Proximity as a Service via Cellular Network-Assisted Mobile Device-to-Device

by

Tianhao Guo

Supervised by John Schormans

Supervision Panel: Eliane Bodanese and Tijana Timotijevic

A thesis submitted to the University of London for the degree of
Doctor of Philosophy

Department of Electronic Engineering
Queen Mary, University of London
United Kingdom

November 2019

TO MY FAMILY

Abstract

The research progress of communication has brought a lot of novel technologies to meet the multi-dimensional demands such as pervasive connection, low delay and high bandwidth. Device-to-Device (D2D) communication is a way to no longer treat the User Equipment (UEs) as a terminal, but rather as a part of the network for service provisioning. This thesis decouples UEs into service providers (helpers) and service requesters. By collaboration among proximal devices, with the coordination of cellular networks, some local tasks can be achieved, such as coverage extension, computation offloading, mobile crowdsourcing and mobile crowdsensing.

This thesis proposes a generic framework Proximity as a Service (PaaS) for increasing the coverage with demands of service continuity. As one of the use cases, the optimal helper selection algorithm of PaaS for increasing the service coverage with demands of service continuity is called ContAct based Proximity (CAP). Mainly, fruitful contact information (e.g., contact duration, frequency, and interval) is captured, and is used to handle ubiquitous proximal services through the optimal selection of helpers.

The nature of PaaS is evaluated under the Helsinki city scenario, with movement model of Points Of Interest (POI) and with critical factors influencing the service demands (e.g., success ratio, disruption duration and frequency). Simulation results show the advantage of CAP, in both success ratio and continuity of the service (outputs).

Based on this perspective, metrics such as service success ratio and continuity as a service evaluation of the PaaS are evaluated using the statistical theory of the Design Of Experiments (DOE). DOE is used as there are many dimensions to the state space (access tolerance, selected helper number, helper access limit, and transmit range) that can influence the results. A key contribution of this work is that it brings rigorous statistical experiment design methods into the research into mobile computing.

Results further reveal the influence of four factors (inputs), e.g., service tolerance, number of helpers allocated, the number of concurrent devices supported by each helper and transmit range. Based on this perspective, metrics such as service success ratio and continuity are evaluated using DOE. The results show that transmit range is the most dominant factor. The number of selected helpers is the second most dominant factor. Since different factors have different regression levels, a unified 4 level full factorial experiment and a cubic multiple regression analysis have been carried out. All the interactions and the corresponding coefficients have been found.

This work is the first one to evaluate LTE-Direct and WiFi-Direct in an opportunistic proximity service. The contribution of the results for industry is to guide how many users need to cooperate to enable mobile computing and for academia. This reveals the facts that: 1, in some cases, the improvement of spectrum efficiency brought by D2D is not important; 2, nodal density and the resources used in D2D air-interfaces are important in the field of mobile computing. This work built a methodology to study the D2D networks with a different perspective (PaaS).

Acknowledgments

First and foremost, I want to thank my primary supervisor John Schormans. He has a golden heart who gave me constant support, patience, and tolerance and granted me to academic visit for one year. John has often made me feel privileged to be his PhD student. He has shown me how research should be carried out and also encouraged me to stay in academia to continue my research. There are many things John has taught me, but nothing is more precious than acting as a role model not only in academia but also in life. As Confucius said, "The gentlemen nourishes the roots. With the roots established, the way grows." During PhD, John taught me by his own example what are the roots of a gentleman: the kindness, tolerance and dedication for the family. There is a Chinese saying: "One day as a teacher, one life as a father." John has deeply changed the way I think and live. After graduation, I hope I can ask for John's suggestions on how to be a teacher, a husband , a father and a gentleman.

A similar gratitude belongs to my secondary supervisor Eliane Bodanese and Tijana Timotijevic. Their work ethic and high standard nurtured my academic skills and aesthetic. They shared their expertise and provide precious advices in the field of wireless networks.

I would like to thank Dr Yue Cao, who accepted my visiting and taught me a complete workflow. He gave me many precious advices to be a researcher and I am deeply touched by his self-discipline and enthusiasm to explore use cases and bridge the gap between academia and industry.

I want to express my special gratitude to my mother. Without her, I cannot go through depression and finish the PhD study. My father also provided me solid support on my research and shared his experience as a PhD.

I also want to thank my best friend, my brother, and soul mate Wu Jinze. There is too much I want to say but so little that can be expressed. He is like a mirror and helps me understand myself. He has witnessed every step of my growth and gave me solid helps at critical moments.

Jiawan Xu gives me a feeling of home by her company. She is a caring and independent girl with a good skill of cooking. We established a home together with five guinea pigs: tofu, pudding, sheep, flower and koala. Every time I leave the campus, Jiawan and guinea pigs are the lighthouses which navigate me back home in this big city.

I also want to thank all the members of the network group. Dr Bo Zhong, Dr Yuanwei, Dr Lifeng, Dr Yansha, Dr dantong , Dr Liumeng once gave me precious suggestions. Dr Xueke accommodated me when I first came back to London from visiting. My special gratitude belongs to Dr Xingyu who gave me all the suggestions and helps he can provide when I started my PhD study. The special gratitude also belongs to my roommate Jinze who takes care of everything when I was not in London and Xiaoshuai who plays table tennis with me.

It is my honour to stay in this great city and this great university for four years to study. It is challenging and exciting. Thank you to everyone I have met along the way.

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP 3rd Generation Partnership Project

$AdjR^2$ Adjusted R-Squares

API Application Programming Interface

AR Access Request

BS Base Station

BT-CMAB Budgeted and Time-Limited Combinatorial Multi-Armed Bandit

CAP ContAct based Proximity

D2D Device-to-Device

DOE Design Of Experiments

DTN Delay/Disruption Tolerant Networking

EDRB Event-Driven and Role-Based mobility model

EFR Emergency First Responders

EMS Emergency Medical Services

EWMA Exponential Window Moving Average

Fo-SDD Fog-assisted Secure Data Deduplication

GUI Graphical User Interface

H Helper

HetNets Heterogeneous Networks

IOT Internet of Things

LW Levy Walk mobility model

MaaS Mobility-as-a-Service

MAB Multi-Armed Bandit

MAR Mobile Augmented Reality

MCC Mobile CrowdSourCing

MCS Mobile CrowdSensing

MEC Mobile Edge Computing

MIMO Multiple Input Multiple Output

mmWave millimeter Wave

MNOs Mobile Network Operators

MOOP Multi-Objective Optimization Problem

MPAD Mobility Prediction based Adaptive Data

NOMA Non-Orthogonal Multiple Access

NS-3 Network Simulator 3

ONE Opportunistic Network Environment simulator

PaaS Proximity as a Service

POI Points Of Interest

$PredR^2$ Predicted R-Squares

PSN Pocket Switched Network

R Requester

RD Random Direction mobility model

RNIS Radio Network Information Service

RPGM Reference Point Group Mobility model

RTT Round Trip Time

RW Random Walk model

RWP Random WayPoint

SR Smooth Random mobility model

TTI Time To Intercept

UE User Equipment

WDMM Working Day Movement Model

Chapter 1

Introduction

In recent years, the explosive growth of miscellaneous devices along with various service demands[4], has brought numerous challenges, e.g., the growth of data traffic by orders of magnitude, ubiquitous coverage and incredible variety of service requirements to current network infrastructures.

The need to increase the network capacity and enrich various services have been commonly recognized as key features in the future networks. Several enabling technologies, e.g., massive Multiple-Input and Multiple-Output (MIMO), millimeter Wave (mmWave), Non-Orthogonal Multiple Access (NOMA), femtocell and Heterogeneous Networks (HetNets) have been extensively studied in the literature[4–7] and partly applied.

Compared to other technologies, Device-to-Device (D2D) is a much cheaper solution[8], by enabling direct communication between mobile devices without necessarily involving cellular links in data transmissions. Thanks to the developing features of mobile phones, User Equipment (UE) with such powerful resources will be able to play an attractive role as service suppliers in proximity, other than as requesters in nature.

D2D communication can use various air interfaces (in-band underlay D2D, in-band overlay D2D, network-assisted out-band D2D, and autonomous out-band D2D). In-band underlay means D2D networks and cellular networks use the same licensed spectrum. In-band overlay means D2D networks and cellular networks use different licensed frequencies, which means that Mobile Network Operators (MNOs) need to allocate dedicated frequencies for D2D communication. Out-band means that D2D networks work in unlicensed bands, such as Bluetooth and WiFi-Direct. If the infrastructure centrally controls the transmission, it is network-assisted out-band D2D; otherwise it is autonomous out-band D2D. A detailed comparison is elaborated in Table 1-A.

- Since licensed bands are owned by MNOs, both kinds of in-band D2D air-interfaces are network-assisted, which means **a dedicated D2D architecture is needed.**[1]
- **In-band D2D air-interfaces have not been adopted by smartphone manufactures.** Although a LTE-Direct chipset has been developed by Qualcomm[9], there is no smartphone manufacture adopting it, to our best knowledge.

Table 1-A: Comparison of Different D2D Air Interfaces

Characteristics	In-Band Underlay D2D	In-Band Overlay D2D	Network-Assisted Out-Band D2D	Autonomous Out-Band D2D
Dedicated D2D architecture	Yes	Yes	Yes	No
Enabled by phone manufactures	No	No	Yes	Yes

The maximum transmission range of in-band D2D air interface (LTE-Direct) is designed to be 500 m[10] because dedicated in-band D2D technique can use its own bands which are suitable for longer transmit range without interference. However, with the consideration of shadow loss, multi-path effects, and sensitivity of battery life, the transmission range of LTE-Direct implemented in phones in urban areas cannot reach 500 m. Therefore, the common setting of D2D transmission range is 150 m[11].

[1]For example, a D2D server in the edge for network-assisted D2D networks

1.1 History of Endeavour on D2D

This thesis reviews the brief history of earlier works on D2D. Through the history, different understandings of the concept "D2D" are elaborated.

Mobile networks where communication is set up with peer-to-peer manner opportunistically are known in the literature as opportunistic networks[12]. These networks are based on direct communication between devices, which forward messages from one device to another when a communication opportunity (contact) takes place. Therefore, opportunistic networks cannot guarantee end-to-end connectivity. Conventional network protocols, such as TCP, require continuous connectivity between the communicating parties throughout a transmission, so this requirement cannot be met easily by mobile devices. Delay/Disruption Tolerant Networking (DTN), a new network architecture enabling communication in the absence of end-to-end connectivity, is suitable in opportunistic networks.

The concept of Pocket Switched Networks (PSN) was proposed in 2005[13]. PSN are networks formed by mobile devices carried by people. An example of wireless interface of devices in this experiment[13] is Bluetooth. Due to the mobility nature of users and limited transmit ranges of mobile devices, the links of PSN are intermittent. So PSN also belong to the paradigm of opportunistic networks.

In contrast, in the dense scenario where the peer-to-peer manner communication can always find a route to maintain stable end-to-end connectivity, ad hoc[14] or mesh networks would be a suitable choice for TCP service provision. According to my previous work[15] in Xidian University, edge computing such as TCP proxy implemented in service providers can support TCP-IP services in complicated wireless environments. However, with even more challenging conditions, e.g., a 5 hours' disruption, the service would normally fall into the Delay/Disruption Tolerant Networking paradigm.

Different from the works mentioned above, which use out-band wireless interfaces, K. Doppler[8] began to focus on the mobile network with in-band underlay air interfaces and named this as a Device-to-Device network. Different from out-band and in-band overlay D2D, in-band underlay D2D has attracted more attention in academia partly because it has an explicit research goal: improving performance such as spectrum efficiency by coordinating transmit power based on distances[16, 17] among D2D network users and cellular network users.

In recent years, D2D was standardized by the 3rd Generation Partnership Project (3GPP) Release-12[18], with basic functions such as discovery, one-to-one communication, one-to-many communication and UE relaying. This standardization of D2D in 3GPP is called "proximity service", which implies the ambition that MNOs want to be "proximal smart service enablers". Furthermore, 3GPP has always been trying to integrate commercial cellular networks with public safety networks. D2D is a way of narrowing the evolution gap because D2D can be a fallback solution[19, 20]. For example, it can increase the robustness when base stations fail in disasters.[2]

1.2 Motivation

As one of the candidate technologies in the future networks, whether Device-to-Device (D2D) communication can be widely implemented in industry is uncertain. To the best of our knowledge, there has been no smartphone vendor adding an in-band D2D wireless air interface such as an LTE-Direct chipset[10] into their smartphones yet. Moreover, there is no MNO implementing dedicated D2D architecture (D2D servers) and protocol stacks for in-band or out-band D2D networks. However, other competing technologies[4] such as massive MIMO, HetNets and mmWave have been implemented or have been planned to be implemented.

[2]Please refer to the proactive movement PaaS in future works.

1.2.1 Limits of D2D

There is no doubt that D2D communication has the potential to expand the coverage of cellular networks by UE-relay[11]. It seems that in-band underlay D2D air interfaces can increase spectrum and energy efficiency by power coordination[16, 17] and D2D can reduce latency by direct communication. However, the improvement does not solely depend on MNOs but **depends heavily on the users.**

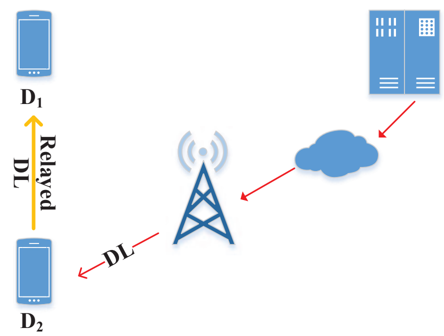


Figure 1.1: D2D Communication with a Remote End

- **For the service of content offloading, D2D technology cannot always improve spectrum efficiency and it is not persuasive enough.** For example, when the two ends of the communication are not in proximity, as shown in Fig 1.1, content is downloaded (DownLink, DL) from a remote server to a device (D_1). In this case, D2D communication only plays a role of relaying (Relayed DL) or forwarding (D2D with a remote end), but cannot improve spectrum efficiency since the contents have already been downloaded by the cellular network once.

  However, if the contents are reused by multiple devices (popularity), and D2D networks and cellular networks use the same licensed spectrum, D2D technology can increase spectrum efficiency[16, 17]. Essentially, D2D technology would instead be considered as releasing the burden of cellular networks rather than increasing the spectrum efficiency. In this case, if the D2D communication releases the traffic on cellular networks by out-band D2D air-interfaces and the unlicensed bands are not severely congested, it is more economical and “spectrum efficient”.

- **For other proximal services, the number of devices with D2D functionality and the transmission range of D2D air interfaces must reach a certain level.** If the two ends of communication are in proximity, namely proximal services within the same cell, the direct communication between pairwise devices can obviously release the traffic on cellular networks. But if there are only a few pairwise D2D technology enabled users within a cell, the performance gain of D2D networks is not enough to let MNOs pay attention to implementing the D2D architecture.

  The increase of D2D transmission range relies on allocating more dedicated resources (for example, bands with lower frequencies and longer transmission ranges) to in-band D2D air interfaces but apprently, the motivation of resource allocation is not convincing. For the current transmission range of D2D (50 m), if the nodes are populated densely but proximal services are sparse, the contents can only be retrieved when two devices opportunistically **contact** each other or when there is a link between these two devices (in the manner of ad hoc or mesh). Multiple hops means occupying more specturm resources than single hop D2D communication. It can be even worse if nodes are populated sparsely which cannot maintain a stable multi-hop link. A message transfer is mostly dependent on the nature of node mobility and sporadic encounter events between devices (users) as they congregate and disperse[21]. The latency can be hours or days[22–25].

  Even for the proximal service of reused content offloading, if the density of devices is insufficient, it is easy to imagine how much popular content can be stored on devices since storage spaces of a few mobile devices provide no advantage compared to content servers at edge.

This thesis proposes a one hop proximal service and experiments verify that the performance of D2D networks is highly reliant on the density of D2D enabled devices and the transmission range of D2D air interfaces. Since improving the efficiency of D2D technology without consideration of proximal services has already been extensively researched[16, 17], **exploring the potential of proximity services** to attract more users and reversely forcing the network operator to allocate more resources to D2D air interfaces could be a worthwhile endeavour in the future.

1.2.2 Distinctive Potential of Proximal Services via D2D Technology

Telecommunication service providers possess resources that influence our daily lives like traditional infrastructure service providers such as water or electricity providers. However, they do not have the monopoly right like other infrastructure providers. On the contrary, the competition between MNOs is fierce because of the high speed development of technologies in IT field, just like the Internet companies. Telecommunication service providers are also not like the Internet companies, since the scale of the market of traditional telecommunication is fixed and nearly saturated[26] in developed countries. The number of people who pay for broadband services and mobile services are basically fixed. Also, the amount of the service fees are relatively fixed no matter how many generations of telecommunication technologies have been updated. MNOs have to keep running like mice in the running wheel in the competition environment, namely **"Red Ocean Traps"**.[3] MNOs have always been trying to be **"smart enablers"** like Internet companies rather than **"dumb pipes"** because smart enablers can create profits without limit (**"Blue Ocean"**).

Far beyond increasing efficiency of the cellular networks like other competing technologies, **the distinctive potential of D2D technology is knitting our lives together with proximity services by sharing proximal resources.** D2D would also be the fallback solution for push-to-talk services[27] when the infrastructure falls in the disaster scenario. These are irreplaceable features that other competing technologies cannot

[3] Competition in "red ocean" is an economy terminology, which means a "zero sum game".

provide.

Exploring the potential of proximity services via D2D technology is promising to boost the endeavour in both academia and industry since it can make our lives better and safer, and infrastructure providers can be “smart proximal service enablers” with the unlimited use cases of proximal services. Since various smart devices have become an indispensable part of daily lives and the resources of mobile devices are spilling over, the broad social consensus[28] and dramatic rise of sharing economy (like Uber[29] and Airbnb[30]) would sooner or later shed a light on the field of mobile computing.

1.3 Research Focus, Nature of PaaS, and Research Aims

1.3.1 Research Focus

It is a novel way to study the D2D communication together with the proximal services. For any service, at least there must be a service provider and a service requester. Since the focus of this research is the proximal service, the pairwise **devices** should be decoupled as the helper and the requester, and the term **“D2D collaborative systems”** is to represent the aforementioned proximal service system enabled by D2D.

D2D technology plays a role of a bridge knitting people together with proximity resources. Different from traditional researchers in the field of D2D, this thesis focuses not only on the bridge but also on both ends of the bridge, treating them as a system, because the D2D collaborative system encompasses mechanisms through which mobile smartphones interact with each other by D2D communication for a specific proximal service. In D2D fog computing[1] (mobile crowd computing), for example, human behaviour (mobility as a human behaviour), D2D communication (the bandwidth resource), and the computational resource (the CPU frequency) are all indispensable parts of the system.

This is the basic perspective and methodology of this thesis.

The model of our D2D collaborative system for proximal services is called **Proximity as a Service (PaaS)** and D2D communication is the bridge in PaaS. The PaaS model PaaS hides all the complexity of the physical layer, simplifying the bridge as transmit range and bandwidth. The range and bandwidth can be dynamic, changing with the underlying conditions.

This thesis proposes the simplified system model of PaaS, as shown in Fig 1.2, the user or the local business that wants to conduct a local task is named **"a requester" (R)** after proposing a proximal service request. Then the cellular networks select the suitable set of devices (named **"helpers" (H)**) and forms a network-assisted collaborative D2D system in a crowdsourcing manner. The requester will pay MNOs for the proximal service after finishing the task and helpers can get rewards from the MNOs. The outcome of the task can be transmitted by D2D communication or cellular communication.

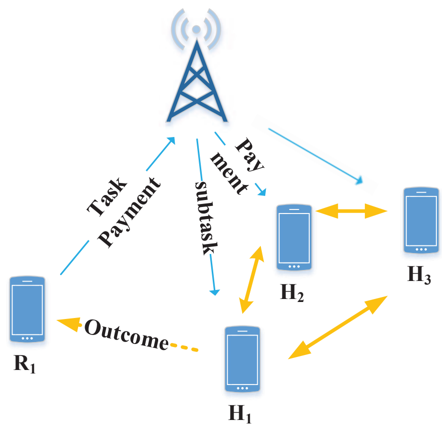


Figure 1.2: Simplified PaaS

1.3.2 Nature of PaaS

1.3.2.1 The Features of PaaS Systems

- **Humans are participants regardless of the kinds of devices.** Devices can be various, such as smartphones, vehicles, and drones, to name a few, but all these devices have interactions with humans. For example, humans have the right to enable or turn off D2D communication. Even if the vehicles and drones are unmanned, these devices still comply with the will of humans.

  In the case that the D2D functionality is enabled by a participant, if the participant proactively takes part in the collaboration systems such as mobile crowdsourcing, the participant is called the **active participant**. If the PaaS is transparent to users, such as the opportunistic mobile crowdsensing (e.g. pothole detection for road surface monitoring [31]), the participant is called the **passive participant**.

  Note that selfishness, altruistic and cooperative behaviour are all basic human nature. An incentive mechanism is to motivate users to participate in the PaaS and exchange their resources. Moreover, privacy and security are also basic human rights, so privacy and security are not out of fashion topics in all D2D networks.

- **Mobility.** Undoubtedly, mobility is the nature of mobilephones, vehicles, and drones. It is well-known that D2D networks cannot operate without D2D discovery because of the mobility nature. Thinking further, mobility is also one of the forms of human behaviours (labour), which has social and geographical features. By exploiting these features, a more efficient discovery mechanism can be designed [32]. Moreover, the movements of the crowd can be treated as "collective intelligence and labour", so mobility itself is a service, namely Mobility-as-a-Service (MaaS)[22, 33]. In other words, **exploiting mobility is exploiting human labour**, so the movement of the participant can also be proactive or passive, based on the feature of the human participant.

The mobility nature of PaaS brings the advantage of ubiquity in service coverage, but it also has the disadvantage of **bringing the opportunistic nature of all PaaS systems, resulting in the disadvantage of service continuity and delay.** However, the disadvantage is brought by passive mobility since passive movement is not centrally controlled but constrained by social and geographical constraints. When the movements of all the mobile nodes are centrally controlled for a mission or nodes are moving proactively for proximal demands,[4] D2D collaborative PaaS can be deployed in the field in a timely fashion because of the pervasiveness and ubiquity of smart mobile devices. Note that, proactive movement of active participants is likely to only apply in use cases with incentive mechanisms or extreme use cases such as disasters and public safety.

- **Resource shortage.** Mobile devices, especially smartphones, have always suffered the shortage of resources such as energy, computation power, storage space, and bandwidth. Moreover, it is unreasonable to take for granted a stable D2D link, which means that all the resources are opportunistic. The D2D communication opportunities in the collaborative D2D networks are brought by the mobility of the crowd and, as aforementioned, mobility is a form of human resource. The opportunities of D2D communication can be scarce when the node density is sparse. Therefore, to complete the tasks (services), the mechanism of PaaS must make the best of these resources. Mechanisms may include helper selection based on mobility prediction and time-sensitivity; resource allocation based on the known current resource status of the helpers.

1.3.2.2 Technical Challenges of Resource Allocation in PaaS

Note that, the main technical challenges of the research is

- **Decision making and resource allocation.** Decision making is the discrete

[4]Please refer the chapter of future works.

form of resource allocation; and resource allocation is a continuous form of decision making. For example, recruiting a device as a helper for a proximal service is a decision. How much resource (bandwidth, storage and computation resources) of the helper are allocated to the proximal service is resource allocation.

- **Performance analysis.** There are many factors that can influence the performance of the PaaS, which means a wide input space. It is a challenging task to bring rigorous statistical experimental design methods into the research of PaaS to extract full information (all the interactions and the corresponding coefficients).

Every decision has a "price". The price can be literally the payment of incentive mechanism. Selfishness is human nature so participants will not share their resource for free. Therefore, the incentive mechanism can also be integrated with the resource allocation algorithm. The price can also be the "opportunity cost" (alternative cost)[5] and "waiting cost" because of the opportunistic nature brought by mobility.[6]

- Proximal resources are opportunistic, which means they are not always available to be allocated.

- If the resource is not available, sometimes the system cannot predict exactly when it will be available again. When a requester meets a helper, the resources of the helper are available. However, this helper might be not perfect (high price, not enough resource). The requester may not know exactly when the "perfect" helper can come, but the requester are assumed to know exactly the time constraint of the task. Therefore, the helper selection is a challenging problem.

- Proactive movement can be designed to alleviate the opportunistic nature of mobile proximal resources. But the incentive mechanism should be considered since human

[5]This is a terminology of microeconomics theory. The New Oxford American Dictionary defines it as "the loss of potential gain from other alternatives when one alternative is chosen." This terminology plays crucial role to make sure that scarce resources are used efficiently

[6]The opportunistic nature can also be brought by wireless channel conditions such as shadowing and interference, not only by mobility.

resources are expensive. An easy case is the public safety scenario.

1.3.3 Research Aims

Since this thesis has a unique perspective on D2D networks, and proposes a D2D collaborative system with helpers and requesters, the research aim is to conduct a complete work of a simple use case of PaaS, which can pave a way that can be used to other use cases. The complete work includes three parts:

- Finding a way to categorize previous use case exploration in proximal services.
- For a simple use case, finding the optimal helper selection or optimal resource allocation scheme in scenarios with the consideration of mobility.
- Exploring the nature of this case by comprehensive performance analysis using DOE and finding how factors influence the performance.

1.4 Contributions in this Thesis

1.4.1 A Use Case of PaaS and a Helper Selection Algorithm

A generic framework of PaaS has been implemented in a simulator [7] and the use case of opportunistic access enhancement via the framework of PaaS has been done[11]. In [11], some phones work as service providers to share communication resources during pairwise contact to other devices which do not have access to the cellular networks. The algorithm for coverage enhancement via the generic PaaS framework is called as ContAct based Proximity (CAP). This name is a pun with the meaning of covering UEs ubiquitously (ideally seamlessly).

[7] Please refer to chapter 3

As mentioned above, human participants, mobility and resource shortage are common features of PaaS systems. These features are also considered in the design of CAP.

- **Human Participants:** The incentive mechanism, privacy and security are not the focus in CAP, but the expense, privacy and security issues of recruiting helpers were considered. For one requester, only a limited number of helpers can be selected to provide the access enhancement service.

- **Mobility:** The geographical and social constraints of passive movement are jointly implemented in CAP. The focus of CAP is to optimally find the limited number of helpers based on the information of mobility learned from contact history. The passive movement drives the opportunistic nature of service provision (service continuity), but service tends to be seamless along with the increase of the number of helpers or increase of the transmission range (change WiFi-Direct to LTE-Direct).

- **Resource Shortage:** The UE-relay service is sharing bandwidth resources. Considering the shortage of bandwidth resources and energy, the system assumes that a helper can only have limited connections with several requesters. The focus of CAP is not bandwidth resource allocation among several requesters based on their urgency or quality requirements, but the bandwidth allocation functionality and consideration of energy have been implemented in the generic PaaS system. Bandwidth allocation could be a possible future work.

- **Proximal Service and Evaluation Metrics:** The service of access enhancement manages to guarantee a minimum access time (requested access duration) within a tolerable period (access tolerance). For example, the setting could be accessing the network for half an hour within two hours deadline. If the minimum access time is completed, it is called a success. If the service is successful, continuity (how often and how long the disruptions happen) is evaluated during the span of accessing time.

1.4.2 Performance Analysis by Design Of Experiment (DOE)

An experiment can be treated as a process with inputs (factors) and outputs (responses). Four outputs and four inputs are carefully chosen so that the responses can reveal the nature of the PaaS system and the influence of factors is meaningful. In this work, three key issues have been solved.

- Statistical experimental design is applied to determine levels of factors and conduct full factorial[34] designs.
- Four polynomial multiple regression equations[34] for four responses are obtained and a comparison of regression lines and simulation results based on the equations are drawn.
- Based on the four polynomial multiple regression equations, a non-linear multi-objective optimization has been completed.

1.4.3 Revised and Published Papers in Queen Mary University of London

1.4.3.1 Published Papers

- T. Guo, G. Huang, J. Schormans, T. Wang, and Y. Cao, “Cap: A contact based proximity service via opportunistic device-to-device relay,” in 2017 International Symposium on Wireless Communication Systems (ISWCS), pp. 337342, Aug 2017.
- Wang, Tong and Li, Pengcheng and Wang, Xibo and Wang, Yunfeng and Guo, Tianhao and Cao, Yue, “A comprehensive survey on mobile data offloading in heterogeneous network,” in Wireless Networks Springer, pp. 1-12, 2017.

1.4.3.2 Revised Paper

- T. Guo, J. Schormans, Lexi Xu, Jinze Wu and Y. Cao, "Proximity as a Service for the Use Case of Access Enhancement via Cellular Network-Assisted Mobile Device-to-Device" IEEE ACCESS.

1.5 Structure of the Thesis

The generic PaaS framework has the potential to be tailored for various use cases. In chapter 2, related works on use case exploration are reviewed. Some use cases only need mirror implementations to the current PaaS framework, which are discussed in the chapter of future works (chapter 5). The implementation of current PaaS in the simulator is introduced in the chapter 3. A complete methodology of the use case of access enhancement of PaaS is elaborated in chapter 4 and chapter 5. In chapter 6, the contributions and unfinished work of this thesis are concluded. Based on the drawbacks of the works, a detailed plan of future work is shown in chapter 7.

Chapter 2

Literature Review

2.1 A Review of Exploration of Use Case of Proximal Services

As mentioned in the section on motivation in chapter 1, exploring the potential of proximity services via the D2D collaborative system is an immensely promising way to improve people's lives and boosting the endeavour into D2D in both academia and industry. Hence, it is necessary to review previous works on use case exploration on proximity services via the D2D collaborative system and pave the way for future directions.

Up-to-now, the exploration of use cases with different service requirements is in full swing, including cellular traffic offloading[35], opportunistic crowd computing[36, 37], mobile augmented reality[38], mobile crowdsourcing[39], target marketing[40], and even block-chain applications (LocalCoin)[41]. In [41], the computational hardness (Bitcoin) has been replaced by social collaboration hardness, which might be a generic measurement of contribution in social collaboration activities. However, to our best knowledge, there is no review on use case exploration on the D2D collaborative system focusing on proximity services with the vision of proximal resources. For any review, the taxonomy is the first and most important thing.

2.1.1 The Choice of Taxonomy

The choice of the taxonomy is of great importance, and the criteria are as follows.

- Since there are various use cases in the paradigm of D2D communication, the boundary needs to be clear to fit the motivation.
- The taxonomy needs to reveal the essence of various related works, disclose the mask of various "fashion" concepts in the literature, and provide readers a closer view to the academic works in the field of mobile computing.
- The taxonomy should be an never out of fashion tool to categorise the previous and future works on proximal service via D2D collaborative system and could be a foundation for exploring new proximal services.
- The taxonomy should try to point out why the authors made the selection of mathematical models and algorithms to tackle their specific problems. Therefore, this review work can be of academic significance.

The D2D technology is a bridge, enabling the sharing of different kinds of proximal mobile resources, which therefore creates various promising proximal services. A certain category of proximal services is a combination of proximal resources and has its unique service requirements, combined with the opportunistic nature in the D2D collaborative system (mobility, wireless channel), which brings fruitful research problems. To conclude these problems, this thesis chooses "proximal opportunistic resources" as the perspective of this survey. The reasons are as follows:

- The nature of all the current and future proximal services is sharing proximal resources, so taking proximal resources as the taxonomy could be not an out of fashion tool to see through myriad concepts and appearance to perceive the essence.

- Mobile devices suffer resources shortages, so resource allocation and decision making are always the key of mobile computing. To think further, the combination of resources, in turn, determines the selection of mathematical models and algorithms. Therefore, using resources as the taxonomy to systematically review the literature can provide readers a closer view to the way to creating mobile computing systems.

2.1.2 Taxonomy

Here, proximity resources are categorized into five aspects:

- **Human Resources:** Human resources can be modelled as workability (with or without task preference). As aforementioned in the common features of the PaaS system, humans are participants regardless of the devices being unmanned controlled or hand held. Mobility (regardless of proactive or passive movement), in a sense, is also a form of human resource. This strongly explains the reason that incentive mechanisms, privacy and security could always be the topics in the field of mobile computing. The explanatory power, in turn, proves that choosing resources as the taxonomy is of academic significance for use case exploration.

- **Sensing Resources:** Sensing resources come from the sensors in the mobile device and are similiar to human resources: sensing resources can be modelled as sensing ability and sensing preference. However, the sensing resource has its own distinctiveness. For example, devices close to each other provide similar data[42]. Environmental context[43] is also important in sensing tasks. For example, a phone needs to record ambient light, but when the phone is put in the pocket, it is energy-wasting to keep sensing, and the data is no longer valuable.

- **Computation Resources:** Intensive computation tasks can be offloaded to nearby devices to save time or save energy. The computation resources can be CPU clock cycles[1].

- **Communication Resources:** Bandwidth and coverage of mobile devices can be shared among each other. For example, in this thesis, PaaS provides coverage extension[11].

- **Storage Resources:** If proximity devices have cached popular contents, contents can be sent to proximal devices directly[44]. For example, in Mobile Augmented Reality (MAR), people who stay in proximity share similar data on the virtual objects. If a device does not have enough storage space for the popular contents, contents can also be stored in a group of nearby devices[45]. It is necessary to implement caching and pre-fetching algorithms so that the needed virtual objects are cached and shared locally.

In all the proximal services, mobile devices acting as helpers inevitably use communication, storage, energy and human resources to provide helps to requesters, so, in some sense, helpers are sharing those resources. However, in this thesis, the definition of sharing resources is restricted into the case that helpers are allocating a particular amount of resources for the usage of requesters. Although it is inevitable that these resources are overlapping with other resources in proximal services, researchers usually take into account only one resource or a few combinations of the resources as the target of optimization, since there are only a few resources playing as the main roles for proximal services. For example, the target of designing a mobile crowdsourcing system is specifically exploiting human resources, although communication resources are also used inevitably.

This review focuses on the distinctiveness of every resource in each category. Energy is not discussed as a category because devices cannot directly share energy.[1]

[1] Energy harvesting or wireless power transfer might be promising in the future, but the distance is too small now.

2.1.3 Human Resources Sharing via D2D

The human resource is an indispensable part of the collaborative D2D system and even mobility is a form of human resource. Current approaches to crowdsourcing are mainly web-based and do not exploit the full potential of local proximal resources[46]. Unlike traditional web-based crowdsourcing, mobile crowdsourcing focuses on selectively recruiting workers in proximity and exploiting their work ability along with their mobility. While mobile crowdsourcing does not depend on any specific underlying network, the D2D based mobile crowdsourcing is highly desired when the tasks need a lot of interactive operations.

In [47], a framework named "Crowd Foraging" was proposed for self-organized mobile crowdsourcing based on opportunistic networks. A mobile task requester can proactively recruit a crowd of opportunistic encountered mobile workers in real time in a distributed manner, which is different from traditional web-based crowdsourcing. In this framework, a requester recruits an arriving worker by jointly considering the ability of workers, timeliness, task profile, and reward of tasks, which make up the utility function for worker selection. The value of the utility function is service quality, indicating the expected service gain.[2] The authors formulated the sequential worker recruitment problem as an online multiple stopping problem and accordingly derive an optimal worker recruitment policy through the dynamic programming principle. The authors conducted extensive trace-driven evaluations (traces of mobility and work ability) to prove the superiority of this work and an Android prototype was implemented, which showed that the framework is time/energy efficient.

In [46], the authors argued that current participatory sensing approaches usually do not consider users as intelligent participants in sensing processes. However, their opinions can complement to captured sensor data, which is a combination of sensing resources

[2]Since the tasks of crowdsourcing are not specific, the evaluation of this kind of service is usually service gain.

and human resources for a better sensing effectiveness.

2.1.4 Sensing Resources Sharing via D2D

The efforts of monitoring and obtaining information from the mobile crowd is called Mobile CrowdSensing (MCS). The difference between Mobile CrowdSourCing (MCC) and MCS is that work ability is narrowed to sensing ability and the evaluation of the task is narrowed to sensing coverage and quality. Since MCS has unique requirements on geographical [42] and environmental context[43], MCS needs to be discussed separately in a single category.

A MCS platform can harvest the data received by a group of people that deliberately express their availability for crowdsensing activities. In terms of the direction of data flow, in MCS system, the data is uploaded, which is different to data dissemination services. The MCS platform generally assigns tasks which are mostly data collection activities[48]. These activities can be conducted automatically by the MCS app without any human interaction (except for initially accepting the task). These actions may also require some active intervention by the participant, e.g., to provide explicit users feedback about an event or proactively move to a certain place (refer to mission-driven movement and demand-driven movement in chapter 7).

Therefore, MCS can be divided into participatory sensing and opportunistic sensing [49]. Participatory sensing requires proactive actions such as moving to a particular place (with specific trajectory), while opportunistic sensing requires minimal involvement such as continuous location sampling and pothole detection[31] with accelerometers.

The work of [43] belongs to opportunistic sensing, which means sensing decisions are application or device-driven, namely, minimal intervention from the user side. The sensing decision is made by device based on a threshold.[3] For example, when the environmental context is unfavorable, the devices never perform sensing and reporting. An unfavorable situation occurs, when, for example, the smartphone is in the pocket, and the device needs to record current ambient light. The sensing utility is calculated by the combination of data collection utility, smartphone sensing potential, and environmental context. Data collection utility is influenced by the average number of samples generated from a sensor in an area during a timeslot.[4] A low value of data collection utility indicates that a large number of samples have been already received at the cloud collector and further reporting is not needed. Vice versa, a high value of data collection utility indicates the need for more samples. Since the goal of [43] is energy efficiency, sensing potential is a function of locally spent energy, which is composed of sensing and communication energy consumption. The evaluation is conducted on the specific simulator CrowdSenSim[50] and the results show that the energy cost for sensing is negligible with respect to the energy spent for communications.

[51] considers the diversity of task locations and user trajectories, and then formulates the optimal task scheduling problem as a continuous path planning problem. Since this is an NP-hard problem, two online heuristic algorithms were proposed to maximize the task quality. The algorithm tries to find the task with the largest ratio of task quality increment to travel cost and hence guides the travel path to areas of high task density.

In [52], the simulation experiments are based on human contact traces during social events (e.g., conference) using Bluetooth devices (e.g., mobile phones, iMotes). Maximizing sensing coverage is modelled as an optimization problem using vertex cover problem in graph theory. An approximation algorithm tailored for the crowdsensing task is proposed. The core of the algorithm is to assign the sensing task to a more "socialized" participant to get a better sensing coverage.

[3]In decision theory, decision is made by threshold and the value of utility.

[4]The time of the experiment is divided into timeslots

A critical problem in MCS is to select appropriate mobile device users to perform sensing tasks, which can maximize sensing coverage and sensing quality. Since the users' utilities are unknown, the challenge is how to effectively learn the users' utilities and make optimal user selection decisions. Budgeted Multi-Armed Bandit (MAB) approaches are suitable for these problems. In [53], a Budgeted and Time-Limited Combinatorial Multi-Armed Bandit (BT-CMAB) model and corresponding algorithm are proposed to tackle this problem. The evaluation results show that the proposed algorithm efficiently utilizes budget and time to achieve a low regret.[5] The innovation of this work is the combination of time sensitivity to previous budgeted MAB.

In [42], the authors proposed a problem: users close to each other in geographical positions usually provide similar sensory data, which is a waste of resources. A scheme in which two users within a limited geographical distance cannot get monetary rewards simultaneously and two user selection heuristic algorithms are proposed. Experiments are based on real location data sets (Brightkite and Gowalla), which proves the algorithms are efficient. In [54], replicate data is solved afterwards, which is data deduplication. Hence a Fog-assisted Secure Data Deduplication framework (Fo-SDD) was introduced, which is a combination of sensing resources and computation resources.

Compared to traditional sensor networks, the features of MCS are as follows.

- MCS devices have more computation, communication, storage and multiple sensing capabilities.
- MCS devices are controlled by human beings, which can further smartly sense the physical world.
- Billions of mobile devices have already been deployed "in the field", which reduces the cost of deployment of dedicated sensing infrastructure.

[5]Low regret is the target in the bandit model

- MCS platforms can perform local analytics to process raw sensing data into intermediate results and then send to a back-end for further usage, which can save bandwidth and storage resources.

- On one hand, the quality of sensing data in terms of accuracy, latency, and confidence can change all the time due to device mobility, capabilities and the preference of owners. On the other hand, MCS is more flexible. For example, by proactive movement, MCS platforms can be quickly deployed in the field.

2.1.4.1 Specific Simulator for MCS

In [55], Network Simulator 3 (NS-3) for crowdsensing simulations is proposed. NS-3 simulates highly accurate estimations of network properties at the expense of scalability. Hence it is extremely challenging to perform simulations with a large number of users.

A specific simulator for MCS was proposed in [50]. The features of this simulator are: scalability (in the order of tens of thousands of users), 3-D geographical information (latitude, longitude, and altitude), and output displayed on the real map (with the help of tools of Google Map). However, since this is a newly proposed simulator, the current version of this simulator has two drawbacks. Firstly, limitation of movement. For example, each node can only move once, and the duration of this move is only between [10, 20] minutes. Secondly, the wireless interface technology of the simulator only contains WiFi technology. There is no essential difference in the design of the simulator between participatory sensing and opportunistic sensing although the authors announced so. The most practical way to enable participatory sensing might be enriching the mobility functionality including proactive group movement in this simulator.

2.1.5 Computational Resource Sharing via D2D

Because of the processing and energy constraint of mobile devices, computation offloading is necessary. The computation tasks can be offloaded to remote servers (cloud computing), local edge servers (Mobile Edge Computing, MEC) and proximal mobile devices. Cloud computing is suitable for execution of heavy tasks, but the distance and congestion of data transmission narrows its application on medium-weight tasks with high latency sensitivity. In this thesis, D2D offloaded execution and MEC offloaded execution are called fog computing. The aims of fog computing are saving energy and time.

Needless to say, computation offloading to local servers is the most intuitive solution for tasks with high latency sensitivity. However, unlike cloud servers, the computation power of MEC servers is limited. For example, in [56], the edge cloud sets prices to maximize its revenue, and for a given price, each user locally makes offloading decision to reach a tradeoff between latency and payment. Moreover, the cellular links cannot always be stable. So it is necessary to research task offloading towards nearby mobile devices, in other words, computation resources sharing via D2D networks.

Devices have the flexibility of choosing among multiple options for task executions, including local execution, D2D offloaded execution, and MEC offloaded execution. The author of [57] developed a hybrid task offloading framework among these three options and accordingly a three-layer graph matching algorithm to handle the choice space. The problem of minimizing the total task execution cost was recast as a minimum weight matching problem over the constructed three-layer graph and hence was solved by the Edmonds's Blossom algorithm. The evaluation was conducted by event-driven simulator Opportunistic Network Environment simulator (ONE). However, in this work, a device was allowed to establish and maintain at most one D2D link during a task offloading slot. If a crowd of devices wants to relieve computation burden of a proximal device in a crowd-sourcing manner, one-to-many or many-to-many D2D should be enabled.

To boost the participation rate, an incentive scheme tailored for computation offloading was proposed[58] with both a payment scheme and a reputation scheme with a truthful auction strategy. The peer-to-peer reputation exchange scheme in this work is like people exchanging money in social collaborative systems such as markets. Collaborating devices get rewarded for their contribution, while selfish ones get sidelined by others. However, just like economy events in the physical world, a market without global information might be losing control and not be safe, just like financial or economical crisis. Therefore, a centralised and network-assisted collaborative computation offloading framework would be a promising research topic.

The aforementioned D2D computational offloading schemes are distributed. A centralised supernode at the base station responsible for scheduling cooperation tasks based on user mobility was proposed in [59]. The coordination of the centralised controller enabled the connectivity-aware task scheduling scheme for D2D computation offloading. A heuristic algorithm was further proposed to ensure a low cooperative task execution time.

Besides mobility uncertainty, the resource availability of each user is also difficult to predict. A network-assisted D2D computation offloading collaboration framework, called D2D Fogging[1], with an incentive scheme and a resource availability prediction was formulated by Lyapunov optimization. By jointly considering the execution and communication energy consumption, the proposed online algorithm achieved superior performance. The object of D2D Fogging is to achieve energy efficient task executions. They devised task scheduling policies in three kinds of system settings in a time frame because nodes are static and cellular links are stable within the time frame. In this work, network (communication) resources and computation resources are jointly formulated. By using the granularity of time frame, the problem of mobility can be simplified. This work was powerful enough to prove the potential of D2D collaboration in sharing computation resources without resources from infrastructure.

Because of the opportunistic nature of proximal resources, D2D offloaded execution has a disadvantage on ensuring timely feedback, which is contradictory to the aim of fog computing[60]. Therefore, PaaS computation offloading should be a systematic framework with global information on tasks, latency evaluation (remote servers, local servers and proximal devices) and available resources. However, previous works on D2D offloaded execution have not comprehensively considered these issues. Some works on MEC have considered latency evaluation and task analysis.

In [61], the authors utilized queuing theory to formulate the energy consumption, execution delay, and price of offloading process in a mobile-edge cloud system. Both wireless transmission and computing capabilities were considered when modelling the energy consumption and delay performance. The innovation of [62] is to use the Radio Network Information Service (RNIS) Application Programming Interface (API) to drive the decision of user equipment (UE) to offload or not. For example, for a particular application, the framework on the MEC server estimates the current value of Round Trip Time (RTT) between the UE and the MEC server, according to radio quality indicators available through RNIS API, coupled with energy consumption or other parameters, to make the offloading decision.

Note that tasks are various:

- Some tasks (subtasks[63], some parts of a task) are offloadable, but some are not[64].
- Different tasks or subtasks have different time sensitivity.
- Some tasks or subtasks are cardinal but some are optional

2.1.6 Communication Resource Sharing via D2D

Cellular traffic offloading and cellular coverage extension might be the most intuitive applications of communication resource sharing PaaS. WiFi networks or Femtocells are another two options, and they have already played an important role in indoor environments.

However, D2D has its unique advantages: **ubiquitous, timely and deployable**. The D2D collaborative system can be applied in environments including outdoor[11] with partial infrastructure coverage or extreme situations[6] such as disasters, emergency, caves, and mines.

Ubiquity, seamlessness and economy are metrics of communication resource sharing in PaaS. In a particular use case, a tradeoff should be considered. For example, in metropolitan areas with a high population density[25], the cellular network needs to disseminate a popular delay-tolerant content to a bunch of nodes. There are two extreme solutions: 1, cellular networks disseminate the content to every node, which is timely but costly; 2, cellular networks transfers the content to only one node (economical), and then this node transmits the content to other nodes opportunistically (not timely[7]). The authors in [25] chose a target set of nodes which are more socially popular than other nodes to offload the content. In this framework, the authors also set a time constraint for the case that a node can still get the content from the cellular network if this node has not received the content through D2D within the required time constraint.

The mobility nature of D2D brings the advantage of ubiquity in coverage extension, but it also brings the disadvantage of intermittence in both cellular offloading and coverage extension. However, when the movements of all the mobile nodes are centrally controlled

[6]Please refer PaaS for Mission and PaaS for demands in chapter 5.

[7]There is a possibility for multi-casting, but it is not sure that there exist end-to-end links to all the nodes simultaneously.

for a mission or nodes are moving proactively for proximal demands,[8] D2D collaborative PaaS can be timely deployed in the field because of the pervasiveness and ubiquity of smart mobile devices nowadays. That is the reason that 3GPP has always been trying to integrate public safety network with LTE by D2D networks.

2.1.7 Storage Resource Sharing via D2D

The network traffic shows the asynchronous content reuse property [65], which means that a few popular contents account for most of the data traffic. Therefore, caching the most popular traffic at the network edge [66], at Femtocell (Femtocaching) [67], or at proximal devices [45, 68–70] would reduce delay and save energy.

The difference between D2D caching and D2D store-and-forward offloading is that the destination of data is unknown when data is cached, and the objective of caching data is improving data access performance[45], rather than saving bandwidth. Researches on caching tend to reach a tradeoff between data accessibility and caching overhead[45]. To cache data at appropriate network locations so that queries in the future can be responded to with less delay, query prediction is of great importance. The prediction can exploit the query history[45] or combine data popularity[71, 72] (geographic, social and local popularity). Moreover, the nodes with high social popularity are more suitable for caching data.

To overcome the limited caching buffer of mobile devices (if this device has high social popularity in the network), a group of proximal nodes [45] can help this node (central node). Popular data is always cached nearer to the central nodes via dynamic cache replacement based on query history. However, the essential contradiction is that the author tried to build a caching framework to ensure queries in the future can be responded to with less delay in a delay tolerant network. However, if a device needs a

[8] Please refer to future work.

quick response, cellular networks would be a better choice. This contradiction limits the application of this framework, which can only be used in areas without infrastructure.

The inspiration of [45] for PaaS is that caching can also be a proximal service. Sometimes a mobile device needs to perform a fog computing task, with a significant amount of required local data, which exceed the storage space of the mobile device and edge server,[9] for example video compression in the neighbourhood[73]. Proximal devices should share not only computing resources but also storage resources. This framework can be called "fog caching", which is designed specially to serve fog computing. Based on the combination of communication resources and computation resources in D2D fogging framework[1], as shown in Fig 2.1, a more realistic framework for large-scale fog computing tasks can be proposed in the future, which would combine three proximal resources (communication, storage, and computation resources).

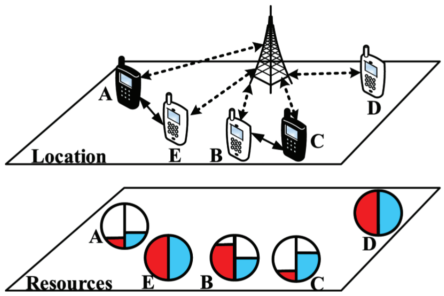


Figure 2.1: Illustration of the combination of communication and computing resources from[1]. The dashed (solid) arrows represent cellular (D2D) links, and the red (blue) color portion represents available CPU (bandwidth) resources.

Caching can also be used in content distribution and retrieval. A promising application is Mobile Augmented Reality (MAR). Augmented Reality (AR) needs to download data in advance for "augmenting" elements in the physical world by computer-generated perpetual information. **The data for AR is spatially registered with the physical world.** Therefore, researches on local caching for MAR at the edge[69] or by collab-

[9] The storage space of the server on the edge might be small and partially used for a particular fog computing task.

orative proximal D2D[70] are necessary and promising. Moreover, MAR also needs computational power, which can be a proximal service combined with several proximal resources, such as storage, computation and communication resources.

File transfer is also a proximal service using sharing storage resources. The authors in [74] discussed the legal problems, trust mechanisms, UI issues, and business model with respect to current device-to-device file transfers services based on third-party applications (SHAREit, Xender, and Zapya). The authors found out D2D networks for file transfer is promising because current applications are enabled by a WLAN interface in a master and slave manner (access point and device manner), which is not efficient compared with specific D2D air interfaces such as WiFi-Direct or LTE-Direct.

2.2 Mobility

In this thesis, mobility is regarded as a form of human resource, but mobility models have particular features that need to be reviewed separately. The most active period of researches on mobility was at late 1990s[21]. However, the mobility of the nodes directly impacts the performance of the protocols, mechanisms, and algorithms of any PaaS in the simulation. Therefore, a better understanding of the mobility pattern of devices is the key to design effective PaaS mechanisms and experiments.

2.2.1 Mobility Traces

Mobility traces are data collected from experiments in the physical world, which can reflect the nature of mobility patterns. Therefore, real-world traces are benchmarks in the research into mobility.

According to [21], there are two ways of trace collection: a polling-based method and an event-based method. In the polling based method, the devices make records at a regular time interval, which is suitable to collect GPS traces.[10] In the event-based method, when two devices come within the communication range of each other, an association record is made; similarly, when devices move out of each other's range, a disassociation record is also made. The latter method is suitable to collect connectivity traces[75].

Apart from the aforementioned traces which can be used in PaaS, some researchers have collected macro-mobility traces by marks of hand-overs in the cellular networks. Macro-mobility traces record devices movement when devices move from one cell to another. The dataset can contain millions of devices[76, 77] and the duration of the experiments can be several months[77], which is suitable for studying the relationship between human mobility and social events or environmental conditions.

One of the well-known repositories of mobility traces is CRAWDAD.[11] The formats of traces in CRAWDAD may not conform to specific network simulators because there is no common standard. Therefore researchers often have to modify and convert the traces to the correct format before using them as input to the simulator[21]. Bonn Motion[12] is a Java-based software which can be used to modify traces.

2.2.2 Movement Models

As [78] defined, movement models try to mimic the mobility of real mobile nodes that change speed and direction with time. Therefore, all the nodes have an initial distribution (starting point for a single node); every node has a velocity $V(t)$; after reaching the destination (or through an epoch, or a step), the node would start to move again (or after a pause time).

[10] wirelesslab.sjtu.edu.cn/

[11] https://www.crawdad.org/

[12] http://sys.cs.uos.de/bonnmotion/

Table 2-A shows the pros and cons of synthetic movement models and mobility traces. In most cases, traces are limited to a university campus or a small conference area. Synthetic movement models are useful when researchers need to simulate proactive movement in various network environments with different air-interfaces.[13]

Table 2-A: Comparison of Synthetic Movement Models and Mobility Traces

Characteristics	Synthetic Mobility Models	Mobility Traces
Enabling proactive movement	Yes	No
Enabling various air-interfaces	Yes	No
Scalable	Yes	No
Similarity with real world movement pattern	Relatively low	Relatively high
Computation complexity	High	Low
Applicability	Not limited	Limited

Movement models can be divided into total randomness models (or called statistical movement models) and constrained randomness, such as: obstacles, pathway, speed limit, group (community), home cell (reference points) and proactive (designed or mapped)). Since PaaS can be divided into opportunistic PaaS and participatory PaaS, in this thesis synthetic movement models are divided into two subclasses viz., proactive movement models and passive movement models.

Proactive movement means devices are moving for a particular mission or along designed paths, for example, a rescue team moving in a disaster area and a crowd of participants moving for a crowdsourcing task. Passive movement means users are moving in their everyday lives, governed by the nature of human tendencies such as sociality and periodicity. A proactive movement model is tailored to a specific mission or demand, so it does not need to imitate any mobility trace. However, a passive movement model is designed to imitate real-world movement and mobility traces are the benchmarks.

[13]Please refer to mission-driven movement and demand-driven movement in chapter 5.

2.2.2.1 Passive Movement Model

The Random WayPoint (RWP) mobility model[14] is a well-known benchmark mobility model because of its simplicity. A node randomly chooses a new destination (x, y) within an area and a random speed[14] after pausing for a specific period at the old destination (at the boundary of the simulation area). RWP belongs to epoch based mobility[21]: during a given period called an epoch, a node moves towards a given destination with the same speed. At the start of the simulation, the nodes are uniformly distributed within the area. At the beginning of the simulation, there is high variability in average neighbour percentage which largely affects the performance results of a simulation. Therefore researchers tend to discard a specific time (commonly called **"warm-up time"**) to nullify the simulation initialization problem.

However, [79, 80] found RWP suffers from border effects, which means that the node density is maximum at the centre region, whereas the node density is almost zero around the boundary of simulation area. To overcome border effects, Random Direction mobility model (RD) was proposed. Instead of selecting a random destination, RD selects a random direction. There is no need to reach the boundary before changing the direction. In Levy Walk model (LW)[81], a mobile node can choose a step length, where a step length and the probability of the length follow a power law distribution$f_L(l) \sim l^{-\lambda}$ $1 < \lambda < 3$. Moreover, mobile nodes can also change the direction and speed at a constant interval or a constant distance in Random Walk model (RW).

The aforementioned mobility models are memoryless because they retain no knowledge of past locations and speed values. However, in most case, the current velocity of a mobile node may depend on its previous velocity because of the physical laws of motion. In the Gauss-Markov mobility model (KM)[82], the velocity at t^{th} is $\vec{V_t}$, which is determined by three parameters: $\vec{V_{t-1}}$, memory parameter, and a random variable with uncorrelated

[14] Normal distribution within a range

random Gaussian process with mean 0 and variance σ^2. It has been observed that nodes in real life move at certain preferred speeds $\vec{V_1}$, $\vec{V_2}$,..., $\vec{V_3}$, rather than any distribution ranging from $[0, \vec{V_{max}}]$. Therefore, in Smooth Random mobility model (SR) [83], the time points of speed change are assumed to be a Poisson process and the speed changes smoothly to the target speed from the set. The probability distribution function of acceleration or deceleration $\vec{a_t}$ is uniform.

Mobile nodes in the aforementioned memoryless and time dependent (memorized) mobility models tend to be 2-dimensional uniformly distributed, which is not realistic. Mobility models may have geographical restrictions, for example, road, pathway and buildings. In the **obstacle mobility model**[84], obstacles are utilized to both restrict node movement as well as wireless transmissions. Apart from geographical restrictions, sociality can also add heterogeneity, which is the realistic nature of human mobility, because the movement of human beings is largely governed by the type of community to which they belong.

The Working Day Movement Model (WDMM)[85] has more constraints to mimic the real world. According to [78], this model belongs to the category of time-variant community and home-cell mobility model. WDMM encompasses 4 different sub-models viz, home activity sub-model, office activity submodel, evening activity sub-model, and transport sub-model. These sub-models are repeated every day which represents the periodic and repetitive movement pattern of humans. Each node is assigned a home location, office location and wake-up time, which determines when a node should start from home. At wake time nodes leave their homes, and use different modes of transport like buses and cars to travel to work with transport sub-model (by walk, by bus or by their cars). The home activity sub-model is used for evening and night while office activity is used for daytime.

A node can move inside an office and is assigned a desk. It stays besides its desk for a certain amount of time (Pareto Distributed) and then randomly selects a new coordinate inside the office where it again waits for a certain amount of time. The movement between the desk and randomly selected coordinate repeats until the work day is over. After work, there are two options: going home or going for activities. The evening activity sub-model is used when nodes move (by transport sub-model) in groups to clubs or pubs where they spend a certain amount of time. After this period, they split and use transport model to go back to their home. However, human mobility is not the same everyday if the simulation is more than a week.

The movement model **Point Of Interests (POIs)**, which is used in the experiments in the chapters 4 and 5, also belongs to the community and home cell mobility model. Some people live near the park and some people live in the city centre. Therefore, people in different communities are assigned with various possibilities for different area.

2.2.2.2 Proactive Movement Models

In the Event-Driven and Role-Based mobility model (EDRB) [86], different groups of individuals are attracted (or repelled) towards (or away from) an event depending on their roles. For example, civilians may flee from a disaster-prone area while rescue workers and army personnel may be attracted towards the event to maintain law and order. The mobility model is based on events with horizons and roles with a given time, which can be represented by a triplet as (Civilian, In Event Horizon, Flee) or (Police, In Even Horizon, Approach). This disaster mobility model reflects the nature of disaster event, where the nodes are highly partitioned. History information from mobility traces cannot reflect movement in a disaster scenario. Therefore, synthetic mobility models are necessary for D2D collaborative networks in public safety use cases.

In the Reference Point Group Mobility model (RPGM)[87], the entire node population is divided into some groups, with each group having a leader. The leader can be a logical center or a pre-defined leader node. The movement of the group center completely characterizes the movement of its corresponding group. For each node, mobility is assigned with a reference point that follows the group movement. The motion vector of group member i at time t is $\vec{V}_i^t$, which can be determined by $\vec{V}_i^t = \vec{V}_{group}^t + \vec{RM}_i^t$, where $\vec{V}_{group}^t$ is the reference and $\vec{RM}_i^t$ is independent and identically distributed (i.i.d). With appropriate selection of predefined paths for group leaders and other parameters, the RPGM model can emulate a variety of proactive mobility behaviours. For example, in the case of mobile crowdsensing and battlefield communication, the whole map is divided into several regions if a single group exclusively occupies a single region; if different groups with different tasks travel on the same field in an overlapping manner, this case might be disaster relief. A variant of RPGM with less randomness can be found in the Structured Group Mobility Model[88].

The Column Mobility Model[89] represents a set of mobile nodes (e.g., row of soldiers marching together) that move in a specific fixed direction with a random vector. This mobility model can be used in searching and scanning activities for rescue and detection in places such as mines. The Pursue Mobility Model[89] emulates scenarios where several nodes attempt to capture single mobile node ahead. The node being pursued can move according to any specific mobility model, and other nodes move with acceleration and randomness.

2.2.3 Mobility Prediction

Since nodes in the real-world are not uniformly distributed in an area because of social and geographical constraints, it is worth predicting the mobility and making the best of it. For PaaS, mobility prediction is a precondition of helper selection and resource allocation. The prediction is based on two pieces of information: social information and

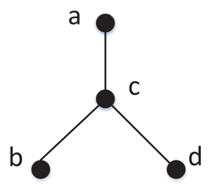


Figure 2.2: Illustration of Centrality

geographical information. Proactive movement is not going to be discussed because the proactive movement is predictable and controllable.

2.2.3.1 Mobility Prediction based on Social Information

This thesis introduces a technology to extract social information from contact history[15] and tries to make the best of that information. Note that, apart from contact history, social information such as common interests and social ties can also be extracted from online social networks.

The basic assumption of social prediction is that nodes with social ties tend to meet frequently and every meeting lasts for a relatively long time. In other words, two nodes are likely to contact again in the future if they contacted frequently in the past. The social information can be extracted from contact history. For example, the contact duration, contact frequency of node A and node B reveal their social strength.

Centrality[16] is the measurement of nodal popularity, which can be calculated in different ways, leading to degree centrality, betweenness centrality, and closeness centrality[2].

- Degree centrality is the number of links. In Fig 2.2, node c is 3 while other nodes are 1. Higher degree centrality means more possible contacts.
- Betweenness centrality measures the number of the shortest paths passing through

[15] So social information can also be called topological information or historical information.
[16] Centrality is a terminology originally used in network theory and is used in sociology

the node. The betweenness centrality of node c is 6 while other nodes are 3 (path has direction). The user who plays a bridge to two groups has a high betweenness centrality.

- Closeness centrality is defined as the inverse of its average shortest distance to all other nodes in a contact graph. The closeness centrality of node c is 1 since its average shortest distance to all other nodes is 1, while the closeness centrality of other nodes are 0.6 since their average shortest path is 5/3. If a node is near to the centre of the graph, it has high closeness centrality.

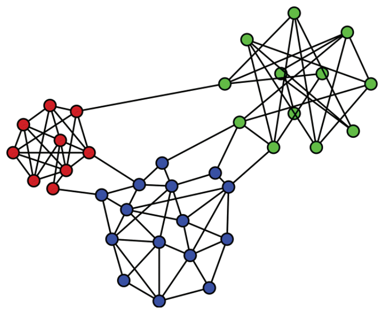


Figure 2.3: Illustration of Community[2]

As aforementioned, people tend to be together as a group. As shown in Fig 2.3, the community is defined as a group of frequently interacting people[2], which is extracted from the contact graph. Community can be detected by various algorithms[90–92] from the contact graph.

2.2.3.2 Mobility Prediction based on Geographical Information

This thesis uses Mobility Prediction based Adaptive Data (MPAD)[3] to illustrate the idea of geographical mobility prediction. There is an intersection between the moving direction and the transmission range of the stationary sink node, as illustrated in Fig 2.4. A closer distance to the stationary sink node indicates a larger communication angle, increasing the contact duration. In this case, distance and moving direction can be adopted as the geometric parameters for candidate helper selection in PaaS.

The Fig 2.4 reveals that:

- Real-time geographic information can be used for mobility prediction. Therefore, geographical information is suitable to make real-time mobility prediction for time-sensitive proximal services because it is easy to predict future contact time.
- To acquire geographic information, the D2D collaborative system needs global information (GPS, distance, and moving direction) and corresponding signalling overhead.

There is another mobility prediction in DTN routing which is innovative for helper selection in PaaS. The mobility prediction named as Time To Intercept (TTI) is proposed in AeroRP[93]. The TTI jointly considers the moving speed, direction and distance of pairwise encountered nodes moving towards the destination. As shown in Fig 2.5, N_i is the node that holds the message for the destination node. When N_i encounters N_j, N_i predicts the mobility of N_j to decide whether to relay the message to N_j. For PaaS, N_d would be the requester and helper selection would be between two helpers (N_j and N_i).

The TTI mobility prediction for N_j is given by $\frac{D_{j,d}-R}{S_j \times \cos\phi_{j,d}}$, where $\phi_{j,d}$ is the relative angle between the moving direction of N_j and the distance $D_{j,d}$ measured from N_j to N_d. If $\frac{D_{i,d}-R}{S_i \times \cos\phi_{i,d}} > \frac{D_{j,d}-R}{S_j \times \cos\phi_{j,d}}$ and $\phi_{i,d} < \frac{\pi}{2}$ and $\phi_{j,d} < \frac{\pi}{2}$, N_j can achieve a faster proximity to N_d.

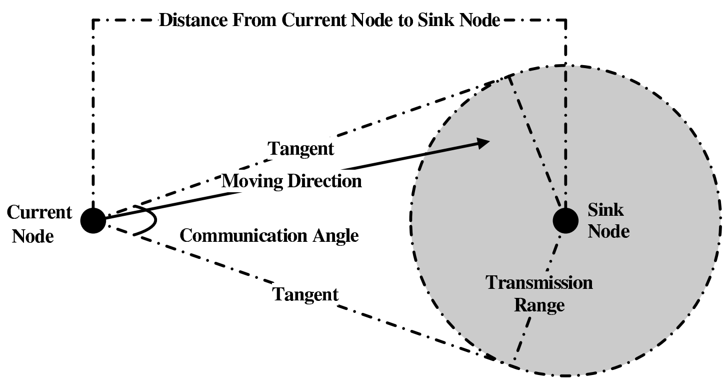


Figure 2.4: Illustration of the Geometric Property in [3]

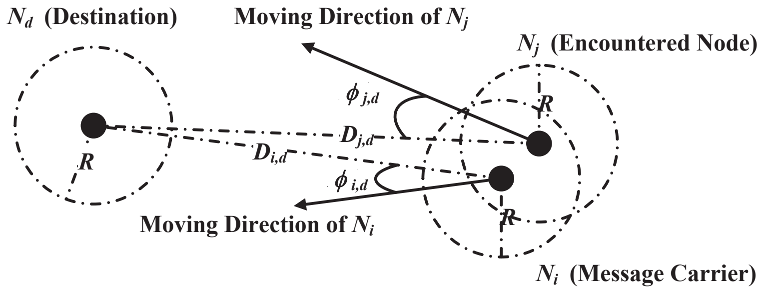


Figure 2.5: Illustration of Nodal Encounter

However, N_i is unable to calculate a negative value of $\frac{D_{i,d}-R}{S_i \times \cos\phi_{i,d}}$ given that $\phi_{i,d} \geq \frac{\pi}{2}$. This implies a case that N_i is moving away from N_d although N_j is moving towards N_d. Therefore, a condition can be added before the comparison, which is $\phi_{i,d} < \frac{\pi}{2}$ and $\phi_{j,d} < \frac{\pi}{2}$.

2.3 D2D Discovery

Because of the mobility nature of mobile devices, D2D devices need to scan and identify their neighbours all the time. For traditional D2D air-interfaces such as Bluetooth, Wi-Fi Direct, and Wi-Fi (ad hoc), there are two existing approaches[10]: location-based technologies and proximity beacon-scan schemes.

Location-based approaches continuously track the users location to determine proximity and use a centralized, cloud-based approach to identify relevancy. For example, a server that is associated with a specific mobile app receives periodic location updates from mobile devices and delivers proximal discovery based on location updates and interests. The advantage of this approach is the unlimited range, but the preconditions are continuous location tracking to access the cloud and signal and energy overhead.

The proximity beacon-scan scheme needs to broadcast relevant information to nearby users through an associated mobile app. Proximal discovery services based on D2D technologies that operate entirely in unlicensed spectrum do not scale well because of uncontrolled interference, a lack of synchronization between devices and the collision-avoidance scheme. Despite these limitations, a proximity beacon is an excellent approach for the requesters.

However, these two existing approaches may still cause mobile app isolation due to proprietary platforms, which need to be integrated into cellular networks with a common language for discovery.

3GPP proposed innovative D2D discovery techniques in the licensed spectrum which can discover approximately 1000s of devices in 500-meter proximity and determine relevancy at the device level without user/app intervention (interoperable discovery across apps, operating systems, devices, and operators).

3GPP does not restrict the algorithms of resource allocation as well as user authentication in the D2D discovery and, on the contrary, it leverages research in this field. For in-band D2D discovery, resource multiplexing between discovery and cellular traffic can be categorized as three ways in [94]:

- **Time:** Time multiplexing is not preferred because of the spreading of the signal in frequency which lowers the power efficiency and hence reduces transmit range.
- **Frequency:** Frequency multiplexing uses narrower bandwidth and hence achieves longer range.
- **Code:** A discovery signal for multiple devices is required and cannot be power controlled, so the near-far effect is inevitable.

The chipset LTE-Direct, as one of the implementations of in-band D2D,[17] uses the uplink resources in an LTE FDD system (which is in-band underlying method) and dedicated frames in an LTE TDD system (which is the in-band dedicated method). It sets aside a small percentage of sub-frames for efficient discovery. All LTE Direct-enabled devices wake-up synchronously during these sub-frames and either broadcast beacons or listen for beacons.

2.4 Incentive Mechanism

As aforementioned, the human resource is one of the common features of PaaS systems, so researches of incentive mechanisms is a never-old-fashion topic. Moreover, every resource has a price, which can also be studied in the paradigm of incentive mechanisms. An incentive mechanism is to motivate users to exchange their resources. If a D2D collaborative system is a metaphor for human society, the incentive mechanism of this system should be an economic policy.

The aims of the incentive mechanism for PaaS are:

- **Improving services of each user**
- **Improving the performance of the whole system**
- **Reaching a tradeoff between the rewards paid for cooperative users and the revenue of MNOs**

If selfishness is assumed as the nature of participants in the system, rewards are necessary. There are two kinds of rewards: reputation rewards and monetary rewards. The difference between reputation rewards and monetary rewards is that, in monetary reward cooperation, the individual who provides services expects a clear and instant

[17] LTE-Direct has been developed by Qualcomm following the specification defined in 3GPP

benefit, like a business contract[95]. In a reputation rewards scenario, the individual spending resources with a long-term and unclear expectation, just like students spending time and money on parties and friendship. Moreover, the reputation rewards are not as general as the monetary rewards since reputation rewards may only be useful in a particular community.

Not only selfishness, altruistic and cooperative behaviour are also basic human nature. Sometimes rewards are not necessary. For a collaborative D2D system within family members and close friends, participants can be assumed as altruistic.

Moreover, punishment can also be applied in the D2D system to boost the participation rate. For example, in [58], selfish users are sidelined by others.

In this chapter, the exploration of proximal services, the movement model, the D2D discovery and the incentive mechanism are reviewed. This chapter builds a background to this research. The review of exploration of proximal services will be continued in the future work.

Chapter 3

Introduction to the Experimental Methodology

The experimental methodology includes the simulator implementation and DOE (experimental design and data analysis).

3.1 Introduction to the Simulator

The ONE (Opportunistic Network Environment simulator) is a discrete event simulation engine developed at Aalto University and is now maintained and extended in Github[96]. It is designed[1] to simulate communication between devices with mobility, especially for opportunistic communication with intermittent connections.

Compared to other well-known network simulators such as NS3 and OPNET, ONE simplifies the MAC layer and the physical layer into transmission rate, transmission range (transmission time is calculated by consideration of transmission range and mobility) and corresponding energy consumption. So adding a new protocol in ONE requires

[1]https://akeranen.github.io/the-one/

comparatively less work. Therefore, researchers are not constrained by legacy technologies, standards, frameworks, and protocols.

ONE can generate node movement using different movement models or import mobility data from real-world traces or other mobility generators, which constrains the node movement to predetermined paths. The Graphical User Interface (GUI) displays a visualization of the simulation state showing the locations, active contacts, and messages carried by the nodes.

3.2 Introduction to the Implementation of Helper Selection

The helper selection module in PaaS inherits the routing module from ONE and three functions were added to enable helper selection for PaaS.

3.2.1 Requester Select Helpers

When a base station receives an access request, the base station selects a set of helpers to serve the requester. The helper selection scheme sorts all the helpers based on the algorithm illustrated in chapter 4 and picks a particular number[2] of helpers to form the unique target helper set for this requester.

3.2.2 Helper Serve Requesters

The helper selects the requesters based on urgency as follows:

- Sort all the current surrounding requesters within the transmit range of this helper based on urgency.

[2]The number of helpers is an input to the experiment.

- Selected a certain number [3] of requesters to serve.
- Serve the requester until the contact ends

3.2.3 Information Update during Contact

There is a class taking care of contact in ONE, which is called “Connection”. When two nodes contact, the “Connection states” between them is updated.

For PaaS, there is no update for the same type of devices, which means only the contact between helpers and requesters counts. Metrics[4] that are updated are as follows: inter-meeting time, contact counts, average inter-meeting time and utility. When the connection is down, contact duration is updated.

3.3 Improvements to the Simulator

This section introduces improvements to the simulator and contributions to the community during the research.

3.3.1 Improvement to the Random Generator in the Simulator

The meaning of randomness is to simulate the uncontrollable factors of an experiment in the real world. Computer simulations usually use a pseudorandom sequence which are generated by a random generator. Random seeds determine the starting point of the pseudorandom sequence.

[3]Helper access limit is an input to the experiment.
[4]These metrics are used in the utility function of the helper selection mechanism, which is elaborated in chapter 4.

In this thesis, the randomness of this simulation comes from two places: random movement and random traffic generation. Therefore, the simulation of this thesis can also be called a two stream simulation[97] because there are two random streams.

In the original simulator, the random seeds are set manually in the configuration file. The configuration file is invoked before each run of the simulation, which is not convenient for massive runs. Moreover, it is inevitable that the manually configured random seeds would be repeated. For example, 10000 runs need 10000 seeds because the uncontrollable factors of the experiment in the real world never repeat, but it would be too laborious to set the random seeds 10000 times.

Therefore, some core codes of the simulator have been re-written, so that the random seeds of the next run are consecutive numbers from the pseudo random sequence of the previous run.

The simulator is written by Java and the Random class generates the same sequence when the seeds are the same.[5] The maximum period of the pseudo random sequence is about 2^{48}, which is 2.814e+14 (about 3 hundred thousand billion).

For the use of random traffic generation, the simulation time is 960 minutes and the access requests are generated with the mean of 1 minute in requesters. Assume that there are 10 inputs, each with 3 levels, the full factorial combination would be 3^{10}. Every combination is assumed to have 30 repetitions. So the number of runs is $960 * 3^{10} * 30 = 1.7 * 10^{10}$, which is much smaller than 3 hundred thousand billion.

Another randomness stream is used in the mobility model. Random numbers are generated for the speed at the beginning of each path, and for the waiting time at the

[5] It means that for the random generator in Java , the seed is the only thing that is configurable.

ending of every path. Assume that, on average, 1 second for a path and 1 second for a waiting time, which is too active for walking speed and city map scenario, would be a strong assumption in the city movement pattern for all the people because it is. For this assumption, the random numbers needed are $960 * 60 * 3^{10} * 30$, less than 10^{11}.

So for both randomness streams, the pseudorandom sequence is long enough. Based on this fact, this research implemented consecutive random seeds by a static "reset" function that can be executed every time when the class is constructed. The simulator now can start the next run at the point at which the previous one stopped (i.e. with those seed values). It is a proper way to approach massive runs with different seeds for each run. This change has been submitted and contributed to the community.

3.3.2 Batch Reading of Results

In the original version of the simulator, after the end of every run, results are written by the Report class into a file. There will be many files with the increase of the number of runs, which means the ONE simulator is not suitable to do large scale performance analysis, especially Design Of Experiment. The simulator was improved to do batch reading file by file and the results can be stored into a CSV file. This change has also been submitted to the community.

3.3.3 Auto-Configuration

In ONE, multiple runs are done in batch mode. The core of the simulator reads the configuration file and then begins the simulation. The number of runs must be pre-set in the command line before the beginning of the simulation. Therefore, the simulator cannot traverse all the possible combinations of inputs with corresponding repetitions of each combination.[6]

[6]In DOE, different combination of inputs may have different repetitions. Each repetition is a single run.

To implement writing the stopping points (new seeds for next runs) to the next run and traversing all the combination of inputs by one long run with many sub-runs, the configuration file should be re-written after every run and there should be multiple layers of loops. For example, if 10 inputs are considered, there should be 10 layers of loops.

3.3.3.1 Process Control

A parameter "auto" is added in the main function, so the choices of the process control of the simulator have default (GUI) mode, batch mode and automatic mode. The number of parameters of the main function reaches 4. The first one is "-b" (batch mode); second one is "nrofRuns" (number of runs for all non-auto modes); the third one is "-auto" (automatic); the forth parameter is the name of configuration files (maybe more than one name).

3.3.3.2 Reading the Configuration File

To traverse all the possible combinations of the inputs, the total layers of the loop might be unknown. It can be implemented by an iteration function or a multiple dimensional array.

Along with the iteration function or the array, all the combinations of inputs in the configuration file need to be set and read. Moreover, the selected combinations can also be efficiently chosen. This change contributes to the community of ONE since this function increases productivity and enables DOE.

3.4 Introduction to Design Of Experiments

Design Of Experiments (DOE)[34] is a statistical and scientific experimental methodology that can rigorously and economically conduct an experiment, extract full information of the relationship between control factors (inputs) and responses (outputs). DOE includes choice of influential factors and responses, designing and conducting the experiment, and experimental data analysis.

In other words, the goal of experimental design is to find the representative points of input factors (how many levels and the values of the levels), which needs the expertise of a certain field and pilot runs. After conducting the experiments, the next step is to find the polynomial regression equations including the main effects of factors and interactions between factors.

To avoid needless repetition of time consuming experiments and extract full information from the experiments, DOE was applied in this research. It can also be called design of anything, which means the design of any task that aims to explore the relation between one or more input variables and output variables. This thesis briefly introduces DOE by a representative simple method: full factorial design.

3.4.1 Introduction to Two-Level Full Factorial Design

For example, there are three input variables and one output variable, $O = f(X, Y, Z)$. In this experiment, each input variable has two levels (for example, X_0 for lower level and X_1 for higher level), and each run has to be repeated four times (based on variances of responses from pilot runs). The simplest experimental design is changing one input variable (X) at a time and controlling the rest $O = f(X, Y_0, Z_0)$, namely one-factor-at-a-time or control variates approach. After eight runs, knowledge of the relationship between O and X is only based on particular combinations of Y_0 and Z_0. Even after 24 runs, the knowledge of the interaction among all the input variables is still unknown.

The experimenters might have confidence of empirical insight into the nature of the experiment. However, to think further, all the combinations of the input factors should be $2^3 = 8$, but there are only $24/4 = 6$ combinations to be observed. The two missing combinations could more or less provide hints about interaction effects of the three input variables. Assuming that the system has a linear response,[7] the response can be expressed as Equation 3.1. In order to solve this expression (eight unknowns), it needs eight equations (eight combinations). Moreover, the information lost could be even worse if there are more variables and each variable has more levels.

$$O = a_0 + a_1X + a_2Y + a_3Z + a_4XY + a_5XZ + a_6YZ + a_7XYZ \tag{3.1}$$

To tackle this problem, a British statistician and biologist named Ronald A.Fisher[98] proposed an idea that **all the input variables can be changed at the same time**, which is called full factorial design. For the same experiment $O = f(X, Y, Z)$, without loss of information, the runs could include "full combinations" ($2^3 = 8$) in full factorial experiment. If it is not the full factorial design, for example, fractional factorial or Taguchi method[34], the number of runs could be even less [97].

Moreover, it is common that the dimensions of inputs (state space) of simulations are more than three and each variable always has several levels. Therefore the aforementioned information loss is inevitable without DOE in simulations.

Table 3-A: Full Factorial Experiment

Order of Run	X	Y	Z	O
1	X_0	Y_0	Z_0	O_0
2	X_1	Y_0	Z_0	O_1
3	X_0	Y_1	Z_0	O_2
4	X_1	Y_1	Z_0	O_3
5	X_0	Y_0	Z_1	O_4
6	X_1	Y_0	Z_1	O_5
7	X_0	Y_1	Z_1	O_6
8	X_1	Y_1	Z_1	O_7

As shown in Table 3-A, the Gaussian elimination algorithm[99] can be used in solving the linear equations. Particularly, in this case, since the change of parameters of the linear equations is simple (X_0, X_1, Y_0, Y_1, Z_0 and Z_1), in practice, the equations can be

[7] Otherwise it would not be a two level experiment

solved easily. For example, when Z_0 changes to Z_1, there are 4 responses O_4, O_5, O_6, O_7 corresponding 4 combinations of other two factors. Intuitively thinking, the main effect of Z (a_3 in Equation 3.1) could be calculated as Equation 3.2. Similarly, a_1, a_2,...,a_7 can be calculated this way[98].

$$Z_m = (O_4 + O_5 + O_6 + O_7)/4 - (O_0 + O_1 + O_2 + O_3)/4 \tag{3.2}$$

3.4.2 Data Analysis

After designing and conducting an experiment, raw data is generated, which needs to be analysed so that the expressions (relationship) between responses and factors with a certain resolution[34] can be found.

3.4.2.1 Introduction to the Resolution of Experiment

The experiments would have more and more complicated interactions between factors (higher orders of interaction) with the increase of the number of input factors and the number of levels. The experiment based on full factorial design can provide full information to find all the interactions, which means the highest resolution. However, in the data analysis, it is not always necessary to find the interactions with the highest resolution since the interactions might not be influential to responses[34]. The concept of resolution is introduced as follows.

For example, for 6 two-level factors, full factorial design needs 64 runs. Confounding effects[8] would be generated if the number of runs is reduced from 64 to 8 for fractional factorial designs. For the experiment with 6 two-level factors, as shown in Table 3-B, 8-run design can reach the resolution of III.

[8] In statistics, a confounding factor is a variable that influences both the dependent variable and independent variable. In DOE, confounding effects means the value of a main effect estimate comes from both the main effect itself and also contamination or bias from higher order interactions.

- **Resolution III:** main effects are confounded (aliased) with two-factor interactions, but no main effect is aliased with any other main effects.
- **Resolution IV:** some 2-factor interactions are confounded with other 2-factor interactions and main effects are confounded with 3-factor interactions, but no main effect is confounded with other main effects nor 2-factor interactions.
- **Resolution V:** 2-factor interactions are confounded with 3-factor interactions and main effects are confounded with 4-factor interactions, but no main effect or 2-factor interaction are confounded with other main effects or 2-factor interactions. Resolution V and resolutions above V are green because the high-order confounded interactions can be neglected (regarded as full resolution). [9]

Table 3-B: Resolution of DOE

	Factors													
Runs	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	Full	III												
8		Full	IV	III	III	III								
16			Full	V	IV	IV	IV	III	III	III	III	III	III	III
32				Full	VI	IV	IV	IV	IV	IV	IV	IV	IV	IV
64					Full	VII	VI	V	IV	IV	IV	IV	IV	IV
128						Full	VIII	VI	V	V	IV	IV	IV	IV

3.4.2.2 Introduction of Regression

The regression analysis can reveal how[10] the factors affect the performance. In the context of regression analysis, inputs factors can be called predictors or regression variables and the outputs can be called the response. Here the regression analysis is simply introduced by two concepts.

"R": R, namely R-square, is a metric of how close the raw data obtained from the experiment are to the regression line. If R is equal to 1, all the predicted values are equal to observed values, which means the regression model explains all the variability of the results of the experiment.

[9]Refer to Minitab (a DOE software) Support for the table and explanation of resolution III, IV, V.
[10]The levels of the factors (linearly, quadratically or cubically).

"P" Null hypothesis of hypothesis test in regression is that the coefficient of a factor (Coef, the slope of the regression line) is equal to zero (no effect). A low p-value (for example, smaller than 0.05) indicates rejecting the null hypothesis. In other words, a factor that has a low p-value is likely to be meaningful because changes in the value of the factor are related to changes in the response variable.

3.5 Summary

In this chapter, we introduce how to do experiments and how to do data analysis. Experiments and data analysis in the next two chapters are based on this foundation.

Chapter 4

PaaS for the Use Case of Access Enhancement

As mentioned in chapter 1, this thesis proposes a D2D collaborative framework PaaS which decouples devices as helpers and requesters in order to focus on the proximal services. Here a simple proximal service for access enhancement is proposed with demands of service continuity via opportunistic D2D communication. User Equipments (UEs) are treated as a part of the network (known as helpers) for service provisioning.

Key techniques here involve necessary knowledge awareness of human mobility, efficient D2D handling for specific services requirements, and consideration of potential security issues (a certain number of appropriate helpers) etc. Among these, the focuses of this thesis are mainly the following issues:

- **How to explore human mobility, so as to identify a certain number of appropriate helpers (in terms of whom to help) for PaaS?** This thesis formulates the helpers selection problem and proposes the ContAct based Proximity (CAP) algorithm, which is a pun for "CAP" with the meaning of covering the demands of UEs ubiquitously (and ideally seamlessly). Fruitful contact history information such as contact duration, contact frequency and inter-meeting duration is captured for human mobility prediction.

- **What are the key factors (inputs) influencing the success ratio and continuity of PaaS (outputs)?** Here, the continuity refers to the least disruption (in terms of frequency and time duration) experienced during a service. Inevitably, D2D connections suffer disruption by their nature[100], because of human mobility and limitation of device transmission range, e.g., WiFi-Direct is generally with 50 meters[101] transmission range. According to the previous work[15], edge computing such as TCP proxy implemented in service providers can enable services in disturbed wireless environments. However, with even more challenging conditions, e.g., a 5 hours' disruption, the service would normally fall into the Delay/Disruption Tolerant Networking [35, 102, 103] paradigm.

4.1 System Model of PaaS for the Use Case of Access Enhancement

This research considered a system with a certain number of Base Stations (BSs), User Equipments(UEs) that play the role of service providers, namely helpers, and UEs that request services from helpers by D2D communication, namely requesters. The helpers are assumed to be willing to help other requesters access the network by D2D communication. The helpers are acting as "mobile access points" or "mobile picocells" with wireless backhaul links[67].

The whole communication process has two hops but only one hop of D2D communication: the first hop is between the requester to the helper; the second hop is D2D communication between the helper to the BS. Helpers are assumed to have stable connections with infrastructure. Nevertheless, the opportunistic contact between a pairwise helper and requester inevitably results in service disruption.

4.1.1 An Example of CAP

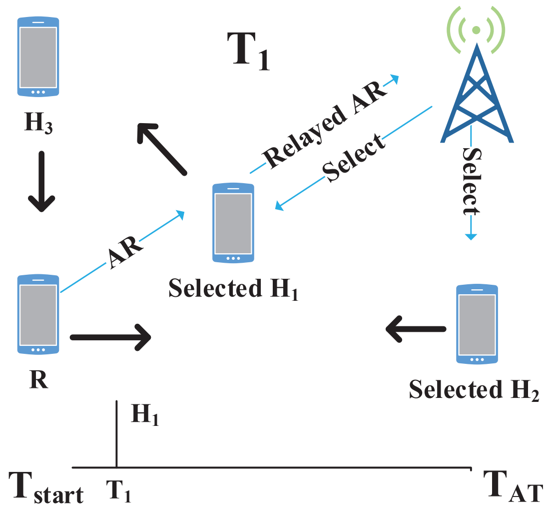


Figure 4.1: Process of D2D Relaying in Time 1

We assume that there are three helpers, one requester and one base station in the system, here is an example of the use case, Fig. 4.1. The requester is assumed to generate access requests deterministically (e.g., generate access request every one minute) with the same requested access duration (T_{RAD}, e.g., half an hour). The access request needs to be served within a period of time, called the access tolerance (T_{AT}, e.g.,2 hours). The time clip of the process of CAP in Fig. 4.1 begins at the start of an access request ($T_{start} = 0$) and finishes at the end of the access tolerance (T_{AT}) of the access request. The moments T_1, T_2 and T_4 are shown to elaborate the access enhancement process.

- At T_1, an access request with T_{RAD} and T_{AT} was generated from a requester. The requester, running D2D air interfaces, broadcasts an access request (AR) to nearby devices periodically through an associated mobile application. As illustrated in Fig. 4.1, when helper 1 (H_1) received the AR and relayed the AR to the base station. The base station then arranged and selected appropriate helpers (H_1 and H_2) for this requester based on metrics (utilities) derived from their contact history.

- At T_2, all the UEs had moved to different locations, as illustrated in Fig. 4.2.

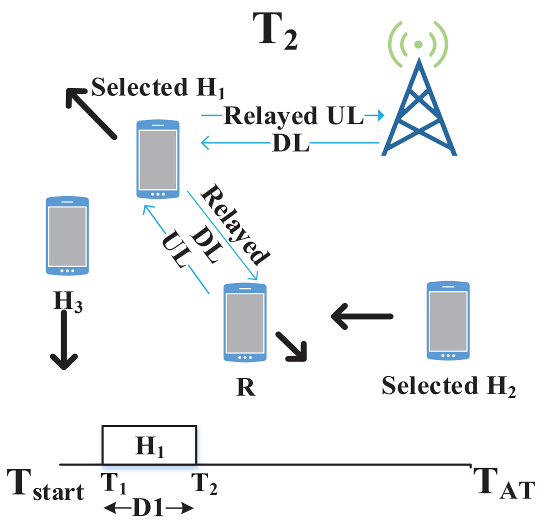


Figure 4.2: Process of D2D Relaying in Time 2

H_1 had helped the requester (R) access the network for a period of duration D_1 ($D_1 < T_{RAD}$). However, H_1 was going to leave R. Noted that "UL" means uplink and "DL" means downlink.

- In Fig. 4.3 taken at T_4, H_2 began to help R at T_3 and finished the service at T_4 ($D_1 + D_2 = T_{RAD}$), which is within T_{AT}.

The system is assumed to be based on one-to-many D2D, which means that a helper can have connections with several requesters but a requester can only communicate with one helper.

4.1.2 Problem Formulation of PaaS for the Use Case of Access Enhancement

It is true that not all helpers will be allocated with the role to provide access service to requesters. Therefore, how to optimally select a certain number of helpers will be of interest. An accurate and concise mathematical description of the helper selection problem is the first step. Note that the notations for problem formulation are shown in Table 4-A.

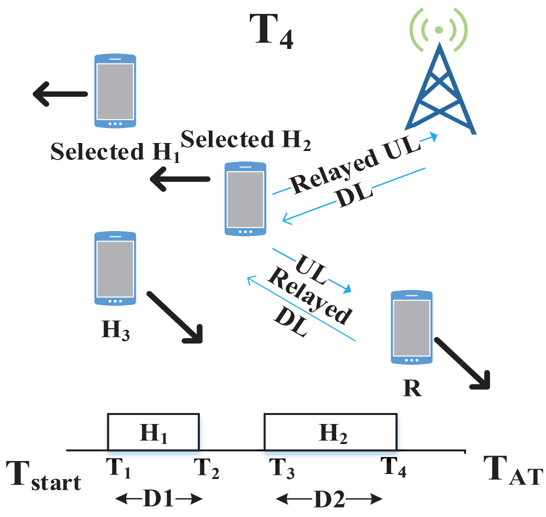


Figure 4.3: Process of D2D Relaying at Time 4

Equation (4.1) shows that at any time, the contact relationship between H_i and R_j can be contacted or not contacted.

Table 4-A: List of Notations for Problem Formulation

T_{AT}	Access Tolerance
T_{RAD}	Requested Access Duration
R_j	Requester j
H_i	Helper i
N_H	Total number of helpers in the system
N_R	Total number of requesters in the system
AR_{jk}	The k th access request of requester j
$T_{AR_{jk}}$	Start time of the k th access request of requester j
$SH_{AR_{jk}}$	Set of selected helper for AR_{jk}
M	limit number of the set $SH_{AR_{jk}}$
S_{AR}	Total number of all the generated access requests
SC_{AR}	Total number of the completed access requests
K	Total number of requester a help can serve at the same time
T_{jk}^{ela}	Elapsed time since AR_{jk} requested
$C()$	If the service is completed, count 1.
$Card()$	Cardinal number of a set

$$H_i(t)R_j(t) = \begin{cases} 1 & \text{if contacted} \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

When base stations receive an access request AR_{jk}, an optimal set of helpers (e.g., in terms of minimized helpers number) is selected as $SH_{AR_{jk}}$ to serve this AR_{jk}. If H_i is in the selected set of helpers $SH_{AR_{jk}}$, the accumulated contact duration between H_i and R_j as denoted by $T(H_iAR_{jk})$, within T_{AT} of AR_{jk}, is shown in Equation (4.2).

$$T(H_iAR_{jk}) = \int_{T_{AR_{jk}}}^{T_{AT}+T_{AR_{jk}}} H_i(t)R_j(t)\,dt. \tag{4.2}$$

The results of the service contributed by the selected helpers ($SH_{AR_{jk}}$) are: completed or not completed. As shown in Equation (4.3), the binary value "1" represents completed, which means that the sum of accumulated contact duration between the selected helpers and the requester R_j is not less than T_{RAD}, while the binary value "0" represents not completed.

$$C(AR_{jk}) = \begin{cases} 1 & \text{if } \sum_{i}^{SH_{AR_{jk}}} T(H_iAR_{jk}) \geqslant T_{RAD} \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

The helper selection strategy can be formulated as an optimization problem. It consists of the objective function (maximizing total number of the completed access requests SC_{AR}), by choosing subsets (selected helpers $SH_{AR_{jk}}$) for all access requests (with the total number S_{AR}), from all helpers in the system. This problem is subjected to two limitations. Firstly, a helper can only serve limited requesters simultaneously; secondly, the number of helpers serving a requester is limited. In other words, the cardinality of the set $SH_{AR_{jk}}$, namely $Card(SH_{AR_{jk}})$, is limited.

$$\max \quad SC_{AR} = \sum_{i=1}^{S_{AR}} C(AR_{jk}) \tag{4.4a}$$

$$\begin{aligned} s.t. \quad & \sum_{j=1}^{N_R} H_i(t)R_j(t) \leqslant K, \quad i = 1, 2, 3 \ldots, N_H \\ & Card(SH_{AR_{jk}}) \leqslant M, \quad j = 1, 2, 3 \ldots, N_R \end{aligned} \tag{4.4b}$$

4.1.3 Discussion on the System

The mobility of the UEs is not prior knowledge, so the information of Equation (4.1) is unknown before the experiment. Without this information, it is hard to estimate the results of the access requests. For example, AR_{j1} denotes the first access request from R_j. If AR_{j1} is completed, it is also hard to estimate the service completion time. A requester can only have one access request at the same time, which means that the AR_{j2} can only be generated after the finishing time of AR_{j1}. In other words, the more access requests completed, the more access requests generated. So the total number of access requests (S_{AR}) is also unknown.

Moreover, when R_j is temporarily isolated, AR_{jk} from R_j is hard to be completed, even if all the encountered helpers serve it, as shown in Equation (4.5). This situation is also unpredictable.

$$\sum_{i=1}^{N_H} T(H_i AR_{jk}) < T_{RAD}. \tag{4.5}$$

Based on the difficulties brought by the unpredictable mobility of the nodes, this optimization problem is hard to solve. Therefore, the heuristic algorithm proposed in the next section which can learn the information of contact history is a feasible approach. The helper can also raise the priority of requesters based on elapsed time T_{jk}^{ela} if the number of concurrent devices is more than K.

4.2 The Algorithm for Helper Selection

The idea of this algorithm is learning the regularity of mobility from contact history, and then, proposing a utility for every helper and requester pair. The utilities are the scores to rank the appropriate helpers for a requester.

Table 4-B: List of Notations for Algorithm

$D_{i,j}$	Historically contact duration between H_i and R_j
$T_{i,j}$	Historically inter-meeting time between H_i and R_j
$C_{i,j}$	Historically contact count between H_i and R_j
$U_{i,j}$	Utility value estimated of H_i and R_j
$UA_{i,j}$	Utility for helper selection, A means average
Γ	aging constant
$N_{i,j}$	Total number of contacts between H_i and R_j

4.2.1 Utility Function Design

The contact history between a helper and a requester is elaborated in Fig. 4.4. $T_{i,j}^{(C_{i,j}=1)}$ means the moment when H_i and R_j contact for the first time. The inter meeting time shown in Fig. 4.4 is $T_{i,j}^{(C_{i,j}=2)}$. For $C_{i,j} \geqslant 2$, $T_{i,j}^{(C_{i,j})} - D_{i,j}^{(C_{i,j}-1)}$ means the encounter gap, namely disruption duration. Note that the $C_{i,j}$ can only be captured when a new contact happens.

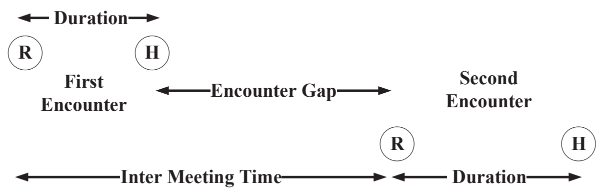


Figure 4.4: Illustration of Contact History

Based on the information obtained from contact history between a pairwise helper and requester, the regularity can be found. For example, there is one requester (R) and two helpers (H_1 and H_2) in the system. H_1 contacts with R frequently, with longer contact duration while with less disruption duration. H_2 rarely contacts R, with shorter contact duration while with longer disruption duration, on average. It is easy to learn that H_1 will be a better choice for R. Moreover, $T_{H_i,R_j}^{(C_{i,j}=1)}$ also implies how fast R_j can get connected to the network. Based on this observation, an empirical utility function can be obtained, as shown in Equation (4.6).

$$U_{i,j}^{'} = \frac{T_{i,j}^{(C_{i,j}=1)} + \sum_{i,j}^{N_{i,j}} (C_{i,j}=2) \left(T_{i,j}^{(C_{i,j})} - D_{i,j}^{(C_{i,j}-1)} \right)}{N_{i,j}} \tag{4.6}$$

Take the above as an example, there is one requester (R_1) and two helpers (H_1 and H_2) in the system. Assuming $T_{1,1}^{(C_{1,1}=1)} = 15$, $D_{1,1}^{(C_{1,1}=1)} = 3$ at the first contact, $T_{1,1}^{(C_{1,1}=2)} = 10$ and $D_{1,1}^{(C_{1,1}=2)} = 6$ at the second contact, while $T_{1,1}^{(C_{1,1}=3)} = 20$ is recorded at the third contact, then $U'_{1,1}$ is calculated as:

$$U'_{1,1} = \frac{15 + (10 - 3) + (20 - 6)}{3} = 12 \tag{4.7}$$

Assuming $T_{2,1}^{(C_{2,1}=1)} = 20$, $D_{2,1}^{(C_{2,1}=1)} = 2$ at the first contact, while $T_{2,1}^{(C_{1,1}=2)} = 20$ is recorded at the third contact, then $U'_{2,1}$ is calculated as:

$$U'_{2,1} = \frac{20 + (20 - 2)}{2} = 19 \tag{4.8}$$

It is conventional that a helper with a larger value of utility has the potential to provide good quality of service, in terms of success ratio and continuity. So the utility should be inverted as Equation 4.9.

$$U_{i,j} = \frac{1}{U'_{i,j}} = \frac{N_{i,j}}{T_{i,j}^{(C_{i,j}=1)} + \sum_{i,j}^{N_{i,j}} (C_{i,j} = 2) \left(T_{i,j}^{(C_{i,j})} - D_{i,j}^{(C_{i,j}-1)} \right)} \tag{4.9}$$

Note that the utility is updated at the beginning of the contact. Contact duration is recorded when the contact is disrupted, while inter-meeting is recorded at the beginning of the contact, similar to $C_{i,j}$, as shown in Algorithm 1.

Algorithm 1 Update Utility

```
if A contact between R_j and H_i is started then
    H_i and R_j update the number of contact C_{i,j}
    H_i and R_j update the inter-meeting time T_{i,j}
    H_i and R_j update the utility U_{i,j}
    H_i reports the information to base stations
end if
if The contact between R_j and H_i is ended then
    H_i and R_j updates the contact duration
    H_i reports the information to base stations
end if
```

The relationship between nodes is relatively steady while the contact may fluctuate, so $U_{i,j}$ will not accurately reflect the proximity (in terms of how often and how continuous) between i and j. This research used $UA_{i,j}$ as the utility to make helper selection. $UA_{i,j}$ is updated by using Exponential Window Moving Average (EWMA) for every connection, as shown in Equation (4.10), where Γ is the weight constant.

$$UA_{i,j}{}^{(new)} = UA_{i,j}{}^{(previous)} \times \Gamma + U_{i,j} \times (1 - \Gamma) \tag{4.10}$$

4.2.2 Access Handling

When base stations receive an access request from a requester relayed by a helper, all the helpers would be sorted based on utilities and the best M helpers would be selected ($SH_{AR_{jk}}$).

A helper can only serve a limited number (K) of requesters simultaneously. If base stations arrange too many devices for this helper, the most urgent requesters would be chosen to serve based on T_{jk}^{ela}.

4.3 Performance Evaluation

Here the evaluation is based on the medium Helsinki city scenario in Opportunistic Network Emulator (ONE)[96], as shown in Fig. 4.5. This research deployed 4 types of interest points, namely Point Of Interests (POIs)[1], on this map with the consideration of the regularity of movement. For example, nodes in area-4 would randomly move to one of the 22 points by map-based shortest path movement pattern (Dijkstra's algorithm).

[1] Please refer to the section of passive mobility model in chapter 2

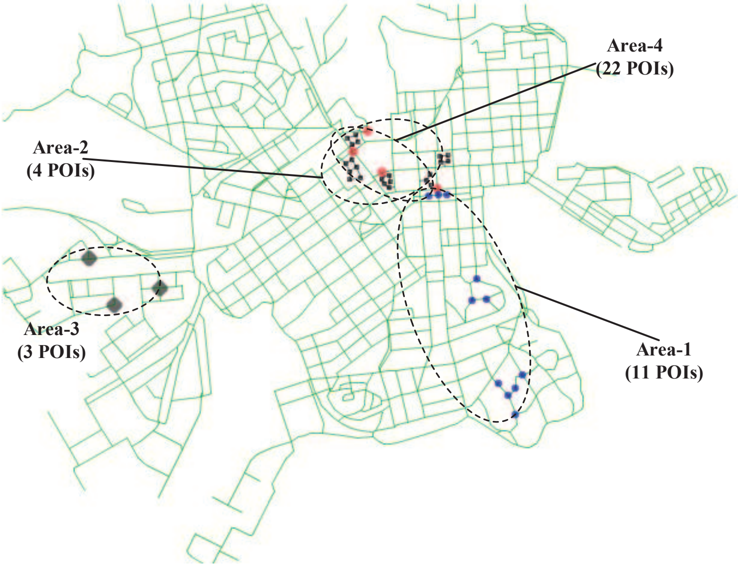


Figure 4.5: Medium Helsinki City Scenario with POI

As shown in TABLE 4-C, there are four groups of requesters and 4 groups of helpers which have different probabilities of being in four areas, namely mobility patterns. The requesters of the first group (R1) spend the majority of time in Area-1 while requesters in group 2 (R2) and group 3 (R3) spend more time in Area-2 and Area-3 respectively. Requesters in group 4 (R4) spend most of the time to move in Area-4 compared to requesters in other groups. Similarly, one hundred helpers also move within 4 POIs with four mobility patterns.

Table 4-C: Probabilities of Movement in Four Areas

Group	Number	Area-1	Area-2	Area-3	Area-4
R1	25	0.7	0.1	0.1	0.1
R2	25	0.1	0.7	0.1	0.1
R3	25	0.1	0.1	0.7	0.1
R4	25	0.1	0.1	0.1	0.7
H1	25	0.7	0.1	0.1	0.1
H2	25	0.1	0.7	0.1	0.1
H3	25	0.1	0.1	0.7	0.1
H4	25	0.1	0.1	0.1	0.7

The D2D air interface is configured as WiFi-Direct (out-band D2D[104]), with 50 m range[101]. This research assumed all the users in the system are moving with speed varying between [0.5∼1.5] m/s, which is the speed range for normal human walking. The one hundred requesters generate network access requests after 10800 s warm-up time[2], so the regularity of the mobility can be learned by base stations. Without loss of generality, the requesters are assumed busy, which means they generate new requests with an interval of 60 s[3] after successful access service or after T_{AT} (access tolerance). All the requests are assumed to require 1800 s (half an hour)[4] of access duration. The simulation time is set as 57600 s (16 hours).[5]

Here this research compared the helper selection algorithm CAP with the "Random Helper Selection" (which means that the M helpers serving requesters are randomly selected from 100 helpers). The algorithm is evaluated by T_{AT} (access tolerance) on:

- **Success Ratio:** The percentage of completed service.
- **Average Disruption Frequency:** The average number of disruptions during each access service.
- **Average Disruption Duration:** The average disruption duration of each access request.

This thesis also evaluated the success ratio through the influence of M (the number of selected helpers) and K (helper access limit).

4.3.1 Differences between CAP and RANDOM on Access Tolerance

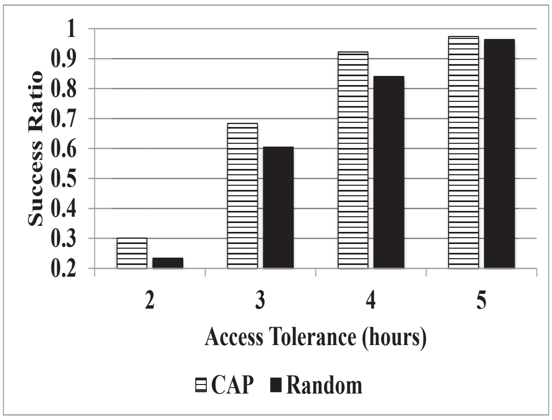
The access tolerance T_{AT} is set as 2 hours, 3 hours, 4 hours and 5 hours respectively, with transmit range of 50 m, 20 selected helpers ($M = 20$) and an access limit of 5 concurrent requesters ($K = 5$). As shown in Fig. 4.6(a), the success ratio of CAP

[2] In this scenario, the mobility pattern usually starts to be steady after 1 hour (statistically). So the warm-up time is set to be 3 hours to avoid any accident.

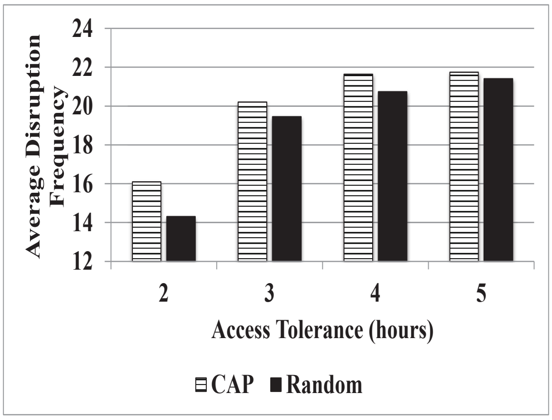
[3] The interval can be constant as 60 seconds. The interval can also be exponentially distributed with the average of 60 seconds. However, results show that these two cases are almost the same. Please refer to pilot runs of next chapter.

[4] Here the "application" has a service requirement with half an hour duration within an access tolerance. The access tolerance would be the input factors in the following sections.

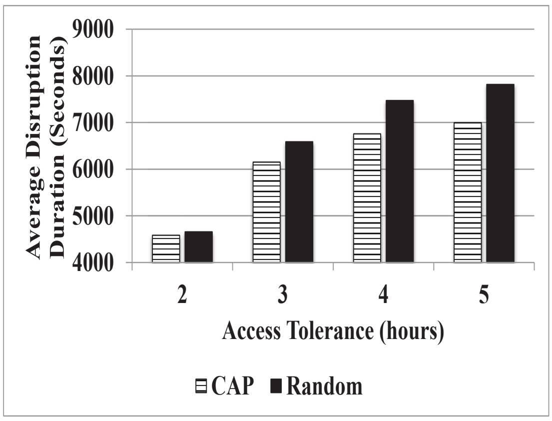
[5] Removing 3 hours of warm-up time, there are still 13 hours. The trends of all results start to be steady when simulation time is 12 hours. 16 hours can avoid any accident.



(a) Success Ratio



(b) Average Disruption Frequency



(c) Average Disruption Duration

Figure 4.6: Influence of Access Tolerance

increases rapidly, along with the growth of access tolerance, from 30% for 2 hours, 68% for 3 hours, 92% for 4 hours to 97% for 5 hours. The CAP outperforms Random, by

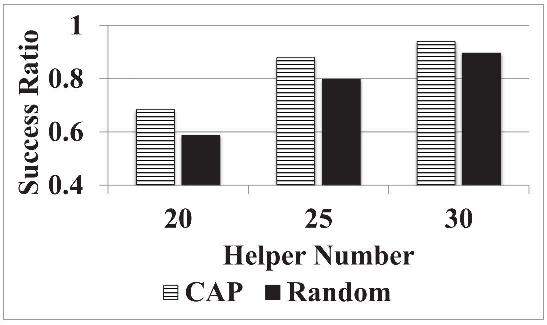


Figure 4.7: Influenced of Helper Number

about 7% for 2, 3 and 4 hours. However, when it comes to 5 hours, Random can also reach 96%. In terms of average disruption frequency, CAP and Random both increase side by side with the length of access tolerance. CAP outperforms Random in average disruption duration and both of them grow with access tolerance, because larger access tolerance can tolerate longer disruption duration.

4.3.2 Influence of Selected Helper Number and Helper Access Limit

Here the number of selectable helpers M is set as 20, 25, 30 respectively, with a transmit range of 50 m, 3 hours' access tolerance ($T_{AT} = 3$) and an access limit of 5 concurrent requesters ($K = 5$). As shown in Fig. 4.7, the success ratio is sensitive to M, with steps of 5 leading to it soaring from 68% to 88% and finally 93% with CAP. CAP outperforms Random in success ratio since appropriately selected helpers are able to successfully relay requesters. In Fig. 4.8, access limit K is set as 1, 3, 5 respectively, with $M = 20$, transmit range of 50 m and $T_{AT} = 3$. CAP also outperforms Random in success ratio. $K = 3$ outperforms $K = 1$ but is the same with $K = 5$, which means that there are rarely five concurrent requesters connected to one helper.

4.3.3 Influence of Transmit Range (LTE-Direct)

The transmit range of LTE-Direct is set as 100 m and 150 m, with 20 selected helpers ($M = 20$), 3 hours' access tolerance ($T_{AT} = 3$) and an access limit of 5 concurrent

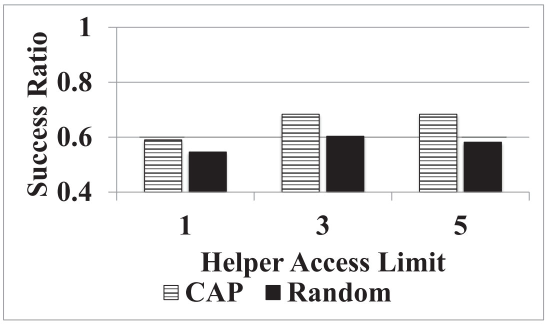


Figure 4.8: Influenced of Access Limit

requesters ($K = 5$). Only continuity of PaaS is shown in Fig. 4.9 since success ratio is always 100% in both cases. Transmit range influences the duration of every contact between pairwise helper and requester, so the average number of disruptions (contacts) during each access service is the same in CAP and Random if the success ratio is 100%. When the transmit range is 150 m, CAP and Random can both work well. Such an observation implies that miscellaneous helpers with longer transmit ranges can provide seamless PaaS. For example, drones implemented with in-band D2D air interfaces suffer less shadow loss[105], multi-path effects and sensitivity to battery life compared to phones.

4.3.4 Discussion on Transmission Range of LTE-Direct

Since the chip-set of LTE-Direct has not been implemented into smartphones yet, there is no real-world experiment. Based on Qualcomm white book[6], the transmit range of LTE-Direct air interface can reach 500 metres. The dedicated in-band D2D technique can use low and unique bands which are suitable for longer range transmission. And in-band D2D networks are without interference, compared with unlicensed band technologies (Bluetooth and WiFi with ISM bands). However, 500 metres is not feasible in the real-world because transmissions are limited by channel conditions such as shadow loss and battery shortage. As it is common to assume that the WiFi interface has the range of

[6]https://www.qualcomm.com/media/documents/files/lte-direct-always-on-device-to-device-proximal-discovery.pdf

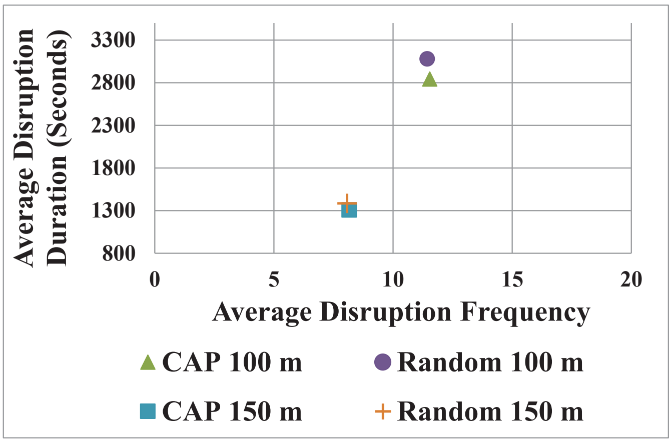


Figure 4.9: Influence of Transmit Range

50 metres, this thesis assumed that the LTE-Direct has a range of 100 metres and 150 metres in the simulation.

However, if the dedicated in-band D2D technique (LTE-Direct) is implemented in devices without the limitations above, for example, drones, the radio channel for Aerial-to-Terrestrial D2D communication[106] can avoid disturbances such as shadow loss and multi-path to some extent. There is no doubt that drones are also power sensitive, but compared with smartphones, the wireless transceiver only consumes a small amount of their total consumption. So the transmit range can reach 200 metres (the same with the range of pico-cells as shown in TABLE 4-D) or much more.

Table 4-D: Transmission Range of Different Kinds of Cells

Cell Type	Typical Cell Size
Macro	1-30km
Micro	200m-2km
Pico	4-200m
Femto	10m

Drones might be everywhere in the future and there is a trend that wireless capability will be implemented everywhere in almost all sorts of devices. For example, companies of drone delivery, car-sharing and bike-sharing can also share network resources which might be temporary access enhancement (can be called "Mobile Hetnet"), which is much cheaper and more convenient. So seamless access enhancement is not impossible through D2D collaborative system in the future.

4.4 Summary

This chapter proposes a simple use case (access enhancement) of PaaS, formulates and proposes a heuristic scheme CAP, to facilitate the selection of appropriate helpers for PaaS. Simulation results show the advantage of CAP, in both success ratio and continuity of the service. Results further reveal the influence of four factors, e.g., service tolerance, number of helpers allocated, the number of concurrent devices supported by each helper and transmit range, on the PaaS through the opportunistic D2D relay.

However, the current performance analysis cannot get full information on the PaaS system (for example, interactions of the inputs and regression equations) and find the optimal set of values of factors. Therefore, in the next chapter, the methodology of DOE is elaborated to evaluate the information more fully.

Chapter 5

Experimental Design and Data Analysis

5.0.1 The Aim of the Experiments

The aim of the experiments in this chapter are to get the nature of the influence of various factors. Experiments are iterative and sequential trials, and in many cases, the results of an experiment are just for pointing out the direction of the next experiment. In this chapter, the first step is the determination of the aim of the experiments; and the second step is the choice of the inputs and outputs of the experiments; then the third step is determination of the levels of factors; the last step is data analysis and optimization.

The goal of the previous chapter and this chapter is to explore the methodology to study the PaaS with the simple use case so it is unnecessary to find every subtle high-order interaction. However, it is still meaningful to find the magnitude of the main effect of every factor and interactions between different factors with proper resolution.

5.1 Choice of Inputs and Outputs of the Experiments

Different from the previous chapter, a comprehensive consideration of all possible influential factors and responses with various perspectives is necessary in the methodology of DOE. There are several metrics that can evaluate the performance of the experiments (outputs of the experiments) and several factors that can influence the performance of the experiments (inputs of the experiments). Determinations of the inputs (factors) and the outputs (metrics) are prior to factorial design.

5.1.1 Choice of the Inputs

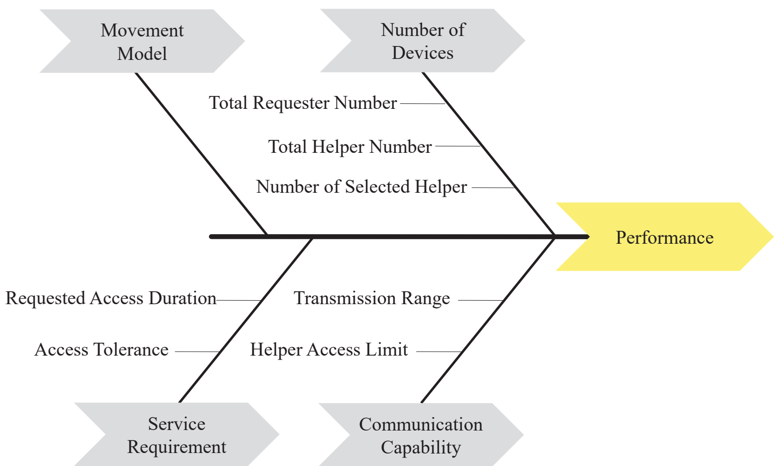


Figure 5.1: The Inputs of the Experiments

A cause-and-effect fishbone is shown in Fig 5.1. There are four categories of inputs:

- **Number of Devices:** The number of devices can influence the density of nodes (nodal density) because the size of the map is fixed. The nodal density has two influences on the performance of connectivity. First, the number of selected helpers determines the chances of contacts for a requester. Second, the number of requesters determine the loads on the helpers. If helpers are overloaded (concurrent requesters exceeding helper access limit), the chance of services[1] are also influenced. Therefore, the number of selected helpers determine the chance of contacts; the scale between the number of requesters and the number of selected helpers determine the load of traffic on the helpers. With the increase of the number of requesters, the number of the total access requesters (total traffic loads) will increase; with the increase of the number of selected helpers, average loads for each helper will decrease.

- **Movement Model:** The movement model has an influence[107] on the connectivity of opportunistic networks because the contacts and disruptions between nodes are caused by the movement. The scale of the map influences the nodal density. The movement model can be a trace or a synthetic movement model.[2] For any synthetic movement model, there are various parameters to be set.

  If the movement model is considered as an unchangeable environment (a configuration)[3] rather than an input factor, this research would be narrowed to a certain case. However, analysis of the main effects of factors and interactions are still with general meaning to demonstrate the methodology of PaaS research.

- **Communication Capability:** Transmission range is another important factor because it directly determines the connectivity between nodes. It is worth mentioning that, for the movement traces, the number of nodes and transmission ranges are fixed. So this research chooses the synthetic movement model. This research focuses on the performance of connectivity (services, contacts and disruptions). Therefore, bandwidth or throughput are not directly considered as an input factor. However, the number of concurrent requesters a helper can serve (helper access limit) is another form of bandwidth. There are three kinds of concurrent problems.

[1] The chance of contacts and the chance of services are different because one helper can only serve a limited number of requesters simultaneously (helper access limit).

[2] This experiment uses a synthetic map-based point-of-interest movement model.

[3] In this case a synthetic map-based point-of-interest movement model

  - **One Requester with Several Helpers:** In this chapter, the system is defined that a requester can only communicate with one helper at a time.

  - **One Helper with Several Requesters:** A helper can only serve a limited number (**K**) of requesters (helper access limit), due to the limitation of air interface resources. A helper ranks all the contacting requesters by urgency (in regard to access tolerance and requested access duration) and selects **K** requesters to serve.

  - **Several Helpers with Several Requesters:** At the same time when helpers choose requesters by urgency, requesters can also sort helpers by workload and requested resources.[4]

  In order to investigate the problem of concurrent connections between one helper and several requesters, the helper access limit was changed and the difference of outputs was checked. A helper can only serve a limited number of requesters simultaneously and the limited number is called the helper access limit in this thesis.

- **Service Requirement:** Generally speaking, connectivity is always the goal of network researches, but the performance is always based on particular service requirements. In other words, the service itself can also influence the performance. A network can support services such as TCP and UDP. Both of these transport layer services require to transfer traffic with complete one-by-one packages (segments). For UDP, the length of the package can influence the results of the experiment; for TCP, the setting of TCP slow start (for congestion control in wired network) can dramatically[15] influence the performance in wireless network scenario. In this chapter, the contacts between helpers and requesters depend on the transient local nodal density and transmission range. For packet-level simulation, stable networks tend to offer TCP-IP services and opportunistic network networks tend

[4]This research has not implemented this many-to-many D2D collaborative framework but the anti-social scheme in the future work will facilitate this.

to offer delay tolerant services. The frameworks of the networks can be varied in this case. Since the PaaS systems is assumed to have one hop D2D communication, in order to get rid of the choice of DTN or TCP-IP, this research proposes a service requirement framework composed of access tolerance and requested access duration, which is not packet-level.

This thesis sets the movement model as one configuration because the experimental work is simply an exploration of methodology. The total number of requesters and helpers are fixed because pilot runs have shown that the loads of a helper are always under 3 concurrent requesters[5] with a reasonable transmission range no matter how many helpers and requesters are under current movement model. Therefore, the number of selected helpers is chosen as one factor. For service requirement, only the proportion matters so the requested access duration is fixed and the access tolerance is set as one factor. So this experiment has four factors: access tolerance, the number of selected helpers, helper access limit and transmission range. In the following sections, these four factors are also called A, B, C and D for simplicity.

5.1.2 Choice of the Outputs of the Experiments

The connectivity and performance can be evaluated by four categories of metrics.

- **Service Completion:** "completion" means: during an access request (Access Tolerance), the service time is more than a given period (requested access duration). The performance metric success ratio is for service completion.

- **Total Service Time:** Service time is not equal to contact time because overloaded helpers can not provide services when contacting requesters; requesters without access request also do not need services when contacting helpers. Total helper access duration + total unfinished duration = the total number of access

[5] Pilot runs have shown that there are up to 11 concurrent requesters, but the cases with 3 concurrent requesters or more are rare.

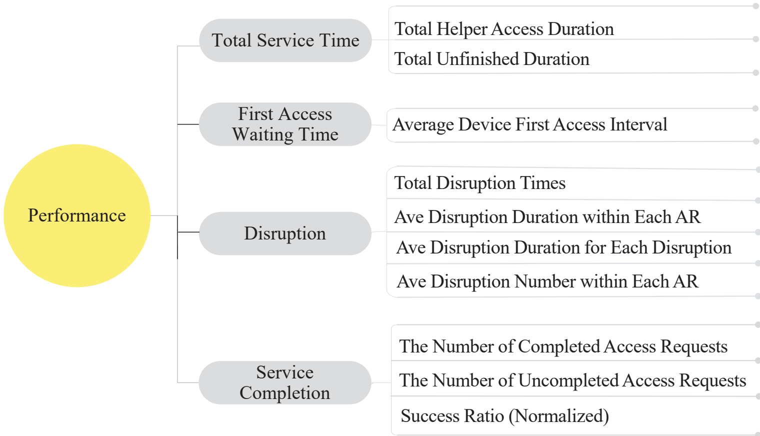


Figure 5.2: The Inputs of the Experiments

requests * requested access duration. Success ratio sometime cannot reveal the performance of the system because the system with good performance always generates more access requests which consequently cannot always increase success ratio (more completed access requests and more uncompleted access requests). This is the reason that total helper access duration and success ratio together can reveal the performance of connectivity. Note that total helper access duration in this experiment is accumulated till the end of the simulation time, which is different from a single service.

- **Disruption:** The continuity can be measured from two perspectives: during each Access Request (AR), the average number of disruptions and the average total length of these disruptions. These two outputs are called the number of disruptions and disruption duration.

- **First Access Waiting Time:** When an access request (AR) is generated in a requester, how much time it takes to get connected with the first helper. Sometimes this metric can be important but generally, connectivity and continuity can measure the performance of this use case.

Above all, four responses are chosen: total helper access duration, success ratio for connectivity and the number of disruptions and disruption duration for continuity. In the following sections, these four response are also called R1, R2, R3 and R4 for simplicity.

5.2 Determination of Levels of Input Factors

The first step of DOE is to determine the number of levels of each factor and the value for each level, which means that pilot runs and regression analyses need to be conducted covering all the possible ranges of each factor.

5.2.1 Determination of Levels of Transmission Range

From previous experience[11], the most influential factor is transmission range, so the first step is determining the levels of transmission range. Moreover, for the first pilot runs, different ways of traffic generation and random seeds are also tested in this section.

The pilot runs have a combination of two inputs: transmission range and two ways of traffic generation. For each combination, 8 runs with different random seeds are repeated. The settings of other factors are shown in Table 5-A. The two ways of traffic generation are:

- **Constant Interval:** After a service, the next access request is generated after 60 seconds (constantly, marked as C 60).
- **Exponential Interval:** After a service, the next access request is generated after an exponential period with an average of 60 s (Poisson process, marked as E 60).

Table 5-A: Settings of Other Factors

Factors	Values
Access Tolerance	10800 s
Helper Access Limit	3
Requested Access Duration	1800 s
The Number of Selected Helpers	20
Total Requester Number	100
Total Helper Number	100

5.2.1.1 Analysis of Influence of Transmission Range, Seeds and Traffic Generation on Total Helper Access Duration

The influences of transmission range, random seeds and traffic generation have been checked by experiments. As shown in Fig 5.3, the increase of transmission range can dramatically increase the total helper access duration. Random sequence successfully brings randomness to the results. Traffic generation has no effect on the total helper access duration.

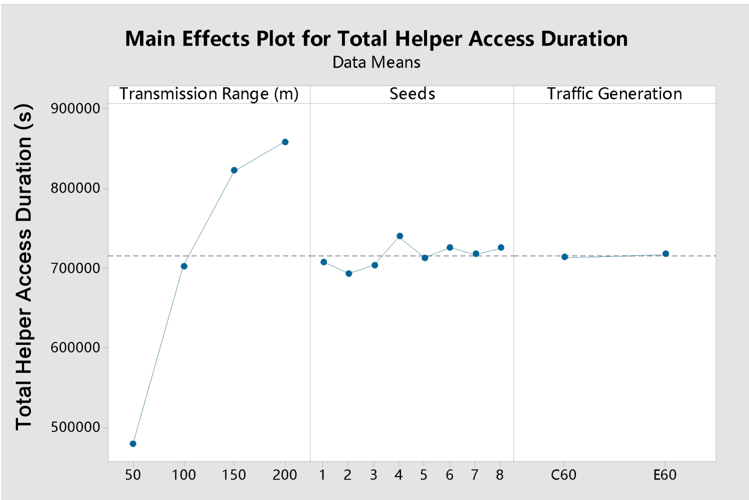


Figure 5.3: Comparison of Main Effects on Total Helper Access Duration

Multiple linear regression was then performed. The P-values for transmission range, seeds and traffic generation are 0, 0.235, 0.756, which means only transmission range has effects on the response.[6] R is 88.46%, which means a nice fit. Adjust R (R-sq(adj))

[6] In many statistic software like Minitab, if P-value is smaller than 0.001, P-value would be shown as 0

is 87.88%, not a big decrease, which means the nice fit is not because of the number of predictors. Predicted R (R-sq(pred)) is 86.88%, not a big decrease, which means it is not an over-fit. Standard error (S = 52481.6, approximately 10% of response) shows that approximately 95% of the observations should fall an area within plus or minus two standard errors (a range of 40%) from the regression line. The residual plot in Fig 5.4 shows the range of fluctuations is larger than a standard error, which means there is still a systematic pattern in the residuals.

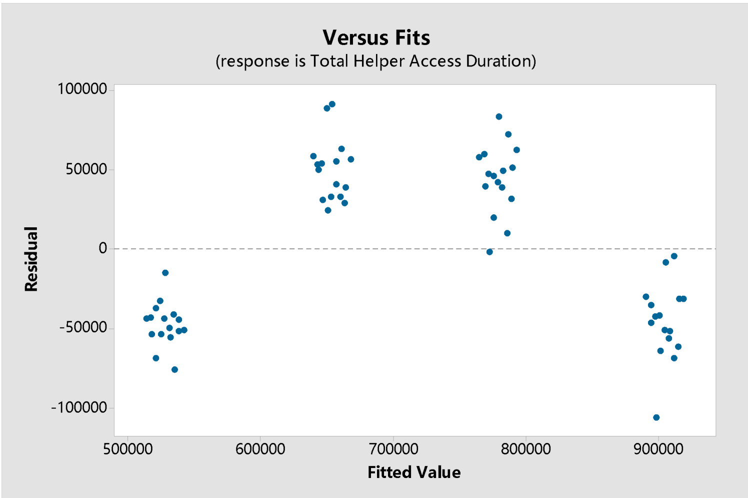


Figure 5.4: Residuals vs Fits Plot for Total Helper Access Duration

Therefore, quadratic multiple regression was then performed. S (standard error) = 20569.8 (approximately 3.5% of the mean of response, which means approximately 95% of the observations should fall an area within up or down 7% from the regression line.), R-sq = 98.26%, R-sq(adj) = 98.14%, R-sq(pred) = 97.97%, which means a nice fit with a precise prediction. P-values for first and second order terms of transmission range are 0.000, 0.001 and 0.456, which means that the second order term really "sneaks" into the residuals of linear regression. As shown in Fig 5.5, residuals are now balanced, which means quadratic regression is enough.

Anyway, a cubic regression analysis is also performed and the P-value of third order term of transmission range is 0.422, which means the effect is without statistic significance. The P-value of random seeds (P = 0.000) shows that random seeds have an effect on the response. The reason is that the second and third order terms can influence the same amount of fluctuations as random seeds and the software over-fits the random pattern generation.

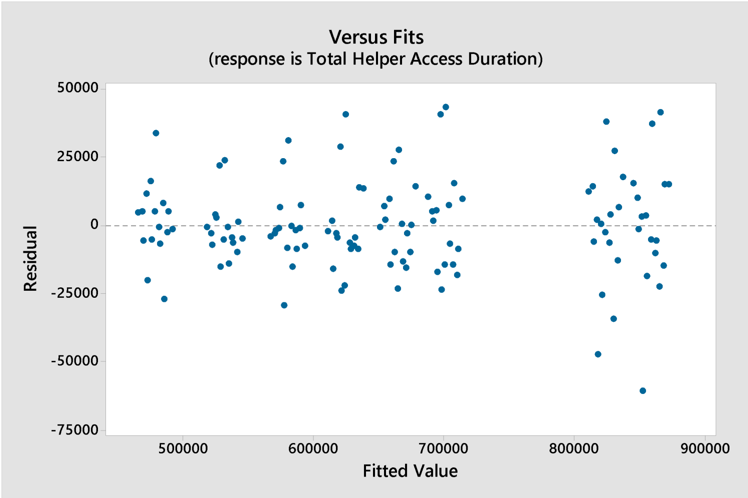


Figure 5.5: Residuals vs Fits Plot of Non-Linear Regression for Total Helper Access Duration

The single linear and quadratic regression for transmission range are also performed, as shown in 5.6 and Fig 5.7.

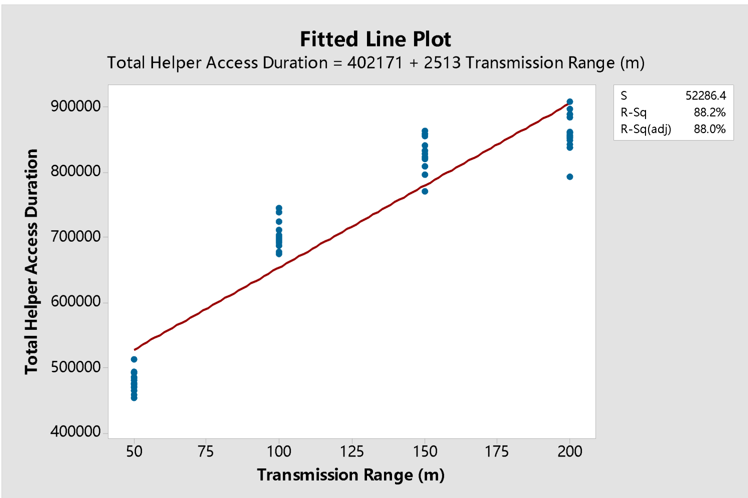


Figure 5.6: Linear Regression Line for Total Helper Access Duration

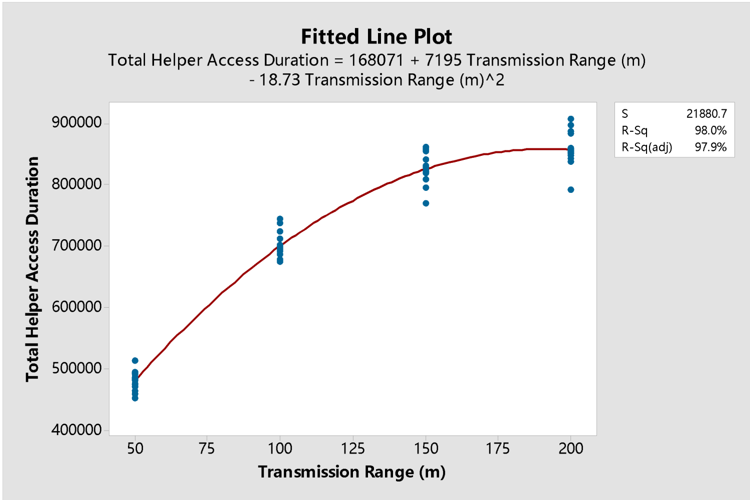


Figure 5.7: Quadratic Regression Line for Total Helper Access Duration

5.2.1.2 Analysis of Influence of Transmission Range, Seeds and Traffic Generation on Success Ratio

As shown in Fig 5.8, transmission range has a dramatic influence (not linearly) on success ratio. The random generator successfully brings randomness to the results. The type of traffic generation has no effect.

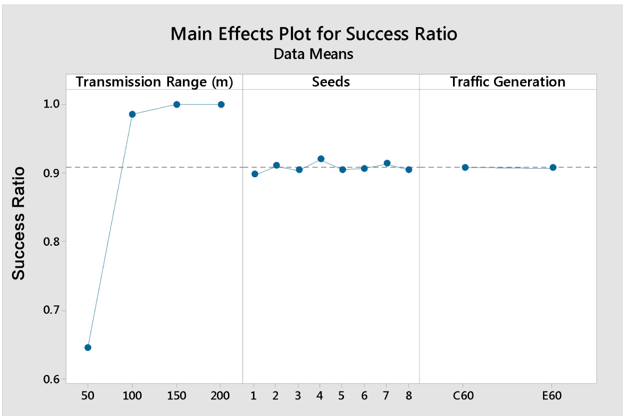


Figure 5.8: Comparison of Main Effects on Success Ratio

The multiple linear regression for success ratio analysis is performed. The standard deviation is 0.09, which means that the approximately 95% of the observations should fall an area within 36% from the regression line. That is not precise for a simulation experiment.

Therefore, a polynomial multiple regression analysis for success ratio is needed. In this case, standard deviation is 0.01, which means it is a precise fit. R-square, adjust R-square and predicted R-square are 99.44%, 99.39% and 99.32%, which means a nice fit but not an over fit. In this case, the P-values for first, second and third order terms of transmission range are all 0.000 so that it is a cubic fit. The residuals are shown as Fig 5.9, which are balanced, the fluctuations are within plus or minus 4%.

The success ratio always needs a higher order regression model than total helper access duration, which means more levels in factorial design. The reason is that this research defines the maximum value of success ratio as 1(normalized), the convergence curve would finally reach 1 (not in a linear way). It is the convergence curve that increases the order of the regression model.

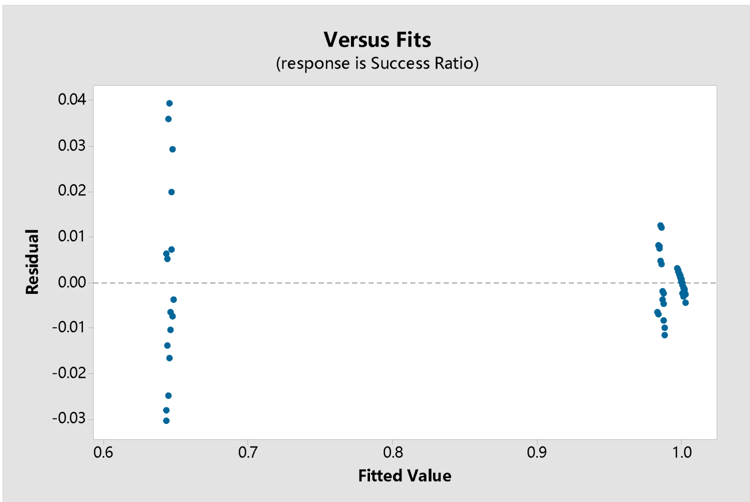


Figure 5.9: Residuals vs Fits Plot of Non-Linear Regression for Success Ratio

5.2.1.3 Analysis of Influence of Transmission Range, Seeds and Traffic Generation on Average Disruption Number for Each Access Request

As shown in Fig 5.10, the increase of transmission range can dramatically decrease the number of disruptions. Random seeds successfully bring randomness to the results. Traffic generations still have no effect.

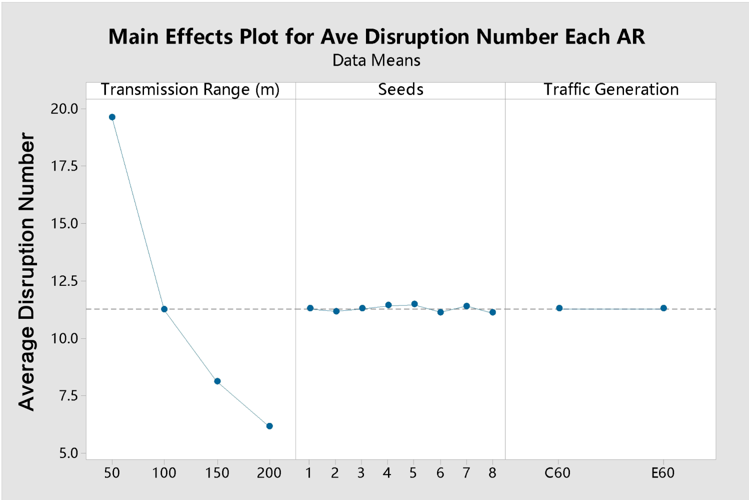


Figure 5.10: Comparison of Main Effects on Average Disruption Number for Each Access Request

To further investigate the influence of these factors, multiple linear and polynomial (quadratic and cubic) regression analyses are performed. For linear analysis, the P-values of transmission range, random seeds and traffic generation are 0.000, 0.939 and 0.979. The standard deviation is 1.72, residual squares are 89.39%, 88.86% (adjusted) and 87.95% (predicted). As shown in Fig 5.10, the means of average disruption numbers range from 19 to 5, 1.72 is too big. As shown in Fig 5.11, the residuals around the value 5 is still above 1. So linear regression is not a good choice.

For quadratic regression analysis, the P-values of transmission range, transmission range * transmission range, random seeds and traffic generation are 0.000, 0.000, 0.800, 0.931. The standard deviation is 0.522, residual squares are 99.05%, 98.98% (adjusted)

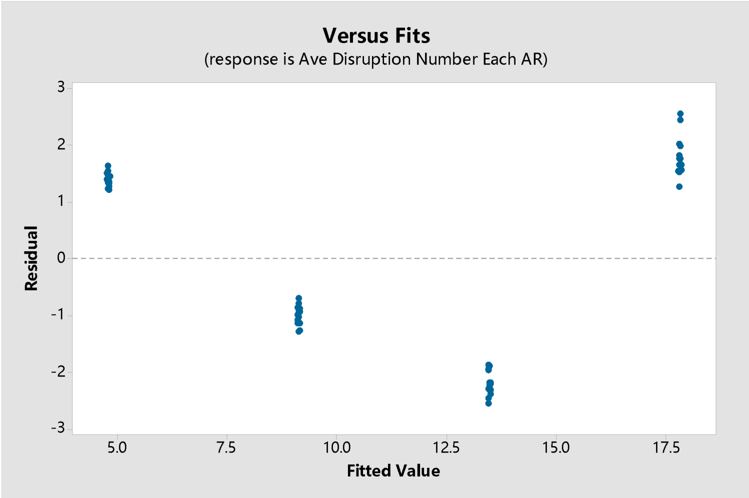


Figure 5.11: Residuals vs Fits Plot of Linear Regression for Average Disruption Number for Each Access Request

and 98.90% (predicted). The residual plot is shown in Fig 5.12. There is still a pattern in the residuals, so cubic regression analysis is necessary.

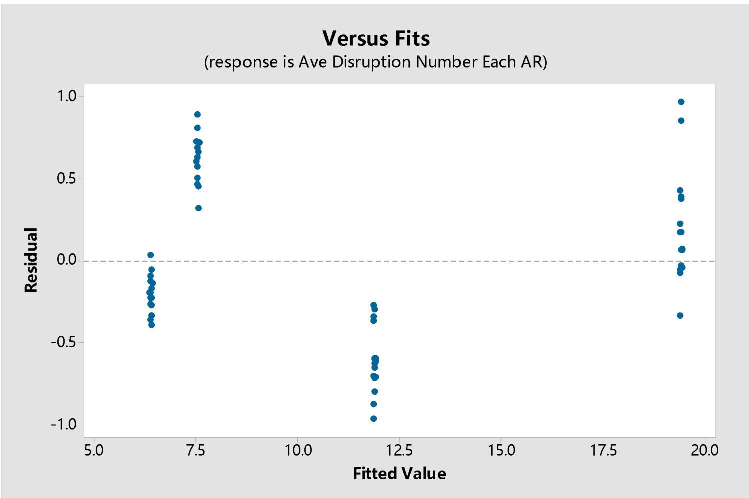


Figure 5.12: Residuals vs Fits Plot of Quadratic Regression for Average Disruption Number for Each Access Request

For cubic regression analysis, the P-values of transmission range, transmission range * transmission range, transmission range * transmission range * transmission range, random seeds and traffic generation are 0.000, 0.000, 0.000, 0.555, 0.840. That means cubic item indeed plays a role. The standard deviation is 0.224, residual squares are

99.83%, 99.81% (adjusted) and 98.79% (predicted). It is not a big decrease for prediction residual square, which means cubic regression is not an over-fit. The residual plot is shown in Fig 5.13.

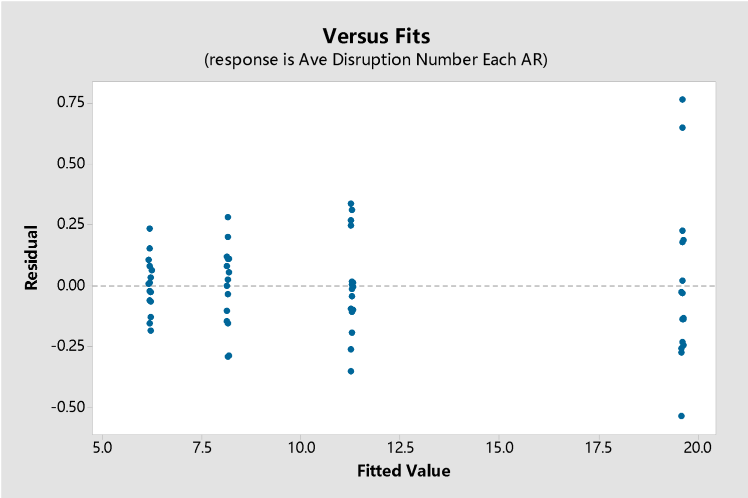


Figure 5.13: Residuals vs Fits Plot of Cubic Regression for Average Disruption Number for Each Access Request

Since a cubic relationship means 4 levels in factorial design, which would bring complexity. So it is necessary to discuss the necessity of cubic regression. Compared with the standard deviations of quadratic and cubic regression analyses, 0.224 is better than 0.552 (5% or 10% of the least value). To have a closer look, as shown in Fig 5.12, all the residuals of quadratic regression are within 10% of the value and the means of the residuals are about 5% of the current value. As for the cubic residual plot (Fig 5.13), the residuals are constantly within 5% of the current value, and the means of the residuals are about 2.5% of the current value.

For standard deviation and residuals, cubic regression might be more precise. But for residual squares, quadratic regression and cubic regression are almost the same. Despite the fact that all the outputs reveal the nature of connectivity of the network, the normalised output, like success ratio, brings a curve that increases the order of regression. The discretized output like average number of disruptions may also bring an increase of regression order. In order to further investigate the features of disruption, a regression analysis of continuous disruption output like average total disruption duration during each access request (AR) is necessary.

5.2.1.4 Analysis of Influence of Transmission Range, Seeds and Traffic Generation on Average Total Disruption Duration for Each Access Request

As shown in Fig 5.14, the increase of transmission range can dramatically decrease disruption duration. Random seeds successfully bring randomness to the results. Traffic generations still have no effect.

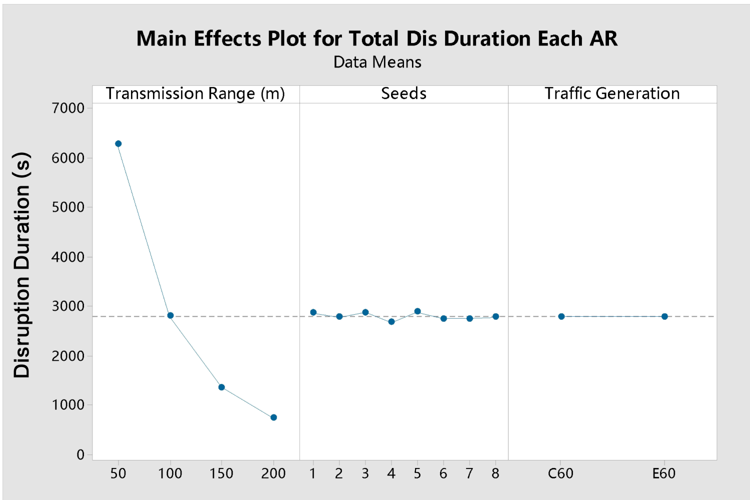


Figure 5.14: Comparison of Main Effects on Total Disruption Duration for Each Access Request

Linear, quadratic and cubic regression analyses are performed. For linear regression, the P-values of transmission range, random seeds and traffic generation are 0.000, 0.811 and 0.991. The standard deviation is 785.208. Residual squares are 88.30%, 87.72% (adjusted) and 86.70% (predicted). For quadratic regression analysis, the P-values of transmission range, transmission range * transmission range, random seeds and traffic generation are 0.000, 0.000, 0.302, 0.960. The standard deviation is 174.661. Residual squares are 99.39%, 99.35% (adjusted) and 99.29% (predicted). Residual squares show that the quadratic regression equation almost perfectly fit the curve. Combined with standard deviation and residuals shown in Fig 5.15, the means of the residuals are about 5% of the current value of the curve. In Fig 5.15, there is a trivial pattern in the residuals but can be neglected.

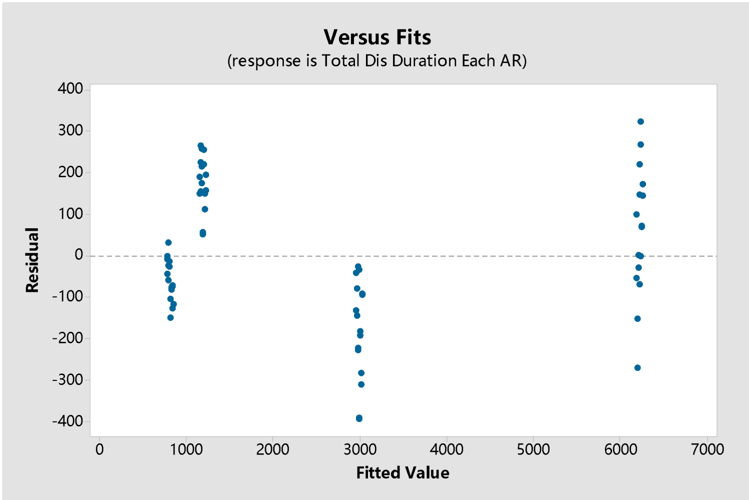


Figure 5.15: Residuals vs Fits Plot of Quadratic Regression for Total Disruption Duration for Each Access Request

5.2.1.5 Conclusion on Influence of Transmission Range, Seeds and Traffic Generation

The analyses in this section investigate the nature of connectivity by two outputs of connections and two outputs of disruption. Random seeds are tested and they successfully bring randomness to the results. The type of traffic generations has no effect on success ratio, which means using one way of traffic generation is without loss of generality.

Whether the 3-order item of transmission range can be neglected or not is based on the complexity of conducting 4-level factorial experiment. For the number of disruptions, the influence of the 3-order item of transmission range is amplified because of the discretization. For success ratio, this research defines the maximal value of success ratio as 1(normalized) and the convergence curve would finally reach 1 (not in a linear way), which increases the order of the regression model to 3.

For outputs without normalization and discretization, linear regressions are reasonable (all residual squares are above 85%); quadratic regressions are precise; cubic regression has no statistical significance (total helper access duration) or trivial effect (total disruption duration during each AR).

5.2.2 Choice of Levels of Inputs

Similarly, determinations of levels of other three factors are elaborated in the appendix and the choice of levels are shown in Table 5-B. Generally, the two-level factorial design is for linear trend; the three-level factorial design is for quadratic trend; the four-level factorial design is for cubic trend, and so on.

In Table 5-B, levels for different factor and response pairs (combinations[7] of A, B, C, D and R1, R2, R3, R4) are various, which means the experiment needs a sophisticated fractional factorial design. The factor C (helper access limit) has no trend.[8] The possible reason might be the settings of other factors, which is a drawback of control variates method and might be eliminated by factorial design. The impact of C might be not obvious unless it interacts with other factors. Since simulations are not costly, all four factors are set with four levels,[9] so that full factorial design including C can be conducted.

Table 5-B: Levels of Factors

	R1	R2	R3	R4
A (7200 s 10800 s 14400 s 18000 s)	4	$3 \to 4$	$3 \to 4$	$3 \to 4$
B (20 25 30 35)	$2 \to 4$	$3 \to 4$	$2 \to 4$	$3 \to 4$
C (1 2 3 4)	$\to 4$	$\to 4$	$\to 4$	$\to 4$
D (50 m 80 m 100 m 150 m)	$3 \to 4$	4	$3 \to 4$	4

[7] Four responses are: Total Helper Access Duration (R1), Success Ratio (R2), Disruption Duration (R3), the Number of Disruption (R4); four factors are: Access Tolerance (A), Number of Selected Helpers (B), Helper Access Limit (C), Transmission Range (D).

[8] There is a jump from 1 to 2, but then, for any response, the trends start to fluctuate.

[9] In the language of DOE, virtual levels (four-level) are given to all factors if the experiment is not costly.

Full factorial design needs to traverse all the combinations of factors. For one response, there are 256 (4^4) combinations because the four factors have four levels respectively. Here the number of repetitions for every combination are set with 5, so the total number of runs for each response is 1280 (5120 runs for four responses, $1280 * 4$). For each response, there are two regression equations: the coded regression equation and the actual regression equation. The actual regression equation can get the actual regression value based on a certain set of values of factors. However, the coefficients of different factors cannot indicate the importance of factors since factors represent different dimensions of quantities. As shown in Table 5-B, for factor A, levels are from 7200 s to 18000 s; for D, levels are from 50 s to 150 s. Coefficients of coded regression equations can indicate the importance of different factors by comparing changes of responses along with the changes of levels.

5.3 Data Analysis

After full factorial experiments, regression equations are obtained. Note that, the levels of factors are set to fit the design space and different factors have different dimensions, scales, and units. Therefore, coded factors that are independent of the associated scale are needed to estimate the change in the response for 1 level change in the factor. Table 5-C shows eight regression equations and Equation (5.1) is an example of the actual regression equation of R1.

$$\begin{aligned} R1 = {} & 1.860E + 16 + (7.792E + 12)A + (-4.536E + 15)B + (5.839E + 15)C + (-1.318E + 15)D \\ & + (-4.912E + 10)AB + (-1.811E + 10)AD + (1.173E + 14)BC + (6.565E + 13)BD \\ & + (1.032E + 14)CD + (-4.567E + 08)A^2 + (1.182E + 14)B^2 + (-3.955E + 15)C^2 \\ & + (1.039E + 13)D^2 + (-1.208E + 12)BCD + (1.742E + 06)A^2B + (6.334E + 05)A^2D \\ & + (-4.156E + 11)B^2D + (-2.114E + 11)BD^2 + (-3.566E + 11)CD^2 + (8.332E + 03)A^3 \\ & + (-1.172E + 12)B^3 + (4.247E + 14)C^3 + (6.991 * 10^3)D^3 \end{aligned} \tag{5.1}$$

Table 5-C: Coded and (Actual) Equations of R1, R2, R3 and R4

Coded (Actual) R1		Coded (Actual) R2		Coded (Actual) R3		Coded (Actual) R4		=
8.973E+05	4.651E+05	1.030E+00	-7.141E+00	2.112E+03	2.692E+03	1.176E+01	-1.220E+01	
-1.456E+04	1.948E+02	-2.440E-02	6.760E-04	-1.666E+02	2.352E+00	-3.050E-02	4.398E-03	A
2.099E+05	-1.134E+05	-2.500E-03	1.718E-01	-6.885E+02	5.049E+01	-1.070E-02	1.490E+00	B
2.553E+04	1.460E+05	-7.200E-03	3.526E-01	-3.180E+01	-2.786E+02	7.800E-03	4.645E-01	C
4.703E+05	-3.295E+04	-2.870E-02	8.352E-02	-1.630E+03	-1.483E+02	-5.260E+00	-7.322E-02	D
-5.278E+03	-1.228E+00	-4.580E-02	-8.826E-06	-1.570E+02	-4.428E-02	-3.282E-01	-6.300E-05	AB
		-1.080E-02	-1.500E-05	-2.591E+01	-2.452E-02	-7.180E-02	-3.400E-05	AC
-1.454E+04	-4.529E-01	-1.182E-01	-3.977E-06	-3.313E+02	-1.389E-02	-8.922E-01	-3.700E-05	AD
-9.913E+02	2.933E+03	-4.200E-03	-2.482E-03					BC
-2.347E+04	1.641E+03	-5.440E-02	-1.184E-03	3.935E+02	1.040E+00	-5.521E-01	-1.384E-02	BD
-2.590E+03	2.579E+03	-9.600E-03	-1.164E-03	6.270E+00	-9.405E-01	-2.560E-02	-3.476E-03	CD
-2.225E+04	-1.142E-02	-6.670E-02	-2.169E-08	-1.647E+02	-6.800E-05	-4.336E-01	-9.418E-08	A^2
-2.825E+04	2.955E+03	-1.340E-02	-1.192E-03	1.485E+02	-5.797E-01	-1.139E-01	-8.196E-03	B^2
-4.330E+04	-9.887E+04	-9.300E-03	-5.518E-02	8.943E+01	1.502E+02			C^2
-1.331E+05	2.598E+02	-7.300E-02	-3.350E-04	8.808E+02	1.480E+00	3.100E+00	3.119E-03	D^2
		3.400E-03	5.596E-08					ABC
		5.010E-02	2.476E-08	2.368E+02	1.170E-04	5.073E-01	2.505E-07	ABD
		9.800E-03	2.428E-08	3.292E+01	8.100E-05	1.008E-01	2.488E-07	ACD
-1.699E+04	-3.021E+01							BCD
9.528E+03	4.356E-05	2.680E-02	1.227E-10	1.272E+02	5.817E-07	2.552E-01	1.167E-09	A^2B
		8.000E-03	1.831E-10					A^2C
2.309E+04	1.583E-05	7.990E-02	5.479E-11	2.467E+02	1.692E-07	6.886E-01	4.723E-10	A^2D
		1.100E-02	3.612E-08	7.761E+01	2.560E-04			AB^2
		1.330E-02	1.091E-06	3.206E+01	2.639E-03			AC^2
		9.570E-02	7.086E-09	3.359E+02	2.500E-05	9.504E-01	7.040E-08	AD^2
-2.923E+04	-1.039E+01	1.400E-02	4.976E-06			1.736E-01	6.200E-05	B^2D
		4.700E-03	2.810E-04					BC^2
-9.908E+04	-5.284E+00	4.250E-02	2.269E-06	-1.372E+02	-7.319E-03	5.450E-01	2.900E-05	BD^2
		1.050E-02	9.300E-05					C^2D
-3.343E+04	-8.916E+00	5.000E-03	1.326E-06					CD^2
3.280E+04	2.083E-07	4.200E-02	2.670E-13	1.235E+02	7.845E-10			A^3
-1.237E+04	-2.931E+01							B^3
3.583E+04	1.062E+04	9.100E-03	2.700E-03	-6.467E+01	-1.916E+01			C^3
-6.060E+04	-4.847E-01	6.270E-02	5.016E-07	-5.167E+02	-4.134E-03	-1.490E+00	-1.200E-05	D^3
	9.935E-01		9.803E-01		9.934E-01		9.944E-01	R^2
	9.934E-01		9.798E-01		9.933E-01		9.943E-01	$AdjR^2$
	9.933E-01		9.793E-01		9.931E-01		9.942E-01	$PredR^2$

As shown in coefficients of coded regression equations in Table 5-C, the factor D is more important than other factors for all responses (except R2) by at least one order of magnitude. For R2 (success ratio), A (access tolerance) is directly associated with service completion since more tolerance time means more chance to satisfy the requested duration. For R4, D is the only factor that matters (by two orders of magnitudes). For R1 and R3, B is the second important factor.

Based on the actual regression equations, regression lines for four factors and four responses are drawn and compared with the results of the experiments in Fig. 5.16. Intuitive comparisons of the trends of regression lines and conclusions of coded factors

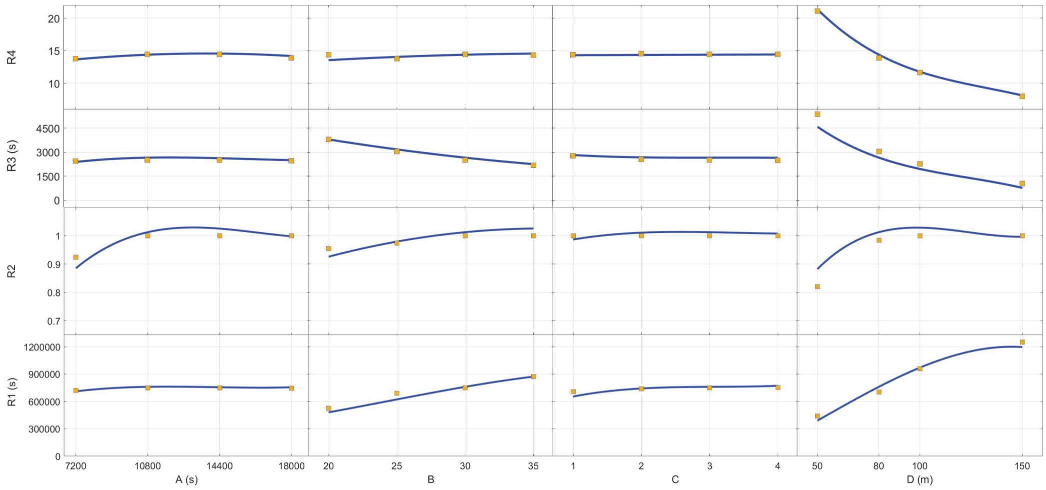


Figure 5.16: Regression Lines of Four Responses and Four Factors

are mutually verified.

Overall, the R-Squares[34] (R^2) of all responses are high (close to 1), which means regression lines fit the results of experiments well and the multiple regression lines derived from the full factorial design verify the high R-squares. The good performance of the fitting of regression lines is not because of the number of predictors and over-fit since the Adjusted R-Squares[34] ($AdjR^2$) and Predicted R-Squares[34] ($PredR^2$) are basically the same with R^2.

The default levels of the four factors of Fig. 5.16 are shown in Table 5-D. Without loss of showing the differences of the trends of regression lines, the aim of the choice of these levels is to show the saturated state of the system, such as the regression line of D for R1 and R2. Therefore, it is necessary to find the optimal settings of factors to maximize the system.

Table 5-D: Default Levels of Factors of Fig. 5.16

Factors	Values
Access Tolerance (R1)	10800 s
Helper Access Limit (R2)	3
The Number of Selected Helpers (R3)	30
Transmission Range (R4)	80 m

R2 is success ratio, the values over 1 are meaningless but it is inevitable for regression lines. The dominance and monotonicity of B and D for all responses are proved by regression lines in Fig. 5.16 and partial derivation.

5.4 Optimization of the Factors

This section addressed the optimization of 4-tuple responses $\{R_1, R_2, R_3, R_4\}$, which is formulated as

$$\begin{aligned}
& max\ R_1(\mathbf{x}) \\
& max\ R_2(\mathbf{x}) \\
& min\ R_3(\mathbf{x}) \\
& min\ R_4(\mathbf{x}) \\
s.t.\quad & \mathbf{x} = (A, B, C, D)^T \in \mathbf{R}^4 \\
& 7200 \leq A \leq 18000 \\
& 20 \leq B \leq 35 \\
& 1 \leq C \leq 4 \\
& 50 \leq A \leq 150
\end{aligned} \tag{5.2}$$

It is a non-linear Multi-Objective Optimization Problem (MOOP)[108]. A common methodology is to convert a multiple-objective optimization problem into the single-objective optimization problem. The importance of four objectives of responses in this system is difficult to be weighted because different responses have different dimensions, scales, and units, which means the method of *weighted sum*[108] is not feasible. Since the four responses are intrinsically consistent and are four perspectives to evaluate the performance, we chose the ϵ-constraint method[108] which is formulated as

$$
\begin{aligned}
& max/min\ R_j(\mathbf{x}) \\
s.t.\quad & \mathbf{x} = (A, B, C, D)^T \in \mathbf{R}^4 \\
& R_i(\mathbf{x}) \geq \epsilon_i\ or\ R_i(\mathbf{x}) \leq \epsilon_i \\
& \forall\ i \in \{1, 2, 3, 4\} \backslash \{j\} \\
& 7200 \leq A \leq 18000 \\
& 20 \leq B \leq 35 \\
& 1 \leq C \leq 4 \\
& 50 \leq D \leq 150,
\end{aligned} \tag{5.3}
$$

where ϵ_i is the lower or upper bound in the last optimization procedure and is one of the non-linear constraints. In each step of iterations, there is only one optimization problem and three constraints which had been optimized in previous steps.

The initial optimal set without constraints was first obtained, and then we impose non-linear constraints from the initial optimal set to iterate, as shown in Table 5-E.

Table 5-E: MOOP Evolution According to ϵ-constraint.

Iteration	RA1	RA2	RA3	RA4
0	1.237E+06	1.034	559.785	8.099
1	1.1713E+06	0.9752	1003.7	8.1853
2	1.1781E+06	1.0134	1054.4	8.8743
3	1.2171E+06	0.9914	898.2377	8.9766
...	...	...	...	...

For R1 and R2, the more the better; for R3 and R4, the less the better. For ϵ-constraint method, there might be no iteration result that is better than other iterations, as shown in Table 5-E. The solution set is a non-dominated set[108], which is referred as Pareto-optimal set[108]. But the differences between iterations are small and a specific iteration can be chosen based on particular needs.

By using the platform proposed in [109] based on genetic algorithms[110], it is convenient to produce a Pareto-optimal set, which is listed in Table 5-F. A confirmation experiment for the second iteration is conducted and the results verify the table (RA1 to RA4: 1.2012E+06 s, 1, 612.323 s, and 8.012).

Table 5-F: Producing a Pareto-Optimal Set of the Factors.

A(s)	B	C	D (m)	RA1 (s)	RA2	RA3 (s)	RA4
12137	32	2	86	8.6822E+05	1.034	2227.767	13.52
12825	35	3	150	1.1838E+06	0.972	560.251	8.174
12140	33	2	131	1.2116E+06	0.987	1046.72	9.248
12314	32	2	94	9.5645E+05	1.031	1927.994	12.378
12096	32	3	92	9.4273E+05	1.032	1992.645	12.673
12143	32	2	86	8.6477E+05	1.034	2240.447	13.499
12195	33	3	120	1.1684E+06	0.998	1267.368	9.939
12188	33	2	108	1.0855E+06	1.016	1539.796	10.928
12302	32	3	98	9.8728E+05	1.027	1840.091	11.998
12060	34	2	102	1.0579E+06	1.021	1626.752	11.53
12179	33	2	126	1.1929E+06	0.992	1157.145	9.566
12137	32	2	87	8.8695E+05	1.034	2172.969	13.358
12174	32	3	105	1.0558E+06	1.018	1641.25	11.219
12362	33	2	99	1.0234E+06	1.023	1716.856	11.783
12143	32	2	90	9.1054E+05	1.034	2081.453	12.928
11876	35	2	136	1.2283E+06	0.98	902.137	8.958
12210	30	4	150	1.1744E+06	0.986	721.616	8.102

In this chapter, we conduct a complete DOE method of performance analysis of the PaaS system for access enhancement. For this process, the most important advantage of DOE is that DOE can obtain regression equations, which is the foundation of the comparison of factors and the optimization.

Chapter 6

Conclusion

As one of the candidate technologies in the 5G era, whether Device-to-Device (D2D) communication can be widely implemented in the industry is uncertain. This research give a comprehensive discussion on the disadvantage and distinctiveness of D2D technology from the point of view of both academia and industry. In summary, the disadvantage is that D2D is highly dependent on users and transmission range of D2D air interfaces; the distinctive potential is that D2D can knit people's life with proximity services together which might be a way to make the mobile network manager a "smart enabler" rather than a "dumb pipe".

This thesis argues that the common features of PaaS are: **human participation, mobility, and resource shortage.** Based on that, this research proposed a literature review about use cases exploration with a perspective of resource allocation; a survey on mobility which is categorized by proactive or passive movement with social and geographical constraints; and an introduction about the incentive mechanism used in the local D2D collaborative system. Moreover, the methodology of PaaS research is introduced including simulation tools, evaluation methods and Design Of Experiment (DOE).

In this thesis, helpers, service requesters, and D2D collaborative mechanisms together form the Proximity as a Service (PaaS) framework. By coordination of cellular networks, proximal services of various use cases can be conducted. Two complete researches of a case (coverage extension with a system model) of PaaS have been done: the algorithm research and the research of the fundamental impact of key factors. The methodology of algorithm research is composed of the problem formulation, the algorithm design, and performance evaluation. There are many dimensions of the state space (access tolerance, the number of selected helper, helper access limit, and transmit range) that can influence the performance of CAP. Therefore, the research of the fundamental impact of key factors utilizes DOE methodology, which is composed of the choice of factors, determination of levels of factor, data analysis, and optimization.

This thesis also proposed several promising future works in the paradigms of D2D and mobile computing. For example, mode selection on fully participatory PaaS and semi-participatory PaaS; social helper selection and anti-social helper selection.

The contributions of this thesis are:

- This work propose a PaaS use case where D2D communication is between requesters and helpers. Helpers are regarded as the service provider and each requester can only receive help from a limited number of helpers. In such a case, helper selection is the main concern and is described mathematically by problem formulation.
- A heuristic helper selection algorithm is proposed and evaluated by experiments.
- For various use case explorations of proximal services via D2D communication, this thesis explored these use cases based on sharing of different proximal resources: network resources, human resources, sensing resources, storage resources, and computation resources.
- This work contributes to the introduction of DOE in the field of mobile computing

and D2D. With the precise multiple regression equations, it is easy to predict the influence of any change of the four factors within a certain domain. Although a particular movement model was chosen in this work, this work demonstrates the general feasibility and advantage of using DOE.

- The ONE simulator is made more suitable for PaaS system and DOE use.
- With polynomial multiple regression equations, finding the optimal values of key factors is turned into a mathematical problem. It is not an NP-hard problem because the regression equations are designed to be polynomial.
- It is intuitive that the longer transmission range and the more the number of selected helpers, the better the performance. This thesis proves the montonicity by a partial derivative.
- Transmission range and the number of selected helpers have dramatic influence on performance and they are monotonic. There is always a particular value of transmission range which fits a particular user density. For example, in the city center, a small range is enough and a long transmission range may disturb proximal cells when D2D chooses in-band air-interface. In the countryside, the transmission range needs to be long and can be long. Therefore, transmission range and user density are the main concerns.

Above all, the main contribution is that this thesis points out a new perspective of D2D research focusing on the proximal services; it provides a complete methodology for the research of PaaS; sheds light (the future trend with broad social consensus of sharing economy, like Uber and Airbnb) on sharing the spilling over resources of mobile devices; and draws a blueprint that can make our lives better and safer with proximal services and boosts the development of D2D technology in both academia and industry.

The drawbacks and unfinished works of this thesis are as follows.

- Since the PaaS system is newly proposed, there is no comparison of performance in the literature.
- The review of the use case exploration has not been finished. It is a challenging task to find methods to make the best of scarce opportunistic proximal resources with a time limit.
- The incentive mechanism has not been integrated with the resource allocation and decision making mechanisms yet.
- Recruiting a device as a helper as a decision has been done but allocation of continuous resources such as storage, computation power, and bandwidth has not been done.
- Different movement models (including proactive movement) have not been applied and compared.

In summary, the current works could be extended to further support the perspective and methodology (D2D with the consideration of proximal services) proposed in this research. This thesis has finished a complete work for one simple use case and paved the way for future works. A comprehensive exploration is indispensable so a detailed work plans of future works is elaborated in the next chapter.

Chapter 7

Future Works

Work plans in the future are small steps forward compared with the use case of access enhancement. The next tep to consider with the movement model, including incentive mechanism, would be the proactive movement PaaS; the next step with continuous resource allocation would be allocating bandwidth; the the next step with algorithm together would allocating bandwidth will be anti-social helper selection PaaS, which is in a dense scenario. In the future, the incomplete literature review of use case exploration would be finished.

7.1 Proactive Movement PaaS

7.1.1 PaaS for Missions

Service coverage for public safety users is critical. In a natural disaster, a Public Safety Network (PSN) should be prepared with a rapidly deployable fallback solutions to ensure the connections of PSN users, as Kumbhar et al. said in[27].

> Law enforcement agencies, fire departments, rescue squads, emergency medical services (EMS), and other entities that are referred to as emergency first

> responders (EFR). The ability of EFR to communicate amongst themselves and seamlessly share critical information directly affects their ability to save lives

EFRs are called public safety users with high priority in this thesis. Ordinary users in proximity would be public safety helpers in the collaborative system of PaaS for access enhancement in a PSN scenario.

If a disaster happens, for example an earthquake, some base stations[1] may fall in earthquake fields. Therefore, far-reaching coverage would be a key metric in this situation. The transmit power of base stations is suppressed to avoid adjacent spectrum collisions in ordinary days, but it can be fully released during the disaster if adjacent base stations cannot work. However, the transmit power of mobile devices is still limited. Therefore, in the case of disasters, the transmission range of devices is assumed to be uplink limited in this thesis. With the help of D2D communication, uplink data of public safety requesters can reach headquarters and dispatchers.

The research problem might include:

- **Design Mission-Oriented Group Trajectory:** The rescue trajectory can be planed before the mission. The features of mission-driven PaaS are shown in Table 7-A.
- **Proximal User Selection:** Prioritization of public safety subscribers and services is critical in a disaster.
- **D2D Mode Selection:**D2D mode selection is necessary because the situation in a disaster zone could change frequently. For example, the D2D network should change between DTN and ad hoc based on nodal density; the D2D network should

[1] PSN infrastructure and commercial infrastructure are assumed as one convergence infrastructure in this thesis.

also change between network-assisted central-controlled D2D and distributed autonomous D2D based on the signal interaction between the PaaS system and residuary base stations.

- **Incentive Mechanism:** Privacy and security might not be priority topics, and monetary rewards would not be important. However, reputation incentive mechanism might be useful since no one wants to bury a hero volunteer. The aforementioned block-chain applications (LocalCoin)[41], where the computational hardness (Bitcoin) has been replaced by social collaboration hardness, which might be a way to measure contribution in rescue activities.

7.1.2 PaaS for Demands

Even if there is no base station destroyed by disasters, the cellular infrastructure and public safety infrastructure still need fallback solutions to handle some incidents, for example, supporting large-scale temporary assembly (concerts, exhibitions, and carnivals), no matter how well the network is designed and implemented. Moreover, the number of public safety users is significantly lower than regular subscribers in any commercial mobile network, but the number of simultaneously active public safety users can be very high, especially in a relatively small area when a major incident occurs. Operators usually send emergency communication devices (vehicles and drones) to the hot zone. Mobile Network Operators (MNOs) can also authenticate nearby users to provide PaaS such as Mobile Picocells or Mobile Access Points for instant capacity and coverage boost, by one-to-many UE-Relay Device-to-Device (D2D) communication technologies with incentive schemes and proactive movement. This solution might be timely, ubiquitous, seamless and economical due to the pervasiveness of phones.

However, since the battery life of a UE is limited, the UE acting as a helper is energy sensitive. The period of PaaS by this helper might be not enough for covering the demands of a requester, which requires a scheme to dynamically authenticate another helper to take-over PaaS job. Moreover, both helpers and users may move, which means that the communication duration of a helper and a requester is uncertain. It is impossible to exploit the extensive resources of wireless capacities without handling the opportunistic nature of D2D communication.

If no one in proximity can take-over this job in an area, users in nearby areas (the whole map is divided into many areas by Geo-Hash[111]) can also proactively move to this area. The demand-driven mobility can also be designed by a trajectory guidance mechanism to optimize the PaaS. Unlike mission-driven mobility, demand-driven movement cannot be planned since demands may occur at any time. However, along with edge computing abilities implemented in helpers, the contact duration and inter-contact time of a helper and a requester can be learned (topological or geographical) by the demand-driven mobility mechanism. Based on the contact duration, helpers can smartly allocate bandwidth to requesters with different demands. Base stations can also allocate more spectrum blocks to helpers if it is the case of the network-assisted in-band dedicated D2D network.

The usage of proactive mobility guidance and resource management will provide fruitful and profound insights to broaden the vision of D2D use case exploration in academia. Table 7-A concludes the features of two different kinds of PaaS systems for comparison.

7.2 Extension to the Use Case of Access Enhancement by Bandwidth

The utility function of the helper selection algorithm in chapter 4 is solely based on contact history. But when the nodal density of one-hop one-to-many opportunistic D2D network reaches a certain extent, the research focus should be on alleviating the burden

Table 7-A: Comparison of PaaS Systems with Different Proactive Movement

Differences	Demand-Driven PaaS	Mission-Driven PaaS
Definition of Helpers and Requesters	Helpers are service providers	Can be two (or more) kinds of functionalities for mission
Relation of Helpers and requesters	Unknown, there might be communities	Known, belong to an organization
Aims of Mobility	Helpers move for requesters demands	All the movements are mission-driven
Alias of Mobility Aims	User-centric Mobility (U-Mobility)	Mission-centric Mobility(M-Mobility)
Plan of the Trajectory	Cannot be planed beforehand	Can be planed beforehand
Applications	temporary assembly (concerts, exhibitions and carnivals)	public safety network, emergency, disaster relief
Relation with Base Stations	Central-controlled by base stations	Mode selection between central and distributed control

of popular nodes since a popular node might be connected by too many proximal devices. So a requester should try to bother the popular helper as little as possible. This can be explained by a metaphor of daily social life **"keep your close friend closer but don't bother your popular eminent friend too much"**. This idea can be called "Anti-Social Helper Selection", which needs the consideration of mode selection, battery consumption and bandwidth.

7.2.1 Anti-Social Helper Selection in D2D Networks

Anti-social is not the only parameter to consider in the decision making process. If the maximum disruption duration of requester A is 2 minutes in a certain use case of PaaS, the requester A should select the helper B, and the helper B should reduce the allocation of resources to other requesters simultaneously, to allocate resources for requester A. Moreover, many-to-many D2D communication can be enabled in the new research so that the urgent requester can occupy more resources from multiple helpers.

Mobility prediction is vital in this work. Since no requester wants to postpone its tasks to the last minute, requester A should entirely rely on helper B and finish the task during the contact duration if no helper would appear after B within the time constraint of the task.

When the service is time-sensitive, or the time constraint is close, geographical prediction is a better choice because it is real-time.[2] Moreover, inter-any-contact time could be a useful way to predict the interval of future contacts. A hybrid mobility prediction scheme combining both geographical and social information is promising in this context.

7.3 Steps to the Future Survey Work

For now, this thesis has not completed the detailed systematic survey to conclude representative use case exploration in this field. The current taxonomy categorizes the proximal services by five resources, which is the basis of a future detailed systematic survey. The steps to make a complete survey would be:

- **Features of Every Proximal Resource:** Describe the unique features of every proximal opportunistic resource and accumulate papers on each category.
- **Combination of Proximal Resources and Propose New Use Cases:** Note that the mobile proximal services are usually sharing several resources simultaneously (resource combinations). Therefore, a review of the up-to-date proximal services by the combination of mobile proximal resources would be necessary. It might be possible that there are some resource combinations which have no corresponding proximal service. Hence it would be our chance to propose new proximal services[3].
- **Selection of Math Models and Algorithms:** Review the selection of math models and algorithms for every representative use case exploration, and find out the reason based on the shortage of proximal resources (targets of the optimization).

[2] Please refer to the discussion on mobility prediction in chapter 2

[3] Please refer to "fog caching", which is proposed in this thesis.

- **An Overview of Performance Evaluation for PaaS:** Every proximal service has its unique evaluation metrics. For example, requirements of metrics in user-centric service would probably be delay, jitter, throughput and packet-loss.[4] However, mobile crowdsensing does not directly serve human beings,[5] which means small bits of delay and disruption would not bother the service performance. Sensing coverage with a time constraint (the deadline of a task), however, would probably be the evaluation metric. The evaluation metrics of proximal services will be categorized based on the taxonomy of service type. Whether a proximal service is user-centric is currently a foundation of the taxonomy. A complete taxonomy and an overview of performance evaluation could be achieved along with the accumulation of literature.

[4]Continuity in the access enhancement service[11] influences the experience of users because this proximal service is user-centric.

[5]Mobile crowdsensing has been studied by researchers from the paradigm of Internet of Things (IoT) and the duration of a sensing activity can be much longer (in the order of hours or days).

Appendix A

Determination of Levels of Factors

A.0.1 Determination of Levels of the Number of Selected Helpers

In order to determine the levels of the number of selected helpers, pilot runs with the following settings (Table A-A) are conduced.

Table A-A: Settings of Other Factors

Factors	Values
Access Tolerance	10800 s
Helper Access Limit	3
Requested Access Duration	1800 s
Transmission Range	50
Total Requester Number	100
Total Helper Number	100

A.0.1.1 Analysis of Influence of the Number of Selected Helper on Total Helper Access Duration

As shown in Fig A.1, the number of selected helpers has a dramatic influence on total helper access duration. It seems the regression relationship is linear. To verify whether

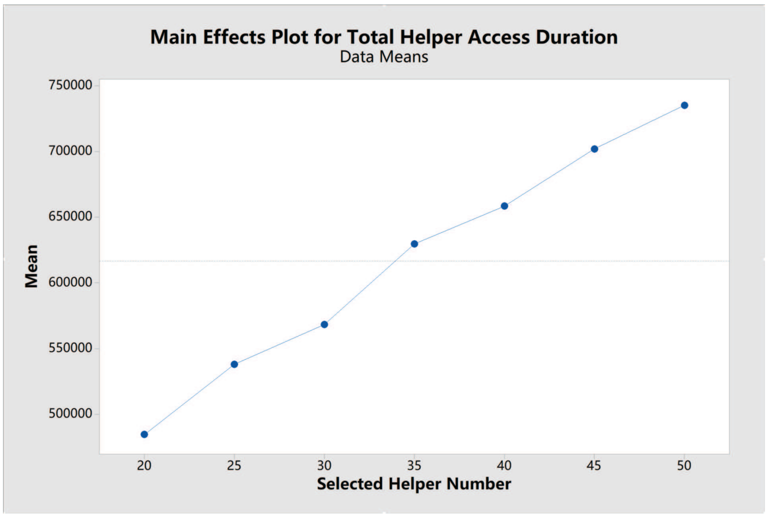


Figure A.1: Main Effect of the Number of Selected Helpers on Total Helper Access Duration

there is a quadratic effect, a single [1] quadratic regression analysis is firstly performed. The P-values of first order and second order items of the number of selected helper number are 0.000 and 0.063 (larger than 0.05), so the regression relationship between the factor and the response is linear.

The result of linear regression is shown in Fig A.2. The standard deviation (less than 2%) and R-squares (above 98%) show that linear regression is a precise fit.

A.0.1.2 Analysis of Influence of the Number of Selected Helpers on Success Ratio

As shown in Fig A.3, the number of selected helpers has a dramatic influence on success ratio and the convergence curve changes the regression relationship to quadratic. To verify it, a quadratic regression analysis was performed.

As shown in Fig A.4, the quadratic regression line can precisely fit this curve. There is a trivial pattern sneaking into the residuals and systematically fluctuating, which means

[1]There is only one factor changed in these pilot runs, so it is single.

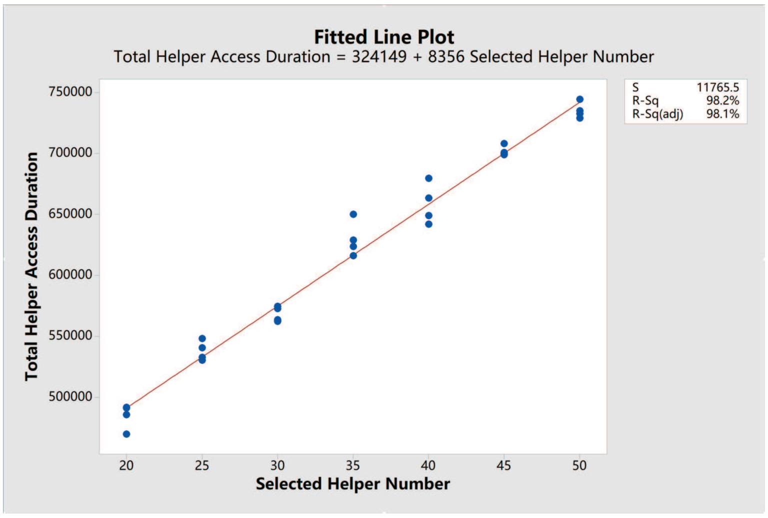


Figure A.2: Linear Regression Line of the Number of Selected Helpers and Total Helper Access Duration

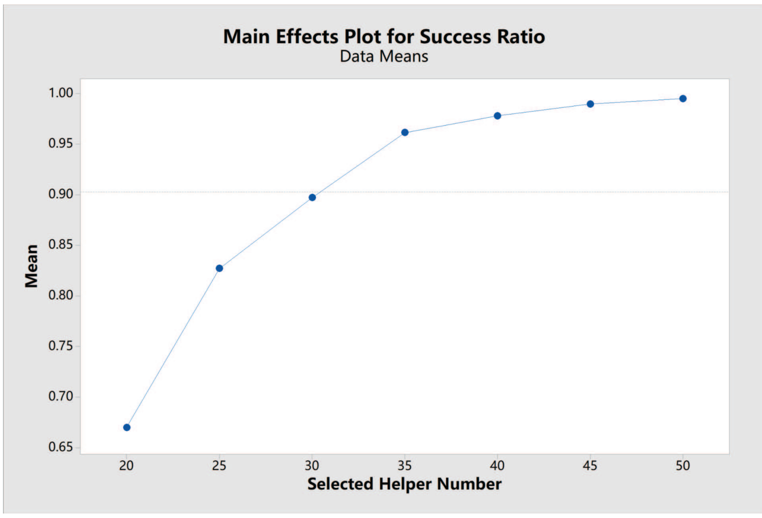


Figure A.3: Main Effect of the Number of Selected Helpers on Total Helper Access Duration

the third order item exists, but its disturbance can be neglected.[2]

[2]The disturbance of third order item is less than 0.02 (standard deviation).

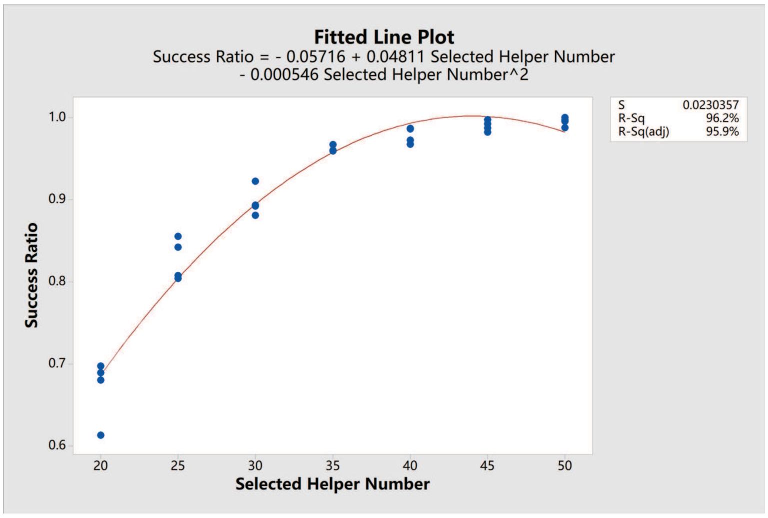


Figure A.4: Quadratic Regression Line of the Number of Selected Helpers and Success Ratio

A.0.1.3 Analysis of Influence of the Number of Selected Helpers on Average Disruption Duration During Each Access Request

The linear and quadratic regression lines are plotted in Fig A.5 and Fig A.6. First of all, the number of selected helpers has a dramatic influence on average disruption duration during each access request. The linear regression line decently (standard deviation is about 5% of the values and R-Squares are above 95%) fits the curve, but there is a pattern sneaking into the residuals.

The quadratic regression line erases the pattern in the residuals and precisely (standard deviation is about 2% of the values of the curve and R-Squares are above 95%) fits the curve.

A.0.1.4 Analysis of Influence of the Number of Selected Helpers on Average Disruption Number During Each Access Request

The linear regression line is plotted in Fig A.7. Different from other outputs, the number of selected helpers has a minor influence on average disruption duration during

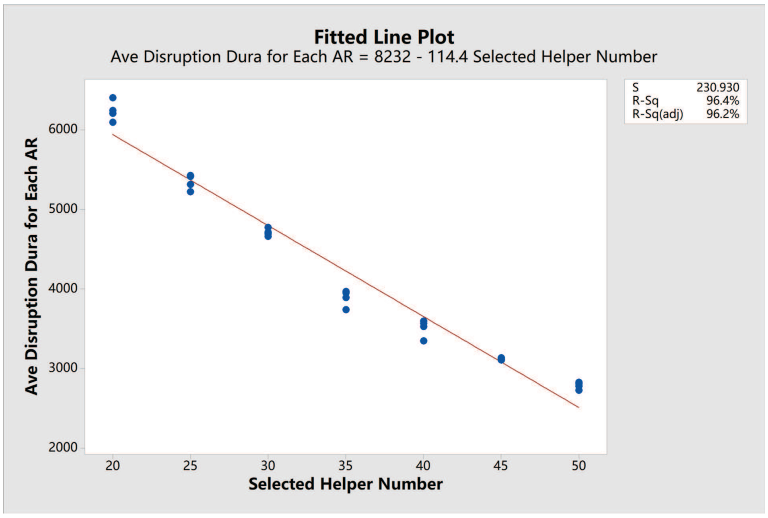


Figure A.5: Linear Regression Line of the Number of Selected Helpers and Success Ratio

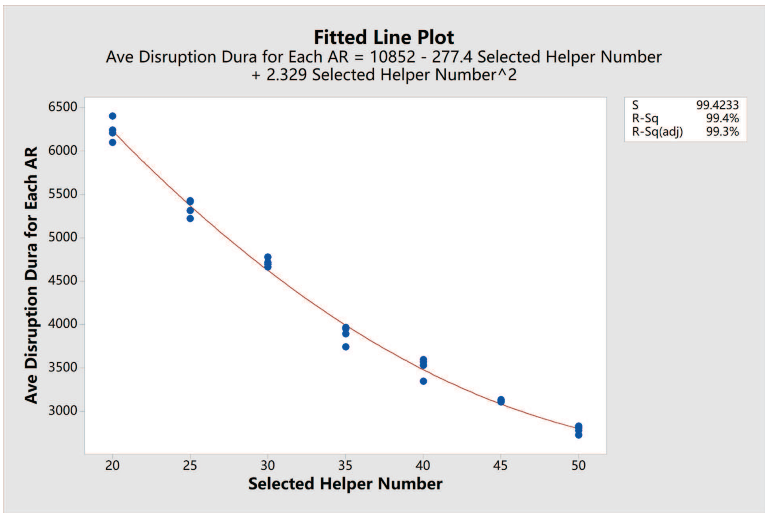


Figure A.6: Linear Regression Line of the Number of Selected Helpers and Success Ratio

each access request. As shown in the regression equation, the coefficient of the number of selected helpers is positive (0.08). The increase of the number of selected helpers does not decrease the average disruption number, which is opposite to the intuitive assumption. However, to think further, a disruption happens when a helper and a requester leave each other. For a certain movement model, only the transmission range can determine the duration of the contact between one helper and one requester. Based on this assumption, the "minor" influence seems to be too large.

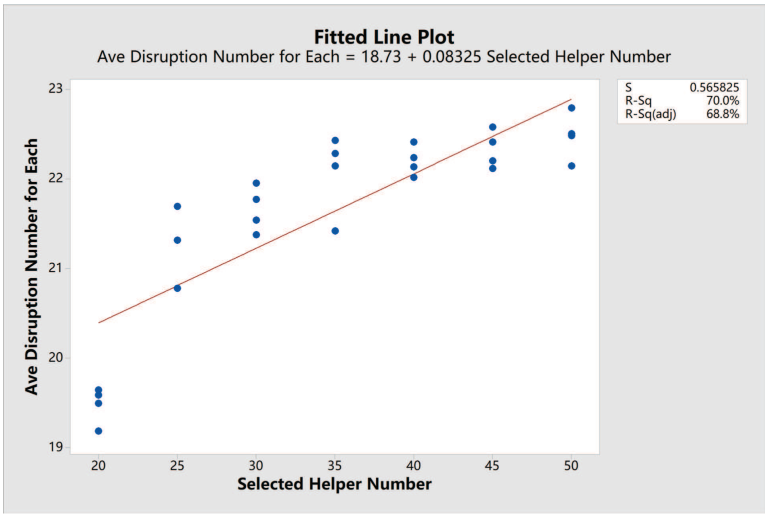


Figure A.7: Linear Regression Line of the Number of Selected Helpers and Success Ratio

Moreover, as shown in Fig A.7, the random seeds bring more randomness for this response. And there is a pattern fluctuating around the fit line. Therefore, a quadratic regression line is plotted in Fig A.8, which fits better.

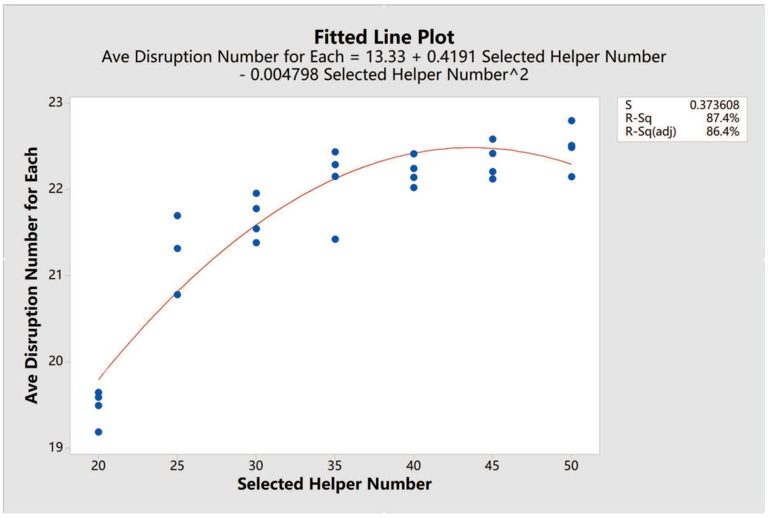


Figure A.8: Quadratic Regression Line of the Number of Selected Helpers and Success Ratio

A.0.1.5 Conclusion on Influence of the Number of Selected Helpers

The assumption before these pilot runs is that the number of selected helpers can linearly influence the connectivity, which is proved by the influence on total helper access duration. The influence on average disruption duration for each access request has a minor quadratic item[3] but the linear regression line is above 96%, which means the quadratic item can be neglected. So the assumption is verified.

For influence on success ratio and average disruption number during each access request, the quadratic items cannot be neglected.

A.0.2 Determination of Levels of Access Tolerance

The outputs are based on service requirements. For example, if the duration of the service for a requester has not reached the requested access duration within the duration of access tolerance, the access request is determined by the system as a failure, which means a decrease in **success ratio**. If the service duration has reached the requested access duration, it is determined as a success. If the requested duration is kept constant, the increase of access tolerance is assumed to increase success ratio.

Total helper access duration is accumulated by all requesters service duration no matter the service is successful or not. For a successful access request, before expiration of the access tolerance, the requested access duration is satisfied, which is constant. For an unsuccessful access request, after the expiration of access tolerance of this access request, the existing access duration is accumulated, which is various. When the access tolerance expires, a new access request will be generated after the traffic generation interval. During the interval, any contact duration is not accumulated into total helper access duration. So service requirement can also influence total helper access duration.

[3] The quadratic item in the regression equation of the output.

Therefore, it is hard to predict the regression relationship, which must be observed in the experiments.

It is obvious that the average disruption number during each access request is influenced by service requirements. If the access tolerance is too small or the requested access duration is too large, contact durations during the access request cannot reach the requested access duration. A requested access duration comprises several contact durations and the number of disruption during each access request is equal to the number of contacts (or $contacts - 1 = disruptions$). Therefore, if the requested duration is kept constant, the increase of access tolerance is assumed to increase the number of disruption. When the success ratio reaches 1, the average disruption number would be constant.

It is the same with the performance of disruption durations. If the requested duration is kept constant, the increase of access tolerance is assumed to increase the average or total disruption durations during each access request, because access tolerance can also be interpreted as disruption tolerance.

To determine the levels of access tolerance, exploratory experiments and regression analyses were done to verify the assumptions one by one. The settings of other factors are shown in Table A-B.

Table A-B: Settings of Other Factors

Factors	Values
The Number of Selected Helpers	20
Helper Access Limit	3
Requested Access Duration	1800 s
Transmission Range	50
Total Requester Number	100
Total Helper Number	100

A.0.2.1 Analysis of Influence of Access Tolerance on Total Helper Access Duration

The linear regression fit line is shown in Fig A.9. The values of total access duration seem to be outliers when access tolerance is equal to 10800 s (3 hours).

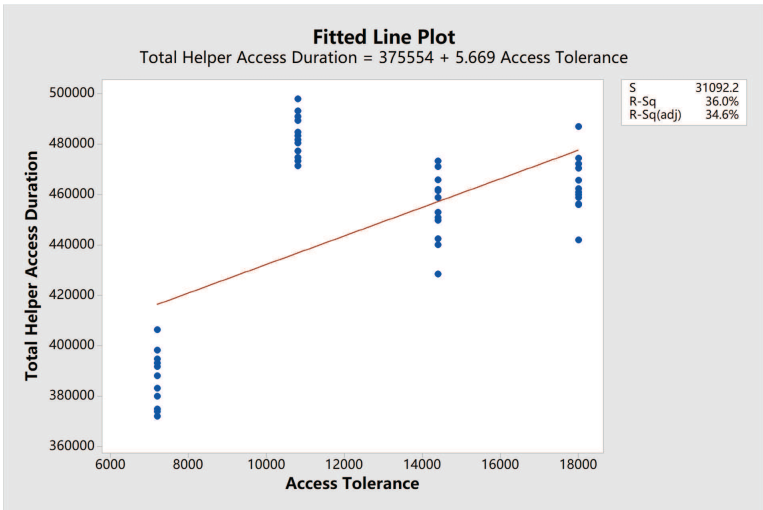


Figure A.9: Linear Regression Fit Line of Access Tolerance on Total Helper Access Duration

A possible explanation is that there is no need to wait for too long time. After 3 hours of "tolerance", if the service is still not completed, the system should abandon this service and start a new one. Compared with longer access tolerance, the total number of access requests increases so the total helper access tolerance increases although success ratio may drop down a little.

To verify this assumption, the linear regression fit for success ratio and total number of access requests are shown in Fig A.10 and Fig A.11. The success ratio keeps increasing from 7200 s to 14400 s but the total number of relayed requests keep basically constant from 7200 s to 10800 s, which verifies the previous explanation. There is a sudden drop of total number of relayed requests from 10800 s to 14400 s and it keeps constant again from 14400 s to 18000 s because success ratio has reached 1.

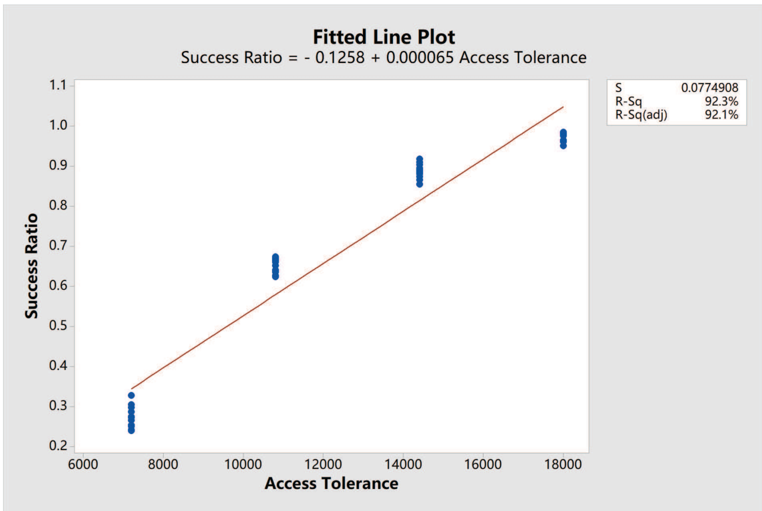


Figure A.10: Linear Regression Fit Line of Access Tolerance on Success Ratio

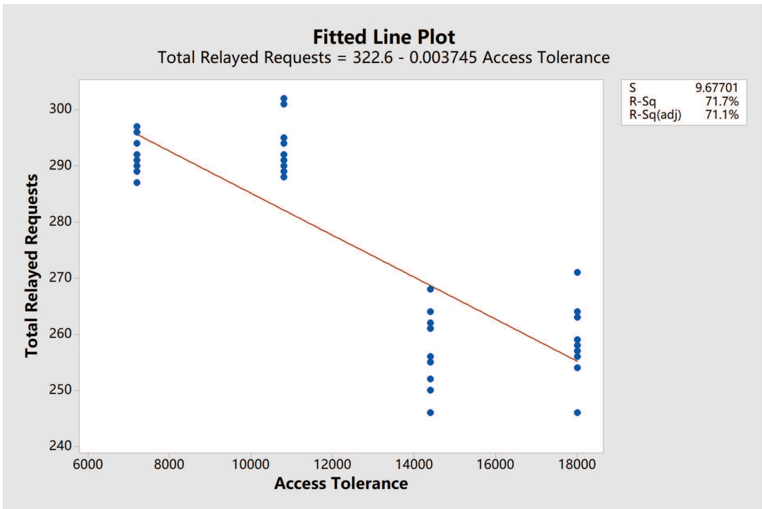


Figure A.11: Linear Regression Fit Line of Access Tolerance on Total Relayed Requests

There should be a highest point near 10800 s but it is not necessary to find it because it is not for optimization. Since the response of the total helper access duration has a curve with highest point and an increase after the curve, a cubic regression should be a good choice. The Fig A.12 and Fig A.13 verify the assumption. The cubic regression line goes through all the medians of the responses which there is no systematic pattern in the residuals. The R-Square is not 100 % because the random seeds brings normally distributed randomness.

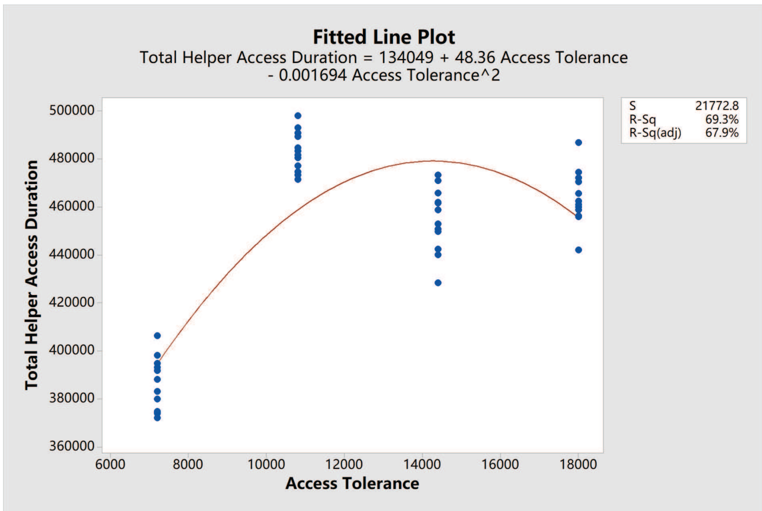


Figure A.12: Quadratic Regression Fit Line of Access Tolerance on Total Helper Access Duration

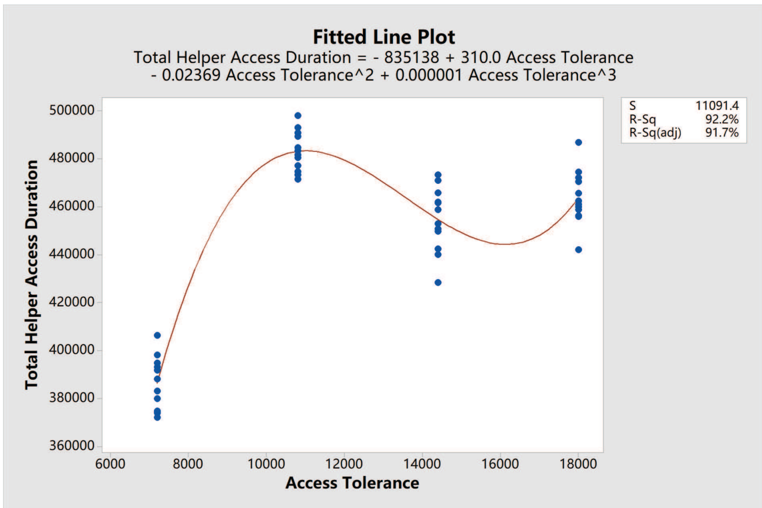


Figure A.13: Cubic Regression Fit Line of Access Tolerance on Total Helper Access Duration

A.0.2.2 Analysis of Influence of Access Tolerance on Success Ratio

As shown in Fig A.10, the success ratio increases linearly until the convergence curve (because of normalization) begins and reaches 1. So quadratic regression should successfully fit the convergence curve. Fig A.14 verifies this assumption.

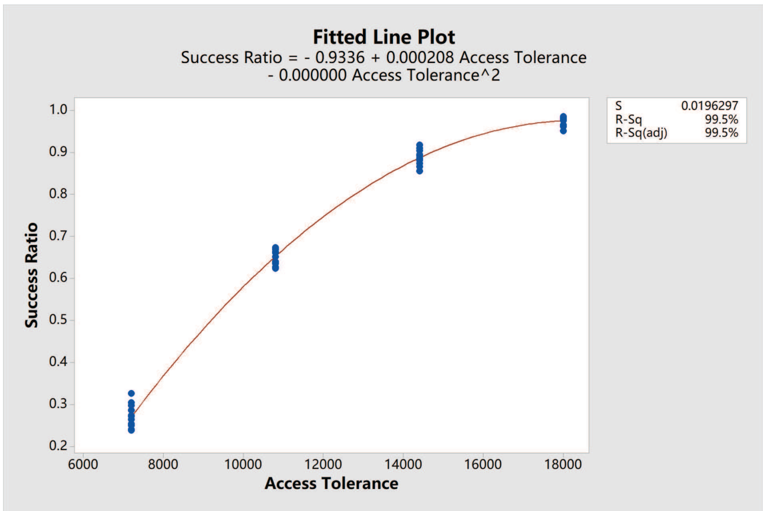


Figure A.14: Quadratic Regression Fit Line of Access Tolerance on Total Helper Access Duration

A.0.2.3 Analysis of Influence of Access Tolerance on Average Disruption Number During Each Access Request

Since the success ratio reaches 1, the average number of disruption would be constant. It is reasonable to assume the response here would be quadratic, just like the response of success ratio. Fig A.15 verifies the assumption with a nice fit.[4]

A.0.2.4 Analysis of Influence of Access Tolerance on Average Disruption Duration During Each Access Request

As for the previous assumption, access tolerance can also be called disruption tolerance. It is also like the average disruption number which becomes constant when success ratio reaches 1. So the assumption is that the average (total) disruption duration during each access request is quadratic. Fig A.16 verifies this assumption.

[4] Although there is a small pattern in the residuals, but quadratic regression is good enough (above 97%).

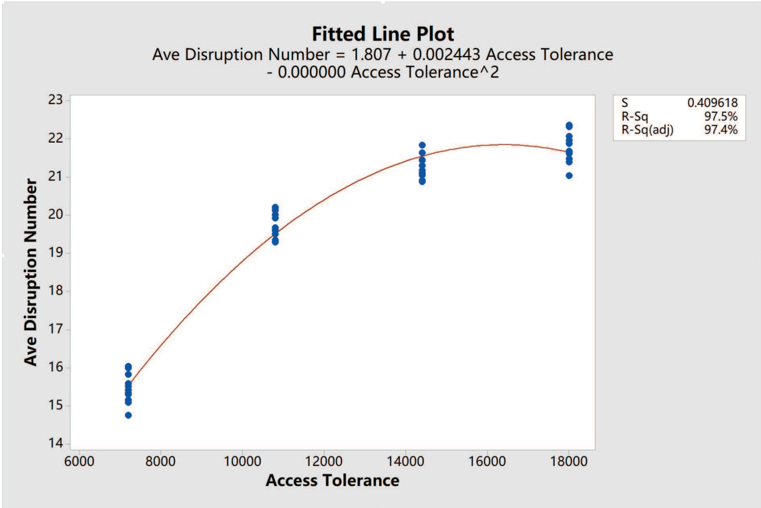


Figure A.15: Quadratic Regression Fit Line of Access Tolerance on Average Disruption Number for Each Access Request

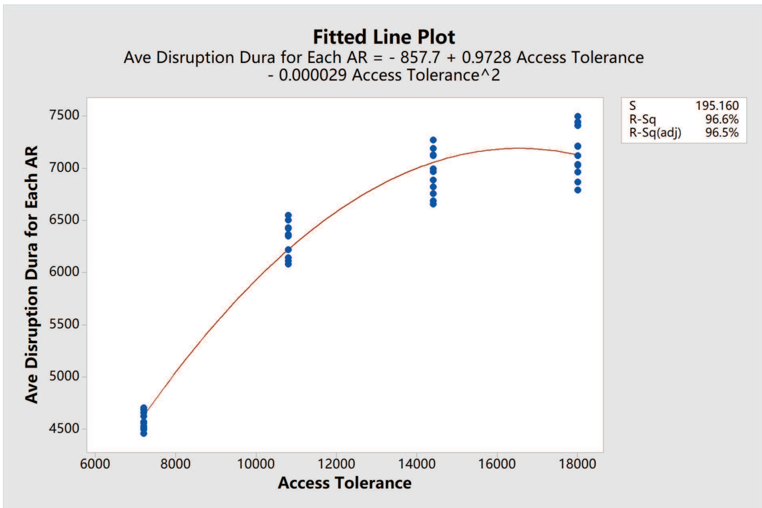


Figure A.16: Quadratic Regression Fit Line of Access Tolerance on Average Disruption Duration During Each Access Request

A.0.2.5 Conclusion on Influence of the Access Tolerance

To study the influence on success ratio, the average disruption number during each access request and disruption duration during each access request were plotted. The quadratic regression line can fits them well. Therefore, 3-level design is enough. But for influence on total helper access duration, 4-level design is necessary.

A.0.3 Determination of Levels of Helper Access Limit

A.0.3.1 Concurrent Problems

The output of the helper access limit is a perspective of concurrent problems in this system. There are three kinds of concurrent problems: 1, one helper contacts with several requester simultaneously; 2, one requester contacts with several helpers simultaneously; 3, several helpers contacts with several requesters simultaneously. The helper access limit is to check out the problem of one helper and several requesters. In this exploratory experiment, the helper access limit is set as 1, 2, 3, 4, 5.[5]

A.0.3.2 The Setting of Parameters

Many settings of parameters have been tried. If the setting of other factors are chosen as Table A-C, there will be no systematic difference[6] from 2 (helper access limit). If the transmission range is above about 120 m, results with "significance"[7] can be observed. However, the jump from 1 to 2 is still the biggest.

Table A-C: Settings of Other Factors

Factors	Values
The Number of Selected Helpers	20
Access Tolerance	10800 s
Requested Access Duration	1800 s
Transmission Range	50
Total Requester Number	100
Total Helper Number	100

The transmission range is not the only way to increase contact opportunities. The number of total requesters can also be increased, for example, 200 or 400. However, in these cases, the success ratio is far lower than 1, which means that average disruption duration may increase with the increase of helper access limit because more access

[5]How many concurrent requesters a helper can serve.

[6]The differences are mainly because of random seeds.

[7]Here, the significance is for systematic difference compared with random fluctuation. When compared with other factors discussed in this chapter, helper access limit has relatively little significance.

requests can be generated. Therefore, the setting of parameters to show the influence of helper access limit is chosen as Table A-D.

Table A-D: Settings of Other Factors

Factors	Values
The Number of Selected Helpers	20
Access Tolerance	10800 s
Requested Access Duration	1800 s
Transmission Range	150
Total Requester Number	100
Total Helper Number	100

A.0.3.3 Analysis of Influence of Helper Access Limit on Total Helper Access Duration

As shown in Fig A.17, the biggest jump is from 1 to 2. The other jumps are relatively trivial.[8]

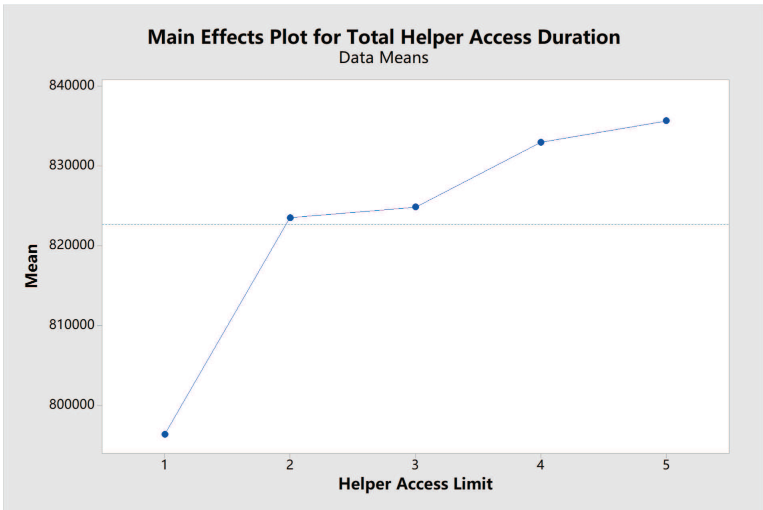


Figure A.17: Main Effect of Helper Access Limit on Total Helper Access Duration

Compared with randomness, the trend of the line is relatively not as obvious as other factors, as shown in Fig A.18. The regression with the highest R-Square (82.4%) is the cubic regression, as shown in Fig A.19. Compared to other factors, this R-Square is

[8]The difference is because of randomness.

relatively low.

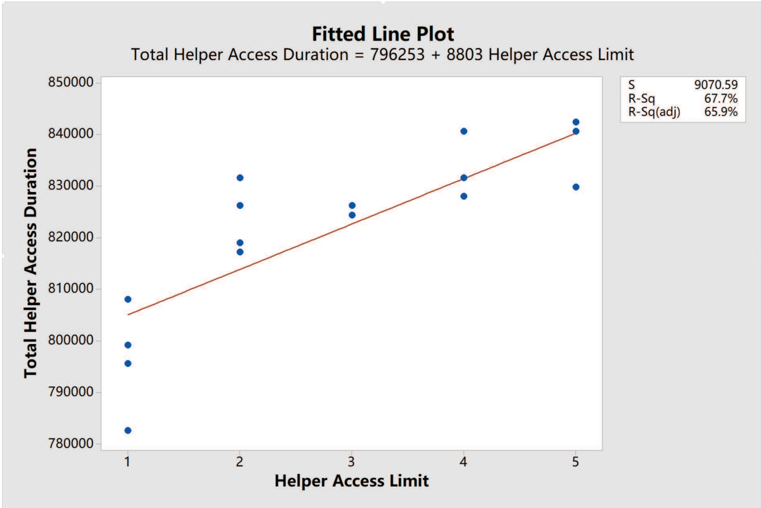


Figure A.18: Linear Regression Line of the Helper Access Limit and Total Helper Access Duration

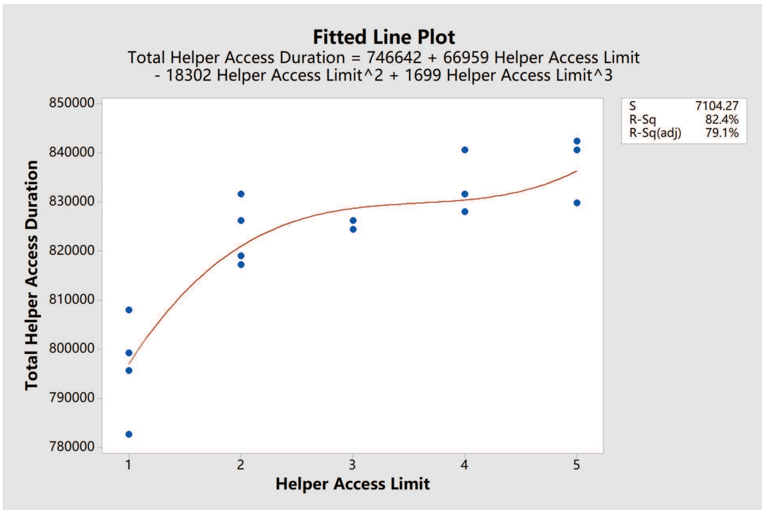


Figure A.19: Cubic Regression Line of the Helper Access Limit and Total Helper Access Duration

A.0.3.4 Analysis of Influence of Helper Access Limit on Success Ratio

In this setting, for all the helper access limit, the success ratio is 100%. The jump from 1 to 2 in total helper access duration comes from more access requests generated.

A.0.3.5 Analysis of Influence of Helper Access Limit on Average Disruption Duration for Each Access Request

As shown in Fig A.20, the cubic regression is the best regression and there are two drops in this curve, from 1 to 2 and from 4 to 5. With the increase of helper access limit, the duration of disruption can decrease so that the service durations would be shorter and there will be more access requests generated.

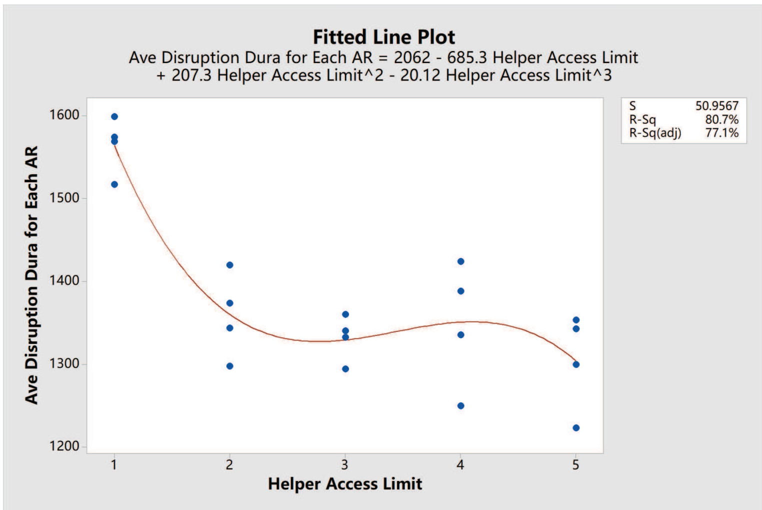


Figure A.20: Cubic Regression Line of the Helper Access Limit and Average Disruption Duration for Each Access Request

A.0.3.6 Analysis of the Influence of Helper Access Limit on Average Disruption Number for Each Access Request

There is no regression model which can fit the curve. As shown in Fig A.21, the average main effect shows there is no trend, just randomness. It make sense because when success ratio is 1, the number of disruptions only depends on how many contacts can finish a service (requested duration), which only depends on the contact durations between helpers and this requester.

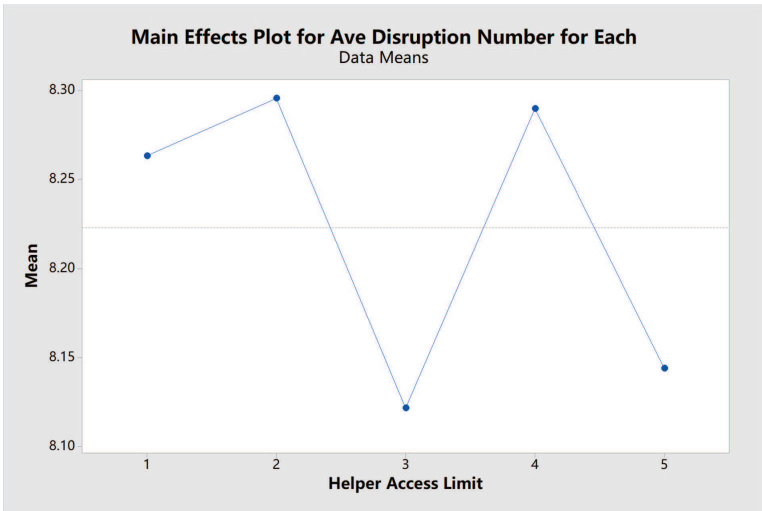


Figure A.21: Main Effect of Helper Access Limit on Average Disruption Number for Each Access Request

A.0.3.7 Conclusion on Helper Access Limit

The simulator can print out the cases that there are more than 5 candidate requesters (up to 12) simultaneously. However, these cases are relatively rare. In the current nodal density, as long as a requester contact a selected helper which can provide services for more than 1 requesters simultaneously, the requester can get the service, which leads to a small difference on all the trends.

From the observation of the main effect of access limit, a conclusion can be drawn as:

- In the current settings (movement model and nodal density), concurrent links between one helper and several requesters do exist[9], but it is not very common. Therefore, contact is almost equal to service because helpers always can serve requesters when they contact. The number of the concurrent links rarely exceeds 2.
- There are indeed trends in all outputs except average disruption number for each access request and all the curves can fit cubic regression.

[9]Even 10 concurrent links can be observed, but it is rare and has little influence on the system.

Bibliography

[1] L. Pu, X. Chen, J. Xu, and X. Fu, "D2d fogging: An energy-efficient and incentive-aware task offloading framework via network-assisted d2d collaboration," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 3887–3901, Dec 2016.

[2] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "A survey of social-based routing in delay tolerant networks: Positive and negative social effects," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 387–401, 2013.

[3] J. Zhu, J. Cao, M. Liu, Y. Zheng, H. Gong, and G. Chen, "A mobility prediction-based adaptive data gathering protocol for delay tolerant mobile sensor network," in *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*, pp. 1–5, IEEE, 2008.

[4] A. Gupta and R. K. Jha, "A survey of 5g network: Architecture and emerging technologies," *IEEE access*, vol. 3, pp. 1206–1232, 2015.

[5] G. Zhao, S. Chen, L. Qi, L. Zhao, and L. Hanzo, "Mobile-traffic-aware offloading for energy- and spectral-efficient large-scale d2d-enabled cellular networks," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 3251–3264, June 2019.

[6] J. Lee and J. H. Lee, "Performance analysis and resource allocation for cooperative d2d communication in cellular networks with multiple d2d pairs," *IEEE Communications Letters*, vol. 23, pp. 909–912, May 2019.

[7] A. Bhardwaj and S. Agnihotri, "Energy- and spectral-efficiency trade-off for d2d-multicasts in underlay cellular networks," *IEEE Wireless Communications Letters*, vol. 7, pp. 546–549, Aug 2018.

[8] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to lte-advanced networks," *IEEE Communications Magazine*, vol. 47, pp. 42–49, Dec 2009.

[9] `https://www.qualcomm.com/invention/research/projects/lte-direct`.

[10] I. Qualcomm Technologies, "Lte direct always-on device-todevice proximal discovery," 2014.

[11] T. Guo, G. Huang, J. Schormans, T. Wang, and Y. Cao, "Cap: A contact based proximity service via opportunistic device-to-device relay," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 337–342, Aug 2017.

[12] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '03, (New York, NY, USA), pp. 27–34, ACM, 2003.

[13] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pp. 244–251, ACM, 2005.

[14] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, MobiCom '98, (New York, NY, USA), pp. 85–97, ACM, 1998.

[15] Z. Jian, G. Tianhao, L. Jilong, and Y. Mingwu, "Adaptive flow control proxy for enhancing downlink tcp throughput in infrastructure wlans," in *Communications and Information Technologies (ISCIT), 2014 14th International Symposium on*, pp. 56–60, IEEE, 2014.

[16] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao, "Efficiency resource allocation for device-to-device underlay communication systems: a reverse iterative combinatorial auction based approach," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 348–358, 2013.

[17] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 136–144, 2014.

[18] `http://www.3gpp.org/specifications/releases`.

[19] `http://www.3gpp.org/news-events/3gpp-news/1455-Public-Safety`.

[20] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys Tutorials*, vol. 16, pp. 1801–1819, Fourthquarter 2014.

[21] S. Batabyal and P. Bhaumik, "Mobility models, traces and impact of mobility on opportunistic routing algorithms: A survey," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 1679–1707, thirdquarter 2015.

[22] G. Chandrasekaran, N. Wang, M. Hassanpour, M. Xu, and R. Tafazolli, "Mobility as a service (maas): A d2d-based information centric network architecture for edge-controlled content distribution," *IEEE Access*, vol. 6, pp. 2110–2129, 2018.

[23] C. P. Mayer and O. P. Waldhorst, "Offloading infrastructure using delay tolerant networks and assurance of delivery," in *Wireless Days (WD), 2011 IFIP*, pp. 1–7, IEEE, 2011.

[24] F. Rebecchi, M. D. de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data offloading techniques in cellular networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 580–603, Secondquarter 2015.

[25] B. Han, P. Hui, V. Kumar, M. V. Marathe, G. Pei, and A. Srinivasan, "Cellular traffic offloading through opportunistic communications: a case study," in *Proceedings of the 5th ACM workshop on Challenged networks*, pp. 31–38, ACM, 2010.

[26] W. C. Kim and R. A. Mauborgne, *Blue ocean strategy, expanded edition: How to create uncontested market space and make the competition irrelevant.* Harvard business review Press, 2014.

[27] A. Kumbhar, F. Koohifar, I. Güvenç, and B. Mueller, "A survey on legacy and emerging technologies for public safety communications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 97–124, 2017.

[28] J. Hamari, M. Sjöklint, and A. Ukkonen, "The sharing economy: Why people par-

ticipate in collaborative consumption," *Journal of the Association for Information Science and Technology*, vol. 67, no. 9, pp. 2047–2059, 2016.

[29] G. Dudley, D. Banister, and T. Schwanen, "The rise of uber and regulating the disruptive innovator," *The political quarterly*, vol. 88, no. 3, pp. 492–499, 2017.

[30] G. Zervas, D. Proserpio, and J. W. Byers, "The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry," *Journal of marketing research*, vol. 54, no. 5, pp. 687–705, 2017.

[31] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo, "Real time pothole detection using android smartphones with accelerometers," in *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, pp. 1–6, June 2011.

[32] B. Zhang, Y. Li, D. Jin, P. Hui, and Z. Han, "Social-aware peer discovery for d2d communications underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 2426–2439, May 2015.

[33] H. Song, R. Srinivasan, T. Sookoor, and S. Jeschke, *Mobility as a Service*, pp. 912–. Wiley Telecom, 2017.

[34] D. C. Montgomery, *Design and analysis of experiments.* John wiley & sons, 2017.

[35] B. Han, P. Hui, V. Kumar, M. V. Marathe, G. Pei, and A. Srinivasan, "Cellular traffic offloading through opportunistic communications: a case study," in *Proceedings of the 5th ACM workshop on Challenged networks*, pp. 31–38, ACM, 2010.

[36] D. Chatzopoulos, M. Ahmadi, S. Kosta, and P. Hui, "Openrp: a reputation middleware for opportunistic crowd computing," *IEEE Communications Magazine*, vol. 54, pp. 115–121, July 2016.

[37] K. Sucipto, D. Chatzopoulos, S. Kosta, and P. Hui, "Keep your nice friends close, but your rich friends closer - computation offloading using NFC," *CoRR*, vol. abs/1612.03000, 2016.

[38] D. Chatzopoulos and P. Hui, "Readme: A real-time recommendation system for mobile augmented reality ecosystems," in *Proceedings of the 2016 ACM on Multimedia Conference*, MM '16, (New York, NY, USA), pp. 312–316, ACM, 2016.

[39] C. Bermejo, D. Chatzopoulos, and P. Hui, "How sustainable is social based mobile

crowdsensing? an experimental study," in *2016 IEEE 24th International Conference on Network Protocols (ICNP)*, pp. 1–6, Nov 2016.

[40] N. Allurwar, B. Nawale, and S. Patel, "Beacon for proximity target marketing," *Int. J. Eng. Comput. Sci*, vol. 15, pp. 16359–16364, 2016.

[41] D. Chatzopoulos, S. Gujar, B. Faltings, and P. Hui, "Localcoin: An ad-hoc payment scheme for areas with high connectivity: Poster," in *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '16, (New York, NY, USA), pp. 365–366, ACM, 2016.

[42] J. Li, Z. Cai, J. Wang, M. Han, and Y. Li, "Truthful incentive mechanisms for geographical position conflicting mobile crowdsensing systems," *IEEE Transactions on Computational Social Systems*, vol. PP, no. 99, pp. 1–11, 2018.

[43] A. Capponi, C. Fiandrino, D. Kliazovich, and P. Bouvry, "Energy efficient data collection in opportunistic mobile crowdsensing architectures for smart cities," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 307–312, May 2017.

[44] T. Wang, P. Li, X. Wang, Y. Wang, T. Guo, and Y. Cao, "A comprehensive survey on mobile data offloading in heterogeneous network," *Wireless Networks*, pp. 1–12, 2017.

[45] W. Gao, G. Cao, A. Iyengar, and M. Srivatsa, "Supporting cooperative caching in disruption tolerant networks," in *2011 31st International Conference on Distributed Computing Systems*, pp. 151–161, June 2011.

[46] P. Mrazovic and M. Matskin, "Mobics: Mobile platform for combining crowdsourcing and participatory sensing," in *2015 IEEE 39th Annual Computer Software and Applications Conference*, vol. 2, pp. 553–562, July 2015.

[47] L. Pu, X. Chen, J. Xu, and X. Fu, "Crowd foraging: A qos-oriented self-organized mobile crowdsourcing framework over opportunistic networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 848–862, April 2017.

[48] J. De Benedetto, P. Bellavista, and L. Foschini, "Proximity discovery and data dissemination for mobile crowd sensing using lte direct," *Computer Networks*, 2017.

[49] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future

challenges," *IEEE Communications Magazine*, vol. 49, no. 11, 2011.

[50] C. Fiandrino, A. Capponi, G. Cacciatore, D. Kliazovich, U. Sorger, P. Bouvry, B. Kantarci, F. Granelli, and S. Giordano, "Crowdsensim: a simulation platform for mobile crowdsensing in realistic urban environments," *IEEE Access*, vol. 5, pp. 3490–3503, 2017.

[51] W. Gong, B. Zhang, and C. Li, "Location-based online task scheduling in mobile crowdsensing," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pp. 1–6, Dec 2017.

[52] P. Nguyen and K. Nahrstedt, "Context-aware crowd-sensing in opportunistic mobile social networks," in *2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems*, pp. 477–478, Oct 2015.

[53] S. Yang, X. Qi, F. Wu, X. Gao, and G. Chen, "Bandit user selection algorithm for budgeted and time-limited mobile crowdsensing," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pp. 1–6, Dec 2017.

[54] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. S. Shen, "Providing task allocation and secure deduplication for mobile crowdsensing via fog computing," *IEEE Transactions on Dependable and Secure Computing*, vol. PP, no. 99, pp. 1–1, 2018.

[55] C. Tanas and J. Herrera-Joancomartí, "Crowdsensing simulation using ns-3," in *Citizen in Sensor Networks*, pp. 47–58, Springer, 2014.

[56] M. Liu and Y. Liu, "Price-based distributed offloading for mobile-edge computing with computation capacity constraints," *IEEE Wireless Communications Letters*, vol. PP, no. 99, pp. 1–1, 2017.

[57] X. Chen and J. Zhang, "When d2d meets cloud: Hybrid mobile task offloadings in fog computing," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.

[58] D. Chatzopoulos, M. Ahmadi, S. Kosta, and P. Hui, "Flopcoin: A cryptocurrency for computation offloading," *IEEE Transactions on Mobile Computing*, vol. PP, no. 99, pp. 1–1, 2017.

[59] Z. Hong, Z. Wang, W. Cai, and V. C. M. Leung, "Connectivity-aware task outsourcing and scheduling in d2d networks," in *2017 26th International Conference*

on Computer Communication and Networks (ICCCN), pp. 1–9, July 2017.

[60] Y. Shi, S. Chen, and X. Xu, "Maga: A mobility-aware computation offloading decision for distributed mobile cloud computing," *IEEE Internet of Things Journal*, vol. 5, pp. 164–174, Feb 2018.

[61] L. Liu, Z. Chang, X. Guo, and T. Ristaniemi, "Multi-objective optimization for computation offloading in mobile-edge computing," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, pp. 832–837, July 2017.

[62] F. Messaoudi, A. Ksentini, and P. Bertin, "On using edge computing for computation offloading in mobile network," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pp. 1–7, Dec 2017.

[63] S. H. Seo and J. Straub, "Comparative analysis of graph partitioning algorithms in context of computation offloading," in *2017 IEEE International Conference on Electro Information Technology (EIT)*, pp. 1–7, May 2017.

[64] A. Zanni, S. Y. Yu, P. Bellavista, R. Langar, and S. Secci, "Automated selection of offloadable tasks for mobile computation offloading in edge computing," in *2017 13th International Conference on Network and Service Management (CNSM)*, pp. 1–5, Nov 2017.

[65] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 176–189, Jan 2016.

[66] S. Goebbels and R. Jennen, "Enhancements in wireless broadband networks using smart caching an analytical evaluation," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pp. 1–5, IEEE, 2008.

[67] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *INFOCOM, 2012 Proceedings IEEE*, pp. 1107–1115, IEEE, 2012.

[68] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 176–189, Jan 2016.

[69] M. Erol-Kantarci and S. Sukhmani, "Caching and computing at the edge for mobile augmented reality and virtual reality (ar/vr) in 5g," in *Ad Hoc Networks*, pp. 169–177, Springer, 2018.

[70] W. Zhang, B. Han, P. Hui, V. Gopalakrishnan, E. Zavesky, and F. Qian, "Cars: Collaborative augmented reality for socialization," in *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications*, pp. 25–30, ACM, 2018.

[71] A. Brodersen, S. Scellato, and M. Wattenhofer, "Youtube around the world: Geographic popularity of videos," in *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, (New York, NY, USA), pp. 241–250, ACM, 2012.

[72] J.-P. Laulajainen, Å. Arvidsson, T. Ojala, J. Seppänen, and M. Du, "Study of youtube demand patterns in mixed public and campus wifi network," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International*, pp. 635–641, IEEE, 2014.

[73] D. Chatzopoulos, K. Sucipto, S. Kosta, and P. Hui, "Video compression in the neighborhood: An opportunistic approach," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2016.

[74] A. Irfan and A. Khalique, "A comparative demonstration and analysis of file sharing applications on android mobile devices," *International Journal of Computer Applications*, vol. 156, no. 12, 2016.

[75] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, pp. 606–620, June 2007.

[76] S. Hoteit, S. Secci, S. Sobolevsky, C. Ratti, and G. Pujolle, "Estimating human trajectories and hotspots through mobile phone data," *Computer Networks*, vol. 64, pp. 296–307, 2014.

[77] F. Calabrese, F. C. Pereira, G. Di Lorenzo, L. Liu, and C. Ratti, "The geography of taste: analyzing cell-phone mobility and social events," in *International conference on pervasive computing*, pp. 22–37, Springer, 2010.

[78] R. R. Roy, *Handbook of mobile ad hoc networks for mobility models*, vol. 170.

Springer, 2011.

[79] C. Bettstetter, "Mobility modeling in wireless networks: categorization, smooth movement, and border effects," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 3, pp. 55–66, 2001.

[80] D. M. Blough, G. Resta, and P. Santi, "A statistical analysis of the long-run node spatial distribution in mobile ad hoc networks," *Wireless Networks*, vol. 10, no. 5, pp. 543–554, 2004.

[81] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the levy-walk nature of human mobility," *IEEE/ACM Transactions on Networking*, vol. 19, pp. 630–643, June 2011.

[82] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for pcs networks," in *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp. 1377–1384, IEEE, 1999.

[83] C. Bettstetter, "Smooth is better than sharp: a random mobility model for simulation of wireless networks," in *Proceedings of the 4th ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, pp. 19–27, ACM, 2001.

[84] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," in *Proceedings of the 9th annual international conference on Mobile computing and networking*, pp. 217–229, ACM, 2003.

[85] F. Ekman, A. Keränen, J. Karvo, and J. Ott, "Working day movement model," in *Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models*, MobilityModels '08, (New York, NY, USA), pp. 33–40, ACM, 2008.

[86] S. C. Nelson, A. F. Harris III, and R. Kravets, "Event-driven, role-based mobility in disaster recovery networks," in *Proceedings of the second ACM workshop on Challenged networks*, pp. 27–34, ACM, 2007.

[87] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang, "A group mobility model for ad hoc wireless networks," in *Proceedings of the 2nd ACM international workshop*

on Modeling, analysis and simulation of wireless and mobile systems, pp. 53–60, ACM, 1999.

[88] K. Blakely and B. Lowekamp, "A structured group mobility model for the simulation of mobile ad hoc networks," in *Proceedings of the Second International Workshop on Mobility Management &Amp; Wireless Access Protocols*, MobiWac '04, (New York, NY, USA), pp. 111–118, ACM, 2004.

[89] M. Sánchez and P. Manzoni, "Anejos: a java based simulator for ad hoc networks," *Future generation computer systems*, vol. 17, no. 5, pp. 573–583, 2001.

[90] P. Hui, E. Yoneki, S. Y. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture*, p. 7, ACM, 2007.

[91] P. Hui, E. Yoneki, J. Crowcroft, and S.-Y. Chan, "Identifying social communities in complex communications for network efficiency," in *International Conference on Complex Sciences*, pp. 351–363, Springer, 2009.

[92] S.-Y. Chan, P. Hui, and K. Xu, "Community detection of time-varying mobile social networks," in *International Conference on Complex Sciences*, pp. 1154–1159, Springer, 2009.

[93] A. Jabbar, "Aerorp: A geolocation assisted aeronautical routing protocol for highly dynamic telemetry environments," in *International Telemetering Conference Proceedings*, International Foundation for Telemetering, 2009.

[94] K. Zou, M. Wang, K. Yang, J. Zhang, W. Sheng, Q. Chen, and X. You, "Proximity discovery for device-to-device communications over a cellular network," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 98–107, 2014.

[95] F. H. P. Fitzek, J. Heide, M. V. Pedersen, and M. Katz, "Implementation of network coding for social mobile clouds [applications corner]," *IEEE Signal Processing Magazine*, vol. 30, pp. 159–164, Jan 2013.

[96] A. Keränen, J. Ott, and T. Kärkkäinen, "The one simulator for dtn protocol evaluation," in *Proceedings of the 2nd international conference on simulation tools and techniques*, p. 55, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009.

[97] R. Jain, *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling.* John Wiley & Sons, 1990.

[98] R. A. Fisher, *The design of experiments.* Oliver And Boyd; Edinburgh; London, 1937.

[99] K. E. Atkinson, *An introduction to numerical analysis.* John Wiley & Sons, 2008.

[100] S. Andreev, J. Hosek, T. Olsson, K. Johnsson, A. Pyattaev, A. Ometov, E. Olshannikova, M. Gerasimenko, P. Masek, Y. Koucheryavy, and T. Mikkonen, "A unifying perspective on proximity-based cellular-assisted mobile social networking," *IEEE Communications Magazine*, vol. 54, pp. 108–116, April 2016.

[101] A. Asadi and V. Mancuso, "Wifi direct and lte d2d in action," in *Wireless Days (WD), 2013 IFIP*, pp. 1–8, IEEE, 2013.

[102] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in sparse vehicular ad hoc wireless networks," *IEEE journal on Selected Areas in Communications*, vol. 25, no. 8, 2007.

[103] Y. Cao, N. Wang, Z. Sun, and H. Cruickshank, "A reliable and efficient encounter-based routing framework for delay/disruption tolerant networks," *IEEE Sensors Journal*, vol. 15, no. 7, pp. 4004–4018, 2015.

[104] X. Lin, J. G. Andrews, and A. Ghosh, "A comprehensive framework for device-to-device communications in cellular networks," *CoRR*, vol. abs/1305.4219, 2013.

[105] A. Orsino, A. Ometov, G. Fodor, D. Moltchanov, L. Militano, S. Andreev, O. N. Yilmaz, T. Tirronen, J. Torsner, G. Araniti, *et al.*, "Effects of heterogeneous mobility on d2d-and drone-assisted mission-critical mtc in 5g.," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 79–87, 2017.

[106] W. Jiang, H. Cao, L. Goratti, M. Wiemeler, and T. Kaiser, "Opportunistic relaying over aerial-to-terrestrial and device-to-device radio channels," in *Communications Workshops (ICC), 2014 IEEE International Conference on*, pp. 206–211, IEEE, 2014.

[107] A. Hess, K. A. Hummel, W. N. Gansterer, and G. Haring, "Data-driven human mobility modeling: a survey and engineering guidance for mobile networking,"

ACM Computing Surveys (CSUR), vol. 48, no. 3, p. 38, 2016.

[108] K. Miettinen, *Nonlinear Multiobjective Optimization.* International Series in Operations Research & Management Science, Springer US, 1999.

[109] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "Platemo: A matlab platform for evolutionary multi-objective optimization [educational forum]," *IEEE Computational Intelligence Magazine*, vol. 12, no. 4, pp. 73–87, 2017.

[110] K. Deb, *Multi-objective optimization using evolutionary algorithms*, vol. 16. John Wiley & Sons, 2001.

[111] S. R. Garzon, B. Deva, B. Hanotte, and A. Küpper, "Catles: a crowdsensing-supported interactive world-scale environment simulator for context-aware systems," in *Mobile Software Engineering and Systems (MOBILESoft), 2016 IEEE/ACM International Conference on*, pp. 77–87, IEEE, 2016.