a reputation-based P2P system. Another scheme was proposed, named Secure and Objective Reputation-based Incentive (SORI), in [100] to establish a reputation-based ad hoc network system via encouraging packet forwarding while having control over the selfish behavior of participating nodes. Interestingly enough, not too long ago, a cognitive radio system was presented in [101] wherein the authors utilized a reputation-based mechanism to identify misbehaviors and mitigate their harmful effects on sensing performance. Hence, it is encouraging to consider reputation-based mechanism as a way to go for modern world secure D2D communications.

6.7 Discussion

6.7.1 Resilience against Sybil Attacks

Both PGP and reputation-based mechanisms have been studied in isolation. However, in this work, we present an approach that aims at dynamically bootstrapping the trust by leveraging PGP and contextual information. Note that any solution based on PGP could be vulnerable to Sybil attacks. The novelty of our approach lies in complementing our flexible PGP-based mechanism with a reputation-based mechanism. This combination naturally minimizes the possibility of Sybil attacks [102] in D2D networks.

6.7.2 Levels of Trust

To choose a certain level of trust, users can be provided with a slide bar that can show options from low to high. The level can be taken into account while making a decision based on reputation score and the potential

mechanism to authenticate prior to exchanging any content in D2D networks.

The level of trust can potentially classify the UEs between highly reputed and least reputed. Depending on the type of application, the user can decide whether to connect with a UE or not. For instance, different proximity-based applications such as social networking, gaming, multimedia content distribution, and public safety may have varying security requirements. The public safety UEs must authenticate only highly reputed users, while user playing online games may be connected with comparatively less reputed UEs.

It is important to note that we consider a user-case of content distribution application for assigning reputation to the UEs. However, this reputation can also be utilized to identify suspicious UEs for any kind of applications.

6.7.3 Content Discovery in D2D

There are two main points concerning content discovery in D2D networks. First, a UE can discover other UEs that are 1-hop away, but in this way we would not be exploiting D2D networks at its full potential. To epidemically disseminate the discovery request within the D2D network, we can consider a mechanism as adopted in opportunistic networks (such as Haggle [103]), thus allowing multi-hop discovery. Second, the discovery phase could lead to serious privacy concerns. To address that, we can consider some privacy-preserving approaches in opportunistic networks, such as [104].

6.8 Chapter Summary

The chapter presents an efficient mechanism to bootstrap trust in D2D networks. We leverage the combination of PGP and reputation-based models

to authenticate D2D users. Incorporating such security features in UEs helps to avoid connections with selfish users. This opens new doors to securely utilize D2D networks in growing proximity-based applications, such as social gaming and P2P social networking. In these scenarios, offloading traffic to D2D networks can provide significant throughput gains over the baseline LTE system, when UEs communicate over cellular links. The simulation results show a mean throughput gain of 133% for 32 D2D links over a 9MHz frequency band.

In the next chapter, we focus on preserving privacy of the UEs, once a trusted link is established with peer UEs. We consider content distribution application where UEs cache a copy of the required content and deliver it to the source UE. Later, we extend our scenario to other caching options such as, cloudlet, IoT gateway, 5G network or a Content Distribution Network (CDN).

Chapter 7

Secure Caching in D2D Networks

In the previous chapter, we discuss the possibility of integrating trust in D2D applications. However, after established communication links with trusted users, the content should be also be secured from possible vulnerabilities, more specifically when they are cached at D2D network or any other location other than the content provider.

In this chapter, we propose a marketplace for providing a number of caching options for a broad range of applications. In addition, we propose a security scheme to secure the caching contents with a simultaneous potential of reducing the duplicate contents from the caching server by dividing a file into smaller chunks. We model different caching scenarios in NS-3 and present the performance evaluation of our proposal in terms of latency and throughput gains for various chunk sizes.

7.1 Introduction

5G cellular networks are taking shape in serving a variety of applications with diverse QoS requirements. These applications range from high data rate video streaming to low data rate IoT communication [105]. Although IoT applications usually generate low data rate traffic, however, the accumulation of data from billions of devices can potentially put burden on the

backhaul network. Concerning this, caching at the edge of a network can potentially solve this problem.

Caching at the network edge to reduce the delivery time of the content is already proposed in literature. CDNs are one approach, which are widely adopted across the world to reduce the network latency in delivering web content. On the other hand, fog computing or edge computing is an approach that introduces resource-rich computation nodes near the end users. These nodes are mainly designed to provide fast computations at the edge of a network but can also be employed as caching servers, especially in wireless environments [106, 107]. D2D communication is another possible place for contents to be cached in wireless environments [8, 56]. Other possibilities include Small cell Base Station (SBS) caching and Macro cell Base Station (MBS) caching [106].

In this chapter, we discuss various caching possibilities in future 5G networks for a broad range of applications, including IoT and normal cellular users. To achieve this, we present a marketplace for 5G service providers and content providers, where 5G service providers can offer caching servers while content providers can use those caching servers to store the content. In addition to providing security of the cached contents, we also consider the possibility of reducing the Internet traffic by uniting the duplicate entries in the caching server. In this regard, we employ convergent encryption, a scheme that encrypts data with its own hash code, to manage the duplicate contents. Duplicate contents are difficult to manage by other end-to-end encryption schemes, such as blind cache [108], designed for caching environments. In those schemes, the encryption of the data being cached restricts the caching server to identify the multiple entries of the same content.

The main contributions of our work are listed as follows.

- We propose different caching possibilities for end users and IoT ap-

plications, apart from those mentioned in [106]. To achieve this, we present a marketplace for both 5G service providers and content providers for offering caching servers and caching contents, respectively.

- We employ convergent encryption as an end-to-end encryption scheme for aforementioned caching environments, which works to manage the duplicate contents as well, saving not only storage at the caching server but also the bandwidth by reducing the number of requests for the same content between caching server and the content provider.
- For handling data duplication, we propose to split the content into smaller chunks so that a part of the content can also be prevented from duplicate storage/requests.
- We measure the latency and throughput for all caching environments by varying the chunk size of the content and taking into account the encryption and decryption of each chunk.

The rest of the chapter is organized as follows. Section 7.2 provides a problem statement to present a gap in the literature. Section 7.3 demonstrates the design overview and the key idea presented in this work. Section 7.4 reports performance analysis. Related work is reviewed in Section 7.5 followed by a discussion in Section 7.6. Finally, the chapter summary is presented in Section 7.7.

7.2 Problem Statement

The motivation behind caching contents near end users is to reduce network latency, as many applications require fast processing and access to their stored data. This holds true for many IoT applications wherein billions of devices connected to the Internet are generating low rate data of

measurements that many end users or applications request frequently. For instance, there are many scenarios wherein a large number of end users run applications that request similar IoT data, such as weather condition and monitoring, among many others. Caching these contents near end users or applications not only reduces the network latency but also reduces the load on the Internet by reducing the number of requests traversing through IoT cloud service providers.

On one hand, caching contents results in network bandwidth, throughput, and latency gains; on the other hand, caching contents on untrusted servers can raise serious privacy and security concerns. For instance, CDNs are abundantly utilized to improve the delivery of the contents by replicating them to the caching servers located geographically near to the end users. However, current CDN technology requires user contents and traffic to be exposed to CDN providers [108], thus compromising user's privacy and security. To address this, a secure solution for caching contents is required, which not only works for IoT applications but also for a range of applications and users using the Internet. End-to-end encryption, such as HTTPS, *i.e.,* Secure Sockets Layer (SSL) or Transport Layer Security (TLS) used with Hyper Text Transfer Protocol (HTTP), is thought as one of the solutions to secure the access of data across the Internet. However, when it comes to caching, the end-to-end encryption restricts the inline transparent caching by the network service provider to serve the previously requested content. Since SSL/TLS works between endpoints, *i.e.,* client and server, aiming to mitigate man-in-the-middle attack, it becomes challenging to secure data when it is stored in a middle server.

Out-of-band cache *a.k.a.* blind cache is one solution proposed by the industry [108] to overcome this problem. Blind cache is an encryption scheme, which shares the key of the encrypted data with the client and places encrypted data to the CDN server (see Figure 7.1). This allows

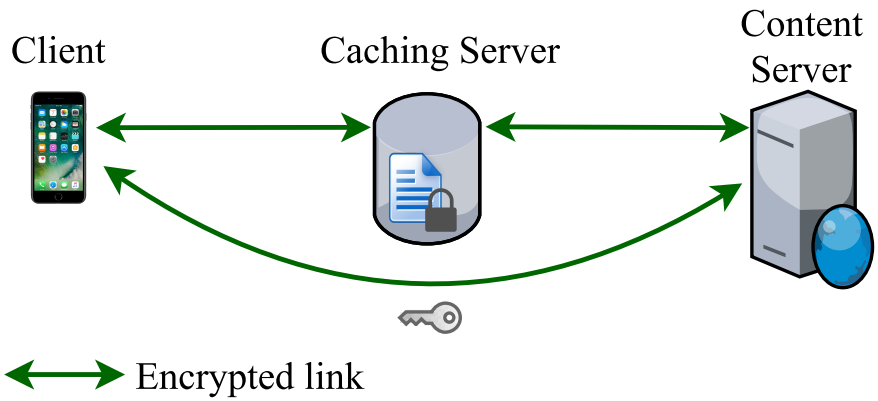


Figure 7.1: Blind Cache: The key is shared to the client on a direct encrypted link between the content server and client. The encrypted data is shared with the caching server and a link to the caching server is sent to the client.

content providers to share the encryption key of the cached content directly with the client while content is accessible through the CDN network. The problem with this solution is the overlapping contents. As CDN is not exposed to the encrypted content so it can possibly save multiple copies of the same content accessed by different users. Moreover, the blind cache solution is proposed for CDNs, which is not only expensive in terms of budget but also requires a prior contract with a CDN provider.

7.3 Design Overview

In this chapter, we present a model wherein any intermediate 5G node, capable of storing the contents, can cache it. This, in turn, boosts the marketplace, where any node between client and server can act as a CDN. This section highlights the design overview of the proposed solution. Figure 7.2 presents potential options to cache the contents near the end users. As we can see, we can cache at the user's premises, 5G cellular networks or CDNs. The entities presented in Figure 7.2 are defined as follows.

- **D2D Nodes:** D2D nodes are smart devices, such as smartphones, in

the vicinity of end users that can be utilized to temporarily cache the contents.

- **Cloudlet:** Cloudlet represents the resource-rich computing resources in the vicinity of the end users [26]. It can be a Local Area Network (LAN) or even a single computer capable of caching the content for end users.

- **IoT Gateway:** An IoT gateway acts as a bridge between the IoT devices and the cloud to transport the information to and fro the cloud. It can be utilized as a potential location for caching contents.

- **5G Intermediate Nodes:** This represents any node in 5G networks, capable of storing data and providing it to the end users. The location of this node can be in the access network with the base station or in the core network. For experimental analysis, we consider this node in the access network.

- **CDN:** CDN is a well-known technology employed to reduce the network latency by caching the contents on the server located at the network edge, close to end-users.

- **Content Provider:** This represents the actual content provider. For instance, it can be an IoT cloud server providing information regarding vacant parking spaces around the city or a server streaming live videos to end users or a server providing weather information after collecting the same from various different IoT sensors.

Once a content provider receives a certain number of requests from the same region for the same content, it caches the content in one of the aforementioned servers, located as close to the end user as possible. To this end, the content can even be cached within the user's premises (see

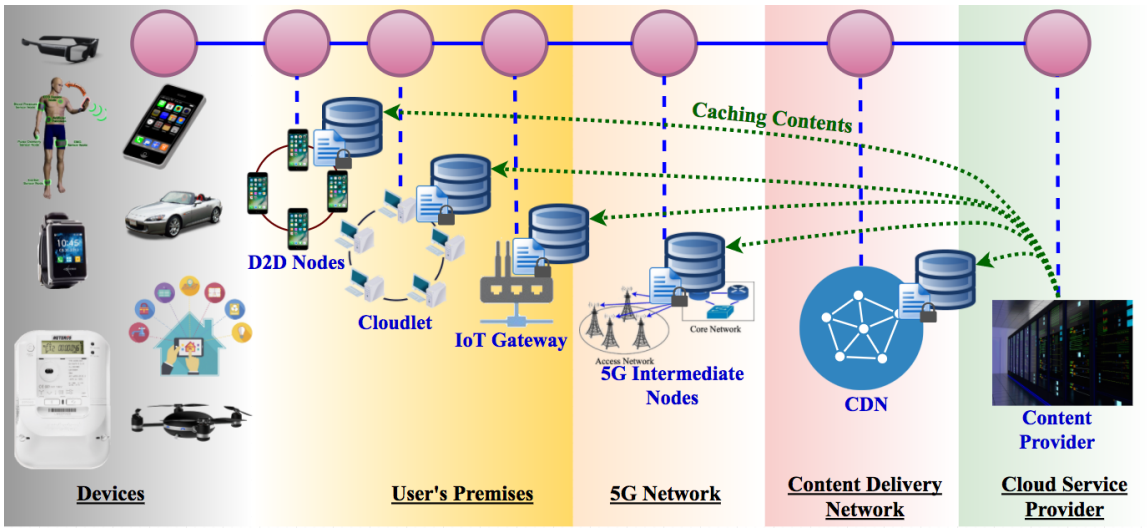


Figure 7.2: Different caching options for a content provider: The data can be cached in any appropriate server between the content provider and clients. The choice of caching server depends upon the requirement of the application or end user, requesting the content.

Figure 7.2). Before caching the content, the content provider divides the content into smaller chunks. The motivation behind dividing the content into smaller chunks is the careful management of overlapping contents. There are many scenarios wherein some users request a content, while others request a part of it. For instance, some users in a region request for a complete book, while others in the same region request for specific chapters. If the content is already cached in the form of chunks, where each chunk corresponds to an individual chapter, then the request for separate chapters can easily be handled by the caching servers. The content provider just needs to communicate the unique identifier of the respective chunk (chapter in this case) to the client. Similar instances of overlapping contents include, but are not limited to, a movie season and its episodes, an album of images and separate images, and results from a search engine for similar queries, *etc.* Dividing a file into smaller chunks not only saves bandwidth in the Wide Area Network (WAN), but also saves storage at the caching servers.

The workflow of our solution is as follows. Upon receiving a request from a client for a particular content, the content provider establishes a secure channel with the client by creating a session key and checks if the content is already present in the caching server. If the content is not already cached but requested frequently, the content provider partitions the content into smaller chunks and calculates a hash code for each chunk. Then, each chunk is encrypted with its own hash code. Technically, we employ convergent encryption [109]. After this, the content is sent to the caching server, *e.g.,* a 5G intermediate node, to be accessed by the client. Then, the content provider sends a list to the client, containing all the hash codes, the encrypted chunk IDs, and the web address of the caching server. In case the content is already cached, the content provider simply shares the aforementioned list with the client through the secure channel protected via a session key.

Henceforth, in this chapter, we evaluate the performance of each caching option in terms of latency and throughput. Moreover, we provide an efficient solution to secure the cached contents to be exposed to the caching server, which leads to preserving privacy of not only end users but also content providers. The partition of the content into smaller chunks solves the problem of caching multiple copies of overlapping content and convergent encryption makes our solution lightweight thereby making it best suitable for a wide range of scenarios, such as IoT applications and much more.

7.4 Performance Analysis

In this section, we analyze performance of our proposed caching scheme in terms of latency and throughput gains. We used the NS-3 network simulator for simulating the proposed model presented in Section 7.3. To cover the broad range of file sizes, we consider three distinct files types including

text, image, and video. For simulations, we consider client as a smartphone requesting aforementioned data files from a content provider. We analyze all the caching options presented in Figure 7.2 to evaluate latency and throughput gains, considering various chunk sizes for each file type. While estimating throughput, we also include the overhead of partitioning the file into smaller chunks inclusive of encrypting and decrypting each chunk of data.

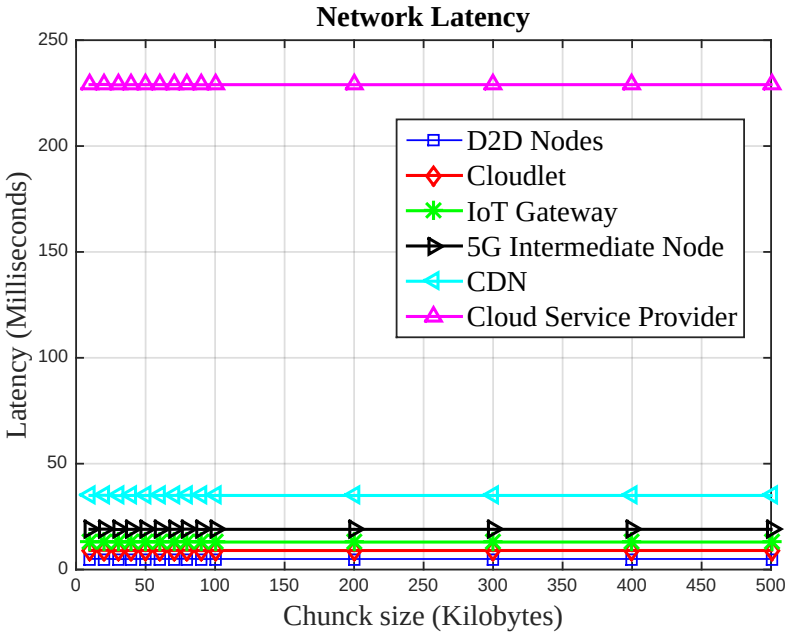


Figure 7.3: Latency of various caching servers: The closer the server is to the client, the lesser is the latency.

Figure 7.3 demonstrates the network latency from the client to each caching server. It can be noticed, the nearer the caching server, the lesser time it takes to serve the client. Moreover, the latency remains independent of the chunk size, which appears true as latency is the property of the network and the nodes involved. It has no relation to the size of the data being transferred between the server and the client. The D2D nodes provide the best latency. However, these resource-constrained caching servers (D2D nodes) cannot store much information nor can they store it for longer du-

rations. Besides this, the mobility factor of D2D nodes makes them lesser suitable for adoption as caching servers. However, there are many scenarios wherein D2D nodes can be exploited as potential caching servers. For instance, remote health care applications, wherein the body sensors or implants do not generate huge amount of data, can exploit smartphone of the concerned doctor to cache the sensors' information of multiple patients a priori, instead of having to establish a one-to-one connection between the doctor and the patients and/or the cloud service provider that is responsible for storing IoT data from various patients.

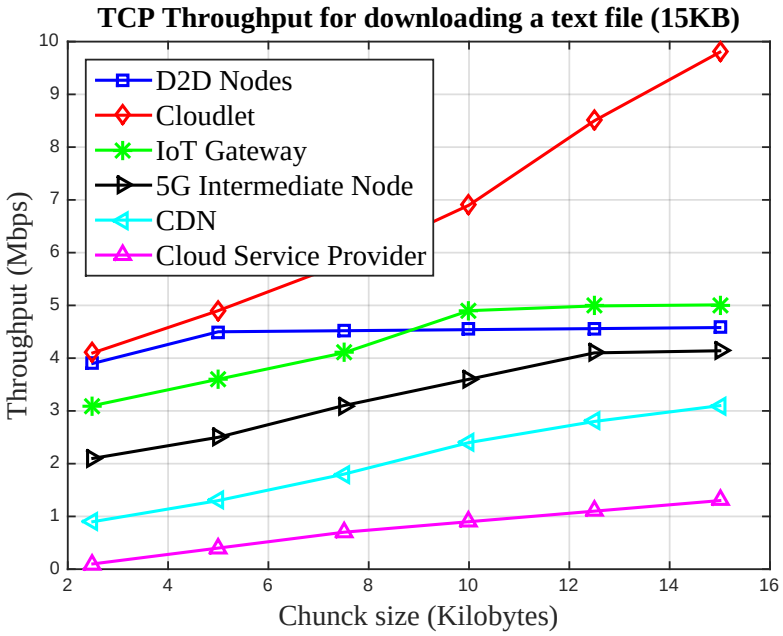


Figure 7.4: Throughput of downloading a text file from different caching servers for varying chunk sizes: The throughput is always less than the optimal when the chunk size is less than the BDP of the link. Moreover, the throughput is, generally, directly proportional to the chunk size, *i.e.,* higher the chunk size, better is the throughput.

Next, we analyze the throughput of the link between the client and each caching server for all aforementioned file types. With throughput, we refer to the Transport Control Protocol (TCP) throughout between source and destination. For simulations, the standard size of a three-page text file is

assumed as 15KB [110], the image is considered as 5MB [111], and the video file size as 50MB for a 10 minutes video of 360p resolution [112]. Figure 7.4 represents the TCP throughput of accessing a text file from different caching servers. The throughput varies with the chunk size the file is actually divided into. It is evident from the Figure 7.4 that the throughput is directly proportional to the chunk size. This is quite expected as increasing the chunk size decreases the number of chunks the file is actually divided into, which consequently decreases the combined overhead of encrypting and decrypting all chunks. Subsequently, an increase in the throughput is observed.

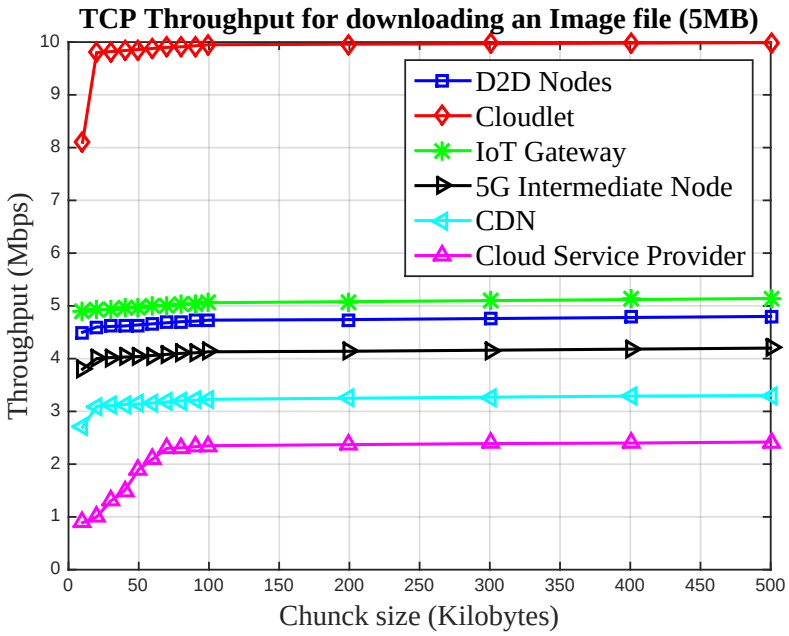


Figure 7.5: Throughput of downloading an image file from different caching servers for varying chunk sizes: Here, too, the throughput is less than the optimal when the chunk size is less than BDP of the link.

Moreover, it can be observed from Figure 7.4 that caching servers located physically close to the end users provide higher values of the throughput. For example, the maximum throughput observed in the case of content providers is about 1.2 Mbps even if the file is not divided into any

chunk. This is due to the reason that the network latency of a link to the content provider is around 230 milliseconds, which results in a higher Bandwidth Delay Product (BDP). For instance, in our scenario, the BDP of the content provider link turns out to be around 60 KB. It is important to note that the aforementioned estimated value of BDP is for the simulation scenario only. For real life scenarios, the BDP is quite inconsistent for a given flow and mainly depends on the network state, which varies with network congestion, buffer overflow, and queuing *etc.*. Smaller chunk sizes, such as 5KB or 10 KB restricts the data pipeline to be fully occupied with data thereby reducing the throughput. On the other hand, low latency links, such D2D nodes, constitute a lesser value of BDP, which allows even smaller chunk sizes to fill the data pipeline completely. For instance, a latency of 6 milliseconds between a client and a D2D cache server with 5Mbps data rate gives a BDP of 3KB. Therefore, a chunk size of 2KB provides much lesser throughput as compared to a 5KB chunk. Subsequent chunk sizes do not provide much throughput improvement on D2D links.

Figure 7.5 and Figure 7.6 present the TCP throughput for scenarios wherein an image file and a video file, respectively, is downloaded from the caching servers, as discussed in Figure 7.2. The throughput is, generally, directly proportional to the chunk size. However, there is an abrupt increase in the throughput when the chunk size surpasses the BDP of the flow.

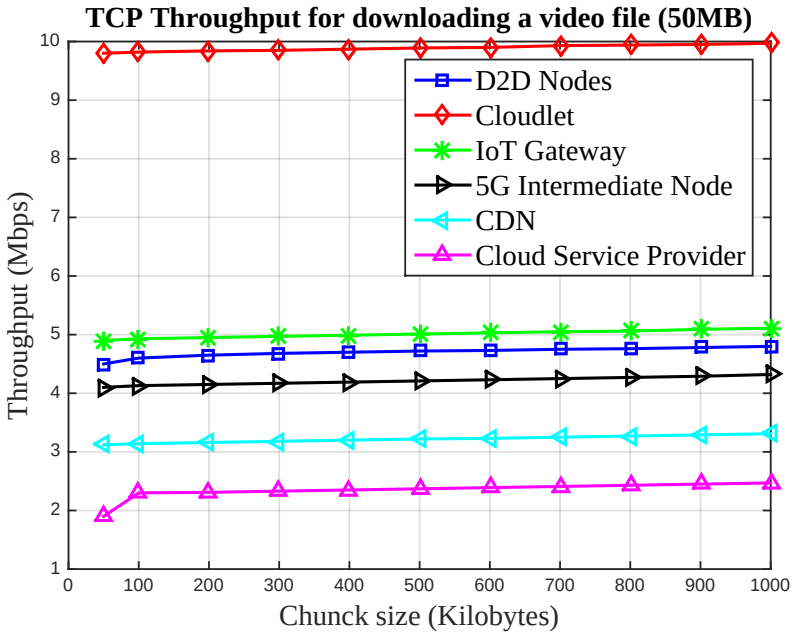


Figure 7.6: Throughput of downloading a video file from different caching servers for varying chunk sizes: Here, too, the throughput is less than the optimal when the chunk size is less than BDP of the link.

7.5 Related Work

One of the most relevant works to our proposal is blind cache [108]. The authors in [108] propose an out-of-band caching scheme for HTTPS traffic. The encrypted data is cached by CDNs while the encryption key is directly provided to the client out-of-band. The idea of an out-of-band caching is very interesting, but there are certain concerns with this approach. The most significant limitation of this approach is the inefficient management of duplicate contents. As the content is end-to-end encrypted, CDNs will store multiple copies of it if requested by multiple users in the region of same CDN server. The second limitation is the delivery of encrypted contents to users using lossy or unstable networks, such as wireless networks with weak signal strength or partial coverage. To re-establish a lost connection, the content provider has to start a new secure session with a new

session key. Moreover, new encryption keys are communicated to the client and the same data has to be cached again with new encryption keys. This is an inefficient utilization of communication and storage resources. On the other hand, our solution of dividing the content into smaller chunks and encrypting each chunk with its own hash code avoids the need of caching data again and again. Moreover, the link can be re-established from where it was interrupted (which depends on the chunk size) due to disconnection or any other similar reason.

The problem of reusability of cached contents is partially addressed by Leguay *et al.* in [113] by introducing a concept called *Cryptocache*. The authors propose to cache the contents based on pseudo-identifiers instead of using real identifiers. However, their solution does not solve the problem of data duplication. Moreover, their solution is not lightweight to be adopted by resource-constrained devices such as a smartphone with a lowly average computational power or IoT devices.

In [114], Mosko *et al.* provide a solution for secure caching in all-encrypted web. Their solution is analogous to blind caching solution proposed in [108]. This solution has same limitations of data duplication and reusability but is lightweight as compared to blind caching solution.

In [115], Yuan *et al.* propose an in-network caching scheme for delivering video files to end-users. The authors propose a request handler (dispatcher) in the network, which has the ability to identify, locate, and manage the in-network caching chunks. The main features of their solution include cache management and adaptive video delivery. However, the authors present their model for video files only and do not discuss the applicability of their approach to a mixed network traffic, more specifically, the case of IoT.

In [116], Engelmann and Elia exploit coded caching to preserve the privacy of end users requesting cached contents. The key features of their solution are confidentiality (user to content linking is not possible) and

hiding the popularity statistics of the cached contents. However, the authors do not discuss the applicability of their approach to lossy or unstable networks wherein the connection can break very frequently.

SDN/NFV based caching scheme is proposed in [117] for future mobile networks. The solution is well defined for wireless networks but the authors do not incorporate end-to-end encryption in their solution.

Similar solutions for securing cached contents are proposed in [118,119]. All these solutions do not seem to be practical for IoT devices having limited resources.

7.6 Discussion

7.6.1 Privacy Issue

Convergent encryption is a well-known cryptosystem that is utilized to efficiently store duplicate files. However, utilizing this scheme to ensure end-to-end encryption in cached contents can raise privacy issues. The caching server can predict the nature of the content by observing encrypted chunks, which could easily be generated from a suspected list of files. To overcome this, the encrypted chunks and hash codes could be masked with random numbers. These random numbers must be shared with the client out-of-band, along with the hash code and encrypted chunk ID.

7.6.2 Caching Location for Efficient Retrieval

The decision on where to cache the content along with the path to the client depends on the requirement of the application the end user is utilizing. If multiple users, in the proximity of a LAN, are accessing the same video from a content provider, the content provider can cache this video in the same LAN instead of caching it to a CDN server or even in a 5G node.

The applications, which require ultra low latency can be served by caching content in a D2D node *e.g.,* a smartphone or any other computationally capable node within user's premises.

7.6.3 Business Model

It is important to note that although the caching decision is taken by the content provider, the other stakeholders involved in providing the content to the client, such as D2D nodes, cloudlet, IoT gateway, and 5G intermediate nodes must obtain some monetary benefits in providing their resources/services as a caching server. In this regard, a business model must be investigated that efficiently transforms the physical resources of a caching server into potential incentives it can gain from the content provider or the end user, who needs low latency and high throughput.

As a starting point, the business model could be along the lines of CDNs with an ability to evolve with the requirements of end users. For instance, in case of caching at D2D nodes, which are already limited in resources (*i.e.,* battery, processing, memory, storage *etc.*), D2D nodes can offer their resources only if they are rewarded with some incentives from the content provider or the end user. The Mobile Network Operator (MNO) can also be included in the business model even if the content is cached within user's premises (such as D2D nodes or cloudlets) since it can significantly save bandwidth of the MNO in both access and core networks.

7.7 Chapter Summary

In this chapter, we present a marketplace for secure caching of the contents in 5G networks. We measure the throughput and latency gains for different file sizes at various locations as caching possibilities. We divide each file into smaller chunks for the possibility of reducing the duplicate contents

at the caching servers. We argue that reducing duplicate contents not only provides the storage gain at the caching servers but also reduces the communication load over the links between caching servers and the content provider. Moreover, our simulation results demonstrate that caching contents near the end users, such as at the cloudlet enables fast delivery of the contents with significant throughput gains. In addition, we also find that dividing a file into larger chunks provides higher throughput gains as compared to smaller chunks.

Chapter 8

Conclusions and Future Work

In this dissertation, we addressed some fundamental issues with D2D communication, which are critical to be addressed before the vast adoption of D2D in LBS, social services and public safety applications. In particular, we benchmark the energy consumption of D2D UEs in computational offloading and different single-hop and multi-hop scenarios. For each scenario, we define a certain threshold for optimal value of energy consumption, depending on the simulation scenario, we considered. Moreover, we propose solutions to bootstrap trust in D2D networks in untrusted environments. More specifically, we propose a PGP and reputation-based mechanism to incorporate trust in D2D networks. Finally, we propose a marketplace for securely caching popular contents in 5G networks.

It is important to note that this dissertation is a portion of our findings during the complete Ph.D. period. In some other works, we proposed an analytical model for energy consumption of TCP, which relates energy consumption to protocol operation cycles. Based on this model, a number of optimization techniques are proposed to reduce energy consumption of TCP [120, 121]. In another work, we propose energy savings in multi-hop D2D networks using cooperative beamforming [3]. Some other works include, but are not limited to, localization and proximity techniques in

LBS, the role of D2D communication in smart cities [122], security issues and challenges in implants and body area networks and more.

In this chapter, we briefly summarize the research contributions of the dissertation and outline some future directions emerging from this work.

8.1 Summary of Contributions

The fundamental contributions of this dissertation are stated here.

Energy Efficiency in Single-hop D2D Communication. In Chapter 3, we investigated the concept of MC for D2D communication. We presented a novel hybrid D2D communication architecture. The central SDN-controller has a global view of the network and consistently handles management of UEs belonging to different MCs. The local controller is responsible for managing the information flow within a single MC. The simulation results show that the concept of the training phase and mature phase can save up to 96.96% energy, once the network is in the mature phase.

Energy Efficiency in Multi-hop D2D Communication. In Chapter 4, we explored the possibility of tuning different parameters in WiFi Direct enabled multi-hop D2D networks. In particular, we proposed a power saving scheme that works on choosing optimal group size and transmit power of the UEs to optimize energy efficiency and throughput. Simulation results demonstrate that medium-sized groups (such as with 4 UEs) perform better in multi-hop scenarios. Moreover, transmitting with optimal power provides inherent security against various attacks. Simulation results reveal that gateway nodes in multi-hop networks are a potential bottleneck against higher throughput, while a large number of clients in a single-hop network potentially reduce the performance.

Computational Offloading to Mobile Clouds. In Chapter 5, we

characterized different offloading schemes in MCC. We presented a mathematical model for these offloading schemes and marked an offloading threshold for task completion time and energy consumption of the source mobile device. The task completion time includes the communication time with cloud and execution time of the task at the cloud. The idea is to offload the task in such a way that both task completion time and energy consumption can be minimized. The results from analytical model are consistent with NS-3 simulations. We conclude that smaller tasks, such as <50KB, are better to be executed locally in the smartphone. The medium sized tasks, such as 50KB to 4MB performs better if they are offloaded to local mobile cloud. For higher data sizes, such as >4MB, it is better to utilize the resources of remote cloud instead of using local mobile cloud as an offloading option.

Bootstrapping Trust in D2D Communication. In Chapter 6, we leverage the combination of PGP and reputation-based models to authenticate D2D users. Incorporating such security features in UEs helps to avoid connections with selfish users. This opens new doors to securely utilize D2D networks in growing proximity-based applications, such as social gaming and P2P social networking. In these scenarios, offloading traffic to D2D networks can provide significant throughput gains over the baseline LTE system, when devices communicate over cellular links. The simulation results show a mean throughput gain of 133% for 32 D2D links over a 9MHz frequency band.

Secure Caching in D2D Networks. In Chapter 7, we present a marketplace for secure caching of the contents in 5G networks. We measure the throughput and latency gains for different file sizes at various locations as caching possibilities. We divide each file into smaller chunks for the possibility of reducing the duplicate contents at the caching servers. We argue that reducing duplicate contents not only provides the storage gain

at the caching servers but also reduces the communication load over the links between caching servers and the content provider. Moreover, our simulation results demonstrate that caching contents near the end users, such as at the cloudlet enables fast delivery of the contents with significant throughput gains. In addition, we also find that dividing a file into larger chunks provides higher throughput gains as compared to smaller chunks.

8.2 Future Directions

The research work described in this dissertation can be extended along several directions.

Validation of Results through SDRs. In this dissertation, we proposed different solutions for energy efficiency and privacy in D2D communication. In future, we plan to validate our results by implementing our solutions to SDRs, such as ExpressMIMO2 by OAI.

Resource Allocation and Interference Mitigation through Network Virtualization. In this dissertation, we propose the concept of SDN controller as a fundamental entity of D2D network for saving energy. In future, we plan to extend the role of SDN controller to virtualize the network resources and perform the radio resource allocation to normal cellular UEs and D2D UEs, such that the interference between them is minimized. In addition, we plan to implement the resource allocation and scheduling scheme in SDRs as well. In this regard, a preliminary work has been recently accepted for presentation in GLOBECOM 2017 [123], which is related to test-bed implementation of end-to-end slicing in 5G networks.

Estimating Proximity Between devices for Energy Efficiency. In Chapter 4 of this dissertation, we propose to dynamically control the transmission power of source UE depending upon the distance between the source and the destination. To further extend this idea, and to accu-

rately estimate the distance between two UEs, communicating on direct D2D links, we plan to propose a hybrid scheme that can be based on the techniques, such as Time of Arrival (ToA), Angle of Arrival (AoA) and Received Signal Strength Indication (RSSI) between UEs. More specifically, we will study the advantage of MIMO in estimating the direct distance between UEs, exploiting a combination of aforementioned techniques with AoA.

Ensuring Privacy and Eliminating Duplicate Contents from Caching Servers. In this dissertation, we propose convergent encryption as a tool to secure cached contents across various caching servers. In future, we aim to present more detailed analyses of our proposal by providing combined storage and security gain over the other out-of-band caching schemes, such as blind cache. In addition, we aim to investigate more encryption schemes that can comply with the possibility of removing duplicate contents while preserving privacy at all ends.

Securing D2D Communication using Physical Layer Security. Physical Layer Security (PLS) is an emerging security scheme in wireless networks, which smartly exploits the imperfections in the wireless medium to safeguard wireless communication. In future, we plan to investigate PLS as a supplement to cryptography. PLS can be used to secure the communication phase of the network while cryptography can protect the processed data after communication phase [124].

Bibliography

[1] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, and J. Yao. 5G on the horizon: Key challenges for the radio-access network. *IEEE Vehicular Technology Magazine*, 8(3):47–53, Sept 2013.

[2] F. Granelli, A. A. Gebremariam, M. Usman, F. Cugini, V. Stamati, M. Alitska, and P. Chatzimisios. Software defined and virtualized wireless access in future wireless networks: scenarios and standards. *IEEE Communications Magazine*, 53(6):26–34, June 2015.

[3] J. Z. Moghaddam, M. Usman, F. Granelli, and H. Farrokhi. Cognitive radio and device-to-device communication: A cooperative approach for disaster response. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec 2016.

[4] T. P. C. de Andrade, C. A. Astudillo, and N. L. S. da Fonseca. Allocation of control resources for machine-to-machine and human-to-human communications over LTE/LTE-A networks. *IEEE Internet of Things Journal*, 3(3):366–377, June 2016.

[5] Antonino Orsino, Giuseppe Araniti, Leonardo Militano, Jesus Alonso-Zarate, Antonella Molinaro, and Antonio Iera. Energy efficient IoT data collection in smart cities exploiting D2D communications. *Sensors*, 16(6):836, 2016.

[6] Louis Columbus. Roundup of internet of things forecasts and market estimates, 2016. *Forbes, December*, 27, 2016.

[7] M. Usman, A. A. Gebremariam, F. Granelli, and D. Kliazovich. Software-defined architecture for mobile cloud in device-to-device communication. In *IEEE 20th International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD'15)*, pages 75–79, September 2015.

[8] M. Usman, M. R. Asghar, I. S. Ansari, F. Granelli, and K. Qaraqe. Towards energy efficient multi-hop D2D networks using WiFi Direct. In *IEEE Global Communication Conference (GLOBECOM)*, December 2017.

[9] W. Alliance. Wi-Fi Peer-to-Peer (P2P) Specification v1.1. *Wi-Fi Alliance Specification*, 1:1–159, 2010.

[10] S. Bluetooth. Available HTTP:http://www. bluetooth. com. *Bluetooth Specification Version 1.1*, 2001.

[11] Z. Alliance. Zigbee Specification. *Document 053474r06*, 1, 2006.

[12] Golrezaei, N. and Mansourifard, P. and Molisch, A.F. and Dimakis, A.G. Base-Station Assisted Device-to-Device Communications for High-Throughput Wireless Video Networks. *IEEE Transactions on Wireless Communications*, 13(7):3665–3676, July 2014.

[13] Golrezaei, N. and Dimakis, A.G. and Molisch, A.F. Device-to-device collaboration through distributed storage. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 2397–2402, December 2012.

[14] Asadi, Arash and Mancuso, Vincenzo. Energy Efficient Opportunistic Uplink Packet Forwarding in Hybrid Wireless Networks. In *Pro-*

ceedings of the Fourth International Conference on Future Energy Systems, pages 261–262, 2013.

[15] Asadi, Arash and Mancuso, Vincenzo. On the Compound Impact of Opportunistic Scheduling and D2D Communications in Cellular Networks. In *Proceedings of the 16th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 279–288, 2013.

[16] Qing Wang and Rengarajan, B. Recouping opportunistic gain in dense base station layouts through energy-aware user cooperation. In *IEEE 14th International Symposium and Workshops on World of Wireless, Mobile and Multimedia Networks (WoWMoM*, pages 1–9, June 2013.

[17] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs. Lte release 12 and beyond [accepted from open call]. *IEEE Communications Magazine*, 51(7):154–160, July 2013.

[18] 3GPP TR 36.843, study on LTE device to device proximity services; radio aspects, v12.0.1, 2014. Last accessed: 2017-10-10.

[19] 5G Americas. Inside 3GPP release 13: Understanding the standards for LTE-advanced enhancements, 2016. Last accessed: 2017-10-10.

[20] Cisco and/or its affiliates. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper, 2015.

[21] Y. Li, L. Sun, and W. Wang. Exploring device-to-device communication for mobile cloud computing. In *2014 IEEE International Conference on Communications (ICC)*, pages 2239–2244, June 2014.

[22] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, and G. Miklos. Design aspects of network assisted device-to-device commu-

nications. *Communication Magazine, IEEE*, 50(12):170–177, Mar. 05 2012.

[23] H. Bagheri, M. Katz, F. H. P. Fitzek, D. E. Lucani, and M. V. Pedersen. *D2D-Based Mobile Clouds for Energy- and Spectral-Efficient Content Distribution*. Springer, Apr. 06 2014.

[24] Hongkun Yang, Fengyuan Ren, Chuang Lin, and Jiao Zhang. Frequency-Domain Packet Scheduling for 3GPP LTE Uplink. *in proceedings of IEEE INFOCOM, San Diego, CA*, pages 1–9, Mar. 14-19 2010.

[25] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O.Spatscheck. 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 (MobiSys'12), NY, USA*, pages 225–238, 2012.

[26] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies. The case for VM-based cloudlets in mobile computing. *IEEE Pervasive Computing*, 8(4):14–23, Oct 2009.

[27] S. Tamoor-ul-Hassan, M. I. Ashraf, and M. D. Katz. Mobile Cloud based Architecture for Device-to-Device (D2D) Communication Underlying Cellular Network. *Wireless Days (WD), IFIP, Valencia, Spain*, pages 1–3, Nov. 13-15 2013.

[28] J. Mass, S. N. Srirama, H. Flores, and C. Chang. Proximal and Social-aware Device-to-Device Communication via Audio Detection on Cloud. *in Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia, Melbourne, VIC, Australia*, pages 143–150, Nov. 25-28 2014.

[29] K. Doppler, C. B. Ribeiro, and J. Kneckt. Advances in D2D Communications: Energy efficient Service and Device Discovery Radio. *2nd International Conference in Wireless Communication, Vehicular Technology, Information Theory and Aerospace and Electronic Systems Technology (Wireless VITAE), Chennai, India*, pages 1–6, Feb. 28 - Mar. 3 2011.

[30] X. Wu, S. Tavildar, S. Shakkottai, T. Richardson, J. Li, R. Laroia, and A. Jovicic. FlashLinQ: A Synchronous Distributed Scheduler for Peer-to-Peer Ad Hoc Networks. *48th Annual Allerton Conference in Communication, Control, and Computing, Allerton, IL*, pages 514–521, Sept. 29 - Oct. 1 2010.

[31] Z. Zhou, K. Ota, M. Dong, and C. Xu. Energy-efficient matching for resource allocation in D2D enabled cellular networks. *IEEE Transactions on Vehicular Technology*, PP(99):1–1, 2016.

[32] Gábor Fodor, Erik Dahlman, Gunnar Mildh, Stefan Parkvall, Norbert Reider, György Miklós, and Zoltán Turányi. Design aspects of network assisted device-to-device communications. *IEEE Communications Magazine*, 50(3):170–177, 2012.

[33] Wi-Fi Alliance. P2P technical group, Wi-Fi Peer-to-Peer (P2P) technical specification v1. 2, 2011.

[34] Wi-Fi Alliance. Retrieved april 1, 2017, 2017.

[35] D. Camps-Mur, A. Garcia-Saavedra, and P. Serrano. Device-to-device communications with Wi-Fi Direct: overview and experimentation. *IEEE Wireless Communications*, 20(3):96–104, June 2013.

[36] C. Funai, C. Tapparello, and W. Heinzelman. Enabling multi-hop ad hoc networks through WiFi Direct multi-group networking. In *2017*

International Conference on Computing, Networking and Communications (ICNC), pages 491–497, Jan 2017.

[37] Y. Wang, J. Tang, Q. Jin, and J. Ma. BWMesh: a multi-hop connectivity framework on android for proximity service. In *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, pages 278–283, Aug 2015.

[38] C. Casetti, C. F. Chiasserini, L. C. Pelle, C. D. Valle, Y. Duan, and P. Giaccone. Content-centric routing in Wi-Fi direct multi-group networks. In *2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 1–9, June 2015.

[39] Daniel Camps-Mur, Xavier Pérez-Costa, and Sebastií Sallent-Ribes. Designing energy efficient access points with Wi-Fi Direct. *Computer Networks*, 55(13):2838–2855, September 2011.

[40] IEEE LAN MAN Standards Committee et al. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. *IEEE Std*, 802, 2007.

[41] Matthew Gast. *802.11 wireless networks: the definitive guide.* " O'Reilly Media, Inc.", 2005.

[42] F. Lassabe, P. Canalda, P. Chatonnay, F. Spies, and O. Baala. A friis-based calibrated model for WiFi terminals positioning. In *Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks*, pages 382–387, June 2005.

[43] Celtrio Sarl. WLAN basics, 2007.

[44] M. Fiore, C. Casetti, and C. F. Chiasserini. Information density estimation for content retrieval in MANETs. *IEEE Transactions on Mobile Computing*, 8(3):289–303, March 2009.

[45] J. Luo, J. Zhang, Y. Cui, L. Yu, and X. Wang. Asymptotic analysis on content placement and retrieval in MANETs. *IEEE/ACM Transactions on Networking*, PP(99):1–16, 2016.

[46] N. Khaitiyakun, T. Sanguankotchakorn, and A. Tunpan. Data dissemination on MANET using content delivery network (CDN) technique. In *The International Conference on Information Networking 2014 (ICOIN2014)*, pages 502–506, Feb 2014.

[47] Q. Gong, Y. Guo, Y. Chen, Y. Liu, and F. Xie. Design and evaluation of a WiFi-Direct based LTE cooperative video streaming system. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec 2016.

[48] Keun-Woo Lim, Woo-Sung Jung, Hanna Kim, Jina Han, and Y. B. Ko. Enhanced power management for Wi-Fi Direct. In *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 123–128, April 2013.

[49] Wael Cherif, Muhammad Asif Khan, Fethi Filali, Sanaa, and Sharafeddine Zaher Dawy. P2P group formation enhancement for opportunistic networks with Wi-Fi Direct. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017.

[50] M. Conti, F. Delmastro, G. Minutiello, and R. Paris. Experimenting opportunistic networks with WiFi Direct. In *2013 IFIP Wireless Days (WD)*, pages 1–6, Nov 2013.

[51] R. I. Ansari, S. A. Hassan, and C. Chrysostomou. Energy efficient relay selection in multi-hop D2D networks. In *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 620–625, Sept 2016.

[52] L. Wei, R. Q. Hu, Q. C. Li, and G. Wu. Energy-efficiency of multi-hop device-to-device communications underlaying cellular networks. In *2014 IEEE International Conference on Communications (ICC)*, pages 5486–5491, June 2014.

[53] P. V. Mekikis, E. Kartsakli, L. Alonso, and C. Verikoukis. Flexible aerial relay nodes for communication recovery and D2D relaying. In *2016 IEEE 5th Global Conference on Consumer Electronics*, pages 1–2, Oct 2016.

[54] E. Datsika, A. Antonopoulos, N. Zorba, and C. Verikoukis. Green cooperative device-to-device communication: a social aware perspective. *IEEE Access*, 4:3697–3707, 2016.

[55] A. Antonopoulos, E. Kartsakli, and C. Verikoukis. Game theoretic D2D content dissemination in 4G cellular networks. *IEEE Communications Magazine*, 52(6):125–132, June 2014.

[56] M. Usman, M. R. Asghar, I. S. Ansari, and F. Granelli. Towards bootstrapping trust in D2D using PGP and reputation mechanism. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6, May 2017.

[57] James OToole. Mobile apps overtake PC internet usage in U.S. `http://money.cnn.com/2014/02/28/technology/mobile/mobile-apps-internet/`, 2014. Published: 2014-02-28, Accessed: 2016-07-21.

[58] S. Abolfazli, Z. Sanaei, E. Ahmed, A. Gani, and R. Buyya. Cloud-based augmentation for mobile devices: Motivation, taxonomies, and open challenges. *IEEE Communications Surveys Tutorials*, 16(1):337–368, First 2014.

[59] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 20152020 white paper. [online]. available:. `http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html`, 2016. Published: 2016-02-03, Accessed: 2016-07-21.

[60] Cong Shi, Vasileios Lakafosis, Mostafa H. Ammar, and Ellen W. Zegura. Serendipity: Enabling remote computing among intermittently connected mobile devices. In *Proceedings of the Thirteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '12, pages 145–154, New York, NY, USA, 2012. ACM.

[61] Frank H.P. Fitzek and Marcos D. Katz. *Mobile Clouds: Exploiting Distributed Resources in Wireless, Mobile and Social Networks*. Wiley Publishing, 1st edition, 2014.

[62] M. Usman, A. A. Gebremariam, U. Raza, and F. Granelli. A software-defined device-to-device communication architecture for public safety applications in 5G networks. *IEEE Access*, 3:1649–1654, 2015.

[63] Google app: Photo translator free. [online]. available:. `https://play.google.com/store/apps/details?id=com.smartmobilesoftware.phototranslatorlight&hl=en`. Accessed: 2016-07-21.

[64] Google app: Global mobile map viewing and navigation for online and offline osm maps - osmand+. [online]. available:. https://play.google.com/store/apps/details?id=net.osmand.plus&hl=en. Accessed: 2016-07-21.

[65] Andres Garcia-Saavedra, Pablo Serrano, Albert Banchs, and Giuseppe Bianchi. Energy consumption anatomy of 802.11 devices and its implication on modeling and design. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '12, pages 169–180, New York, NY, USA, 2012. ACM.

[66] Junxian Huang, Feng Qian, Alexandre Gerber, Z. Morley Mao, Subhabrata Sen, and Oliver Spatscheck. 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*, MobiSys '12, pages 225–238, New York, NY, USA, 2012. ACM.

[67] Trentino map. https://simple.wikipedia.org/wiki/Trentino-Alto_Adige/S%C3%BCdtirol. Accessed: 2016-07-21.

[68] NS-3, Network Simulator. [Online]. Available:. https://www.nsnam.org/. Accessed: 2016-07-21.

[69] Samsung Galaxy SIII Features. [Online]. Available:. http://www.gsmarena.com/samsung_i9300_galaxy_s_iii-4238.php. Accessed: 2016-07-21.

[70] Aaron Carroll and Gernot Heiser. The systems hacker's guide to the galaxy energy usage in a modern smartphone. In *Proceedings of the 4th Asia-Pacific Workshop on Systems*, APSys '13, pages 5:1–5:7, New York, NY, USA, 2013. ACM.

[71] Intel Core i7-3770K Processor. [Online]. Available:. `http://ark.intel.com/products/65523`. Accessed: 2016-07-21.

[72] L. Deboosere, P. Simoens, J. De Wachter, B. Vankeirsbilck, F. De Turck, B. Dhoedt, and P. Demeester. Grid design for mobile thin client computing. *Future Gener. Comput. Syst.*, 27(6):681–693, June 2011.

[73] Roelof Kemp, Nicholas Palmer, Thilo Kielmann, and Henri Bal. *Cuckoo: A Computation Offloading Framework for Smartphones*, pages 59–79. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[74] Byung-Gon Chun, Sunghwan Ihm, Petros Maniatis, Mayur Naik, and Ashwin Patti. Clonecloud: Elastic execution between mobile device and cloud. In *Proceedings of the Sixth Conference on Computer Systems*, EuroSys '11, pages 301–314, New York, NY, USA, 2011. ACM.

[75] Eduardo Cuervo, Aruna Balasubramanian, Dae-ki Cho, Alec Wolman, Stefan Saroiu, Ranveer Chandra, and Paramvir Bahl. MAUI: making smartphones last longer with code offload. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, pages 49–62, New York, NY, USA, 2010. ACM.

[76] Muhammad Usman, Dzmitry Kliazovich, Fabrizio Granelli, Pascal Bouvry, and Piero Castoldi. Energy efficiency of TCP: an analytical model and its application to reduce energy consumption of the most diffused transport protocol. *International Journal of Communication Systems*, pages n/a–n/a, 2015.

[77] B. Gao, L. He, L. Liu, K. Li, and S. A. Jarvis. From mobiles to clouds: Developing energy-aware offloading strategies for workflows.

In *2012 ACM/IEEE 13th International Conference on Grid Computing*, pages 139–146, Sept 2012.

[78] E.E. Marinelli. Hyrax: cloud computing on mobile devices using MapReduce. *Masters Thesis, Carnegie Mellon University*, 2009.

[79] Gonzalo Huerta-Canepa and Dongman Lee. A virtual cloud computing provider for mobile devices. In *Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond*, MCS '10, pages 6:1–6:5, New York, NY, USA, 2010. ACM.

[80] N. Fernando, S. W. Loke, and W. Rahayu. Dynamic mobile cloud computing: Ad Hoc and opportunistic job sharing. In *Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on*, pages 281–286, Dec 2011.

[81] Emmanouil Koukoumidis, Dimitrios Lymberopoulos, Karin Strauss, Jie Liu, and Doug Burger. Pocket cloudlets. *SIGARCH Comput. Archit. News*, 39(1):171–184, March 2011.

[82] Cong Shi, Mostafa H. Ammar, Ellen W. Zegura, and Mayur Naik. Computing in cirrus clouds: The challenge of intermittent connectivity. In *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, MCC '12, pages 23–28, New York, NY, USA, 2012. ACM.

[83] Yujin Li and Wenye Wang. Can mobile cloudlets support mobile applications? In *IEEE INFOCOM*, 2014.

[84] Ying-Dar Lin and Yu-Ching Hsu. Multihop cellular: A new architecture for wireless communications. In *19th Annual Joint Conference of*

the IEEE Computer and Communications Societies (INFOCOM'00), volume 3, pages 1273–1282, Mar. 2000.

[85] A. Asadi, Q. Wang, and V. Mancuso. A survey on device-to-device communication in cellular networks. *IEEE Communications Surveys Tutorials*, 16(4):1801–1819, Fourthquarter 2014.

[86] Avinash Srinivasan, Joshua Teitelbaum, Jie Wu, Mihaela Cardei, and Huigang Liang. *Reputation-and-Trust-Based Systems for Ad Hoc Networks*, pages 375–403. John Wiley & Sons, Inc., 2008.

[87] J. Jiang, S. Zhang, B. Li, and B. Li. Maximized cellular traffic offloading via device-to-device content sharing. *IEEE Journal on Selected Areas in Communications*, 34(1):82–91, Jan. 2016.

[88] Li Xiong and Ling Liu. Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities. *IEEE Transactions on Knowledge and Data Engineering*, 16(7):843–857, Jul. 2004.

[89] S. Ferrante, Q. Zhang, and B. Raghothaman. Capacity of a cellular network with D2D links. In *Proceedings of the 19th European Wireless Conference (EW'13)*, pages 1–6, April 2013.

[90] M. Zulhasnine, C. Huang, and A. Srinivasan. Efficient resource allocation for device-to-device communication underlaying LTE network. In *IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications*, pages 368–375, Oct. 2010.

[91] ITU M.2135-1. Guidelines for evaluation of radio interface technologies for IMT-advanced. Last accessed: 2016-10-12.

[92] IEEE802.org. Multi-hop relay system evaluation methodology (channel model and performance metric), 2007. Last accessed: 2016-10-12.

[93] Emre Telatar. Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, 1999.

[94] Philip R Zimmermann. *The official PGP user's guide*. MIT press, 1995.

[95] S. Capkun, L. Buttyan, and J. P. Hubaux. Self-organized public-key management for mobile ad hoc networks. *IEEE Transactions on Mobile Computing*, 2(1):52–64, Jan. 2003.

[96] Kui Ren, Tieyan Li, Zhiguo Wan, Feng Bao, Robert H Deng, and Kwangjo Kim. Highly reliable trust establishment scheme in ad hoc networks. *Computer Networks*, 45(6):687–699, 2004.

[97] Abderrezak Rachedi and Abderrahim Benslimane. Trust and mobility-based clustering algorithm for secure mobile ad hoc networks. In *International Conference on Systems and Networks Communications (ICSNC'06)*. IEEE, 2006.

[98] Haiyun Luo, Jiejun Kong, Petros Zerfos, Songwu Lu, Lixia Zhang, et al. Providing robust and ubiquitous security support for mobile ad hoc networks. In *Proceeding of The 9th International Conference on Network Protocols*, 2001.

[99] Mawloud Omar, Yacine Challal, and Abdelmadjid Bouabdallah. Fully distributed trust model based on trust graph for mobile ad hoc networks. *Computers and Security*, 28:199–214, 2009.

[100] Qi He, Dapeng Wu, and Pradeep Khosla. SORI: A secure and objective reputation-based incentive scheme for ad-hoc networks. In *IEEE Wireless communications and networking conference (WCNC'04)*, volume 2, pages 825–830, Mar. 2004.

[101] Kun Zeng, Przemyslaw Pawelczak, and Danijela Cabric. Reputation-based cooperative spectrum sensing with trusted nodes assistance. *IEEE Communications Letters*, 14(3):226–228, Mar. 2010.

[102] John R Douceur. The sybil attack. In *International Workshop on Peer-to-Peer Systems*, pages 251–260. Springer, 2002.

[103] James Scott, Jon Crowcroft, Pan Hui, and Christophe Diot. Haggle: A networking architecture designed around mobile users. In *WONS 2006: Third Annual Conference on Wireless On-demand Network Systems and Services*, pages 78–86, 2006.

[104] Muhammad Rizwan Asghar, Ashish Gehani, Bruno Crispo, and Giovanni Russello. PIDGIN: Privacy-preserving interest and content sharing in opportunistic networks. In *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*, ASIA CCS '14, pages 135–146, 2014.

[105] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira. Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges. *IEEE Communications Magazine*, 55(5):80–87, May 2017.

[106] E. Bastug, M. Bennis, and M. Debbah. Living on the edge: The role of proactive caching in 5G wireless networks. *IEEE Communications Magazine*, 52(8):82–89, Aug 2014.

[107] M. K. Kiskani and H. R. Sadjadpour. Secure coded caching in wireless ad hoc networks. In *2017 International Conference on Computing, Networking and Communications (ICNC)*, pages 387–391, Jan 2017.

[108] G Eriksson, John Mattsson, Nilo Mitra, and Zaheduzzaman Sarker. Blind cache: a solution to content delivery challenges in an all-encrypted web. *Ericsson white paper*, 2016.

[109] John R Douceur, Atul Adya, William J Bolosky, P Simon, and Marvin Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In *Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on*, pages 617–624. IEEE, 2002.

[110] R. Radescu and S. Pasca. Experimental results in prediction by partial matching and star transformation applied in lossless compression of text files. In *2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, pages 17–22, March 2017.

[111] Nadezhda Kozhemiakina, Vladimir V Lukin, Nikolay N Ponomarenko, Jaakko Astola, and Karen O Egiazarian. JPEG compression with recursive group coding. *Electronic Imaging*, 2016(15):1–6, 2016.

[112] Digital Rebellion. Video space calculator. Last accessed: September 2017.

[113] Jérémie Leguay, Georgios S Paschos, E Quaglia, and Ben Smyth. Cryptocache: Network caching with confidentiality. In *Communications (ICC), 2017 IEEE International Conference on*, pages 1–6. IEEE, 2017.

[114] Marc Mosko and Christopher A Wood. Secure off-path replication in content-centric networks. In *IEEE ICC 2017 Next Generation Networking and Internet Symposium (NGNI 2017). IEEE*, 2017.

[115] X. Yuan, X. Wang, J. Wang, Y. Chu, C. Wang, J. Wang, M. J. Montpetit, and S. Liu. Enabling secure and efficient video delivery through encrypted in-network caching. *IEEE Journal on Selected Areas in Communications*, 34(8):2077–2090, Aug 2016.

[116] F. Engelmann and P. Elia. A content-delivery protocol, exploiting the privacy benefits of coded caching. In *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pages 1–6, May 2017.

[117] Yaning Liu, Jean Charles Point, Konstantinos V Katsaros, Vasilis Glykantzis, Muhammad Shuaib Siddiqui, and Eduard Escalona. SDN/NFV based caching solution for future mobile network (5G). In *Networks and Communications (EuCNC), 2017 European Conference on*, pages 1–5. IEEE, 2017.

[118] A. Sengupta, R. Tandon, and T. C. Clancy. Fundamental limits of caching with secure delivery. *IEEE Transactions on Information Forensics and Security*, 10(2):355–370, Feb 2015.

[119] K. Thakker, C. H. Lung, and P. Morde. Secure and optimal content-centric networking caching design. In *2015 Second International Conference on Trustworthy Systems and Their Applications*, pages 36–43, July 2015.

[120] M. Usman, D. Kliazovich, F. Granelli, P. Bouvry, and P. Castoldi. A transport layer approach to improve energy efficiency. In *IEEE International Conference on Communications (ICC)*, pages 1–6, May 2016.

[121] Muhammad Usman, Dzmitry Kliazovich, Fabrizio Granelli, Pascal Bouvry, and Piero Castoldi. Energy efficiency of TCP: an analytical

model and its application to reduce energy consumption of the most diffused transport protocol. *International Journal of Communication Systems*, 30(1):e2934–n/a, 2017. e2934 IJCS-14-0785.R1.

[122] M. Usman, M. R. Asghar, and F. Granelli. *5G and D2D Communications at the service of the Smart Cities*. John Wiley & Sons, Inc., 2017.

[123] A. A. Gebremariam, M. Usman, P. Du, A. Nakao, and F. Granelli. Towards E2E slicing in 5G: A spectrum slicing testbed and its extension to the packet core. In *IEEE Global Communication Conference (GLOBECOM)*, December 2017.

[124] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. D. Renzo. Safeguarding 5g wireless communication networks using physical layer security. *IEEE Communications Magazine*, 53(4):20–27, April 2015.

Appendix A

Research Publications

Book Chapters

1. **Muhammad Usman**, Muhammad Rizwan Asghar, and Fabrizio Granelli. 5G and device-to-device communications at the service of Smart Cities, John Wiley & Sons, Inc., 2017. (Accepted)

International Journals

2. **Muhammad Usman**, Dzmitry Kliazovich, Fabrizio Granelli, Pascal Bouvry, and Piero Castoldi. Energy efficiency of TCP: An Analytical Model and its Application to Reduce Energy Consumption of the Most Diffused Transport Protocol, International Journal of Communication Systems, 2017.

3. **Muhammad Usman**, Anteneh A. Gebremariam, Usman Raza, and Fabrizio Granelli. A Software-Defined Device-to-Device Communication Architecture for Public Safety Applications in 5G Networks, IEEE Access 3:1649-1654, 2015.

4. Fabrizio Granelli, Anteneh A. Gebremariam, **Muhammad Usman**, Filippo Cugini, Veroniki Stamati, Marios Alitska, and Periklis Chatzimisios. Software Defined and Virtualized Wireless Access in Future

Wireless Networks: Scenarios and Standards, IEEE Communications Magazine 53(6):26-34, 2015.

5. **Muahmmad Usman**, Muhammad Rizwan Asghar, Imran Shafique Ansari, Fabrizio Granelli, and Khalid Qaraqe. Technologies and Solutions for Location-Based Services in Smart Cities: Past, Present, and Future, (In submission) 2017.

6. **Muhammad Usman**, Ameera Akhtar, Fabrizio Granelli. Mobile Cloud Computing: Remote Cloud or Local Mobile Cloud? A Quantitative Analysis, (In submission) 2017.

7. Javad Zeraatkar Moghaddam, **Muhammad Usman**, Hamid Farrokhi, and Fabrizio Granelli. Cognitive Radio Meets Device-to-Device Communication: A Cooperative Approach for Disaster Response, (In submission) 2017.

International Conferences and Workshops

8. Anteneh A. Gebremariam, **Muhammad Usman**, Ping Duy, Akihiro Nakaoy, Fabrizio Granelli. Towards E2E Slicing in 5G: A Spectrum Slicing Testbed and its Extension to the Packet Core, Globecom 2017 (Accepted)

9. **Muhammad Usman**, Muhammad Rizwan Asghar, Imran Shafique Ansari, Fabrizio Granelli and Khalid Qaraqe. Towards Energy Efficient Multi-hop D2D Networks using WiFi Direct, Globecom 2017 (Accepted)

10. **Muhammad Usman**, Muhammad Rizwan Asghar, Imran Shafique Ansari, and Fabrizio Granelli. Towards bootstrapping trust in D2D using PGP and reputation mechanism, IEEE International Conference on Communications (ICC), Paris, 2017, p:1-6.

11. Javad Zeraatkar Moghaddam, **Muhammad Usman**, Fabrizio Granelli, and Hamid Farrokhi. Cognitive Radio and Device-to-Device Communication: A Cooperative Approach for Disaster Response, IEEE Global Communications Conference (GLOBECOM), Washington, DC, 2016, p:1-6.

12. **Muhammad Usman**, Dzmitry Kliazovich, Fabrizio Granelli, Pascal Bouvry, and Piero Castoldi. A transport layer approach to improve energy efficiency, IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, p:1-6.

13. **Muhammad Usman**, Anteneh A Gebremariam, Fabrizio Granelli, and Dzmitry Kliazovich. Software-defined architecture for mobile cloud in device-to-device communication, IEEE 20th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Guildford, 2015, p:75-79.

14. **Muhammad Usman**, Muhammad Rizwan Asghar, Imran Shafique Ansari, Fabrizio Granelli, Qammer H. Abbasi and Khalid Qaraqe. A Marketplace for Efficient and Secure Caching for IoT Applications in 5G Networks, (In submission) 2018.

15. **Muhammad Usman**, Muhammad Rizwan Asghar, Imran Shafique Ansari, Marwa Qaraqe, Fabrizio Granelli and Qammer H. Abbasi. The need for a business models for V2X communication in future 5G networks, (In submission) 2018.

Appendix B

Vitae

Muhammad Usman was born in Sargodha, Pakistan on May 15, 1987. To pursue his Ph.D., he joined the Future Networks Lab at the Department of Information Engineering and Computer Science (DISI), University of Trento, Italy in November 2014. During Ph.D., he investigated energy efficiency and privacy in D2D communication (presented in this dissertation), under the supervision of Associate Prof. Dr. Fabrizio Granelli and Dr. Muhammad Rizwan Asghar.

He was a Visiting Researcher in the Electrical and Computer Engineering Department (ECEN) at the Texas A&M University, Doha, Qatar from February to May 2017 and from August to December 2017. Prior to joining Doctoral School, he was a Research Assistant at the University of Trento from July to October 2014. He received two M.Sc. degrees in 2014 due to his excellent academic achievements; One in Telecommunications Engineering from the University of Trento, Italy and one in Computer Networks from Sant'Anna School of Advanced Studies, Pisa, Italy. The M.Sc. was a part of double degree program between the two universities. He carried out his research on "Benchmarking Energy Efficiency of TCP" during his Master's thesis at the University of Luxembourg, Luxembourg. He obtained his B.E. degree in Electronics Engineering from the School of Electrical Engineering and Computer Science (SEECS) at National University of Sciences and Technology, Islamabad, Pakistan in 2008. During his career, he held many positions in both academia and industry.

His research interests include D2D communication, software-defined networking, energy efficiency, privacy and security.

Homepage: https://sites.google.com/view/muhammad-usman