Providing Location-privacy in Opportunity Mobile Social Network

by

Rui Huang

A Thesis submitted to the Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements for the degree of

**MASTER OF APPLIED SCIENCE**

in Electrical and Computer Engineering

Ottawa-Carleton Institute of Electrical and Computer Engineering

University of Ottawa

Ottawa, Canada

September 3, 2017

© Rui Huang, Ottawa, Canada, 2017

Abstract

Users face location-privacy risks when accessing Location-Based Services (LBSs) in an Opportunistic Mobile Social Networks (OMSNs). In order to protect the original requester’s identity and location, we propose two location privacy obfuscation protocols utilizing social ties between users.

The first one is called Multi-Hop Location-Privacy Protection (MHLPP) protocol. To increase chances of completing obfuscation operations, users detect and make contacts with one-hop or multi-hop neighbor friends in social networks. Encrypted obfuscation queries avoid users learning important information especially the original requester’s identity and location except for trusted users. Simulation results show that our protocol can give a higher query success ratio compared to its existing counterpart.

Another protocol is called Appointment Card protocol (ACP). To facilitate the obfuscation operations of queries, we introduce messages called the Appointment Card (AC). The original requesters can send their queries to the LBS directly using the information in the AC, so that the query time cost of ACP is similar as no-privacy protocols ensuring that the original requester being not detected by the LBS. Also, a path for reply message is built when the query is sent, thus saving lots of time in replying querries. Simulation results show that our protocol can reach a higher query success ratio comparing to the existing protocol.

We also create a new OMSN simulator involving social network to provide a better evaluation for OMSN protocols. It shows more efficient and reliable performance in our experiments.

Acknowledgements

First, I would like to express my sincere gratitude to Prof. Amiya Nayak for his continuous support, motivation, advice and immense knowledge throughout my thesis work. I could not have imagined having a better mentor for my Masters study.

Also, I would like to thank my family and friends, especially Coco Lin, for providing me with unfailing support and encouragement throughout the process of researching and writing my thesis. This accomplishment would not have been possible without them. Thank you.

Table of Contents

[Abstract i](#_Toc500498519)

[Acknowledgements ii](#_Toc500498520)

[Table of Contents iii](#_Toc500498521)

[List of Figures vi](#_Toc500498522)

[List of Tables viii](#_Toc500498523)

[Chapter 1 Introduction 1](#_Toc500498524)

[1.1. Motivation 1](#_Toc500498525)

[1.2. Objectives 2](#_Toc500498526)

[1.3. Contribution 2](#_Toc500498527)

[1.4. Thesis Organization 4](#_Toc500498528)

[Chapter 2 Background and Related Work 5](#_Toc500498529)

[2.1. Mobile Ad hoc networks 5](#_Toc500498530)

[2.2. Binary Spray and Wait 5](#_Toc500498531)

[2.3. Location-Based Services 5](#_Toc500498532)

[2.4. Location-privacy protocols 6](#_Toc500498533)

[Chapter 3 Multi-Hop Location-Privacy Protection 8](#_Toc500498534)

[3.1. System Model 8](#_Toc500498535)

[3.2. Details of MHLPP 9](#_Toc500498536)

[3.3. Requirement parameters 13](#_Toc500498537)

[3.4. Privacy Analysis 15](#_Toc500498538)

[3.5. Complexity discussion 16](#_Toc500498539)

[3.6. Performance Analysis 17](#_Toc500498540)

[3.6.1. Query success ratio 19](#_Toc500498541)

[3.6.2. Number of Hops 21](#_Toc500498542)

[3.6.3. Security 23](#_Toc500498543)

[Chapter 4 Appointment Card Protocol 26](#_Toc500498544)

[4.1. System Model 26](#_Toc500498545)

[4.2. Appointment Card Protocol Overview 27](#_Toc500498546)

[4.3. Appointment Card 28](#_Toc500498547)

[4.4. AC life cycle 30](#_Toc500498548)

[4.5. System parameters 31](#_Toc500498549)

[4.5.1. Obfuscation distance 31](#_Toc500498550)

[4.5.2. Friends obfuscation distance 32](#_Toc500498551)

[4.5.3. Distributing segment 33](#_Toc500498552)

[4.5.4. Generating Period 33](#_Toc500498553)

[4.5.5. AC timeout 33](#_Toc500498554)

[4.5.6. Avoiding time 33](#_Toc500498555)

[4.6. Protocol Details 34](#_Toc500498556)

[4.6.1. Generating appointment cards 34](#_Toc500498557)

[4.6.2. Exchange distributing appointment cards 34](#_Toc500498558)

[4.6.3. Exchange ready appointment cards 36](#_Toc500498559)

[4.6.4. Sending queries 37](#_Toc500498560)

[4.6.5. Sending replies 38](#_Toc500498561)

[4.6.5.1. The LBSP part 38](#_Toc500498562)

[4.6.6. The first agent 39](#_Toc500498563)

[4.6.7. Intermediate agents 40](#_Toc500498564)

[4.6.8. The last agent 41](#_Toc500498565)

[4.7. Appointment number 43](#_Toc500498566)

[4.7.1. Creator appointment number 43](#_Toc500498567)

[4.7.2. Agent appointment number 43](#_Toc500498568)

[4.7.3. Timeout 43](#_Toc500498569)

[4.8. Example 44](#_Toc500498570)

[4.8.1. Exchanging appointment cards 46](#_Toc500498571)

[4.8.2. Queries and Replies 57](#_Toc500498572)

[4.9. Experiment 61](#_Toc500498573)

[4.9.1. Average query success ratio 61](#_Toc500498574)

[4.9.2. Average reply success ratio 64](#_Toc500498575)

[4.9.3. Total number of query relay 65](#_Toc500498576)

[4.9.4. Memory cost 67](#_Toc500498577)

[4.9.5. Distributing appointment cards 69](#_Toc500498578)

[Chapter 5 OMSN routing protocol simulator 71](#_Toc500498579)

[5.1. Vertex reduce algorithm 71](#_Toc500498580)

[5.1.1. Ignorable vertex and reserved vertex 72](#_Toc500498581)

[5.1.2. Vertex reducing 73](#_Toc500498582)

[5.1.2.1. Remove ignorable vertices 73](#_Toc500498583)

[5.1.2.2. Tidy reserved vertices connections 74](#_Toc500498584)

[5.1.2.3. Iterations 74](#_Toc500498585)

[5.1.3. Vertex assembling 75](#_Toc500498586)

[5.2. Finding nodes in a range 76](#_Toc500498587)

[5.2.1. Static nodes 77](#_Toc500498588)

[5.2.2. Mobile nodes 78](#_Toc500498589)

[5.2.3. Complexities 78](#_Toc500498590)

[REFERENCES: 80](#_Toc500498591)

List of Figures

[Figure 3.1 The selection of the next friend 14](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498592)

[Figure 3.2 The number of encryption with various inner radius 17](#_Toc500498593)

[Figure 3.3 The query success ratio comparison between HSLPO and MHLPP 20](#_Toc500498594)

[Figure 3.4 The number of hops comparison between HSLPO and MHLPP 23](#_Toc500498595)

[Figure 3.5 Locating probability entropy comparison between HSLPO and MHLPSP 25](#_Toc500498596)

[Figure 4.1 Whole Process Example 27](#_Toc500498597)

[Figure 4.2 AC's life cycle 30](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498598)

[Figure 4.3 obfuscation distance 31](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498599)

[Figure 4.4 Friends-obfuscation distance 32](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498600)

[Figure 4.5 constitute query 38](#_Toc500498601)

[Figure 4.6 constitute replies 38](#_Toc500498602)

[Figure 4.7 the reply of the first agent 40](#_Toc500498603)

[Figure 4.8 the reply of the second agent 41](#_Toc500498604)

[Figure 4.9 The reply of the last agent 43](#_Toc500498605)

[Figure 4.10 David's query 58](#_Toc500498606)

[Figure 4.11 Elizabeth forwards a reply 59](#_Toc500498607)

[Figure 4.12 Bob forwards a reply 59](#_Toc500498608)

[Figure 4.13 Charlie forwards a reply 60](#_Toc500498609)

[Figure 4.14 average query success ratio (10 minutes) 62](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498610)

[Figure 4.15 average query success ratio (20 minutes) 63](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498611)

[Figure 4.16 average query success ratio with 50-meters communication ratio 64](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498612)

[Figure 4.17 average reply success ratio with 50-meters communication ratio 65](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498613)

[Figure 4.18 average total number of forwarding queries at 20 minutes 66](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498614)

[Figure 4.19 average query buffer needed at 20 minutes 67](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498615)

[Figure 4.20 average number of carried queries per user 69](#_Toc500498616)

[Figure 4.21 Average number of ready ACs per user 70](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498617)

[Figure 4.22 Average number of exchanging ACs per minute 70](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498618)

[Figure 5.1 Remove ignorable vertices 73](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498619)

[Figure 5.2 Tidy reserved connections 74](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498620)

[Figure 5.3 grid size for still nodes 77](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498621)

[Figure 5.4 length of grid 78](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498622)

List of Tables

[Table 3.1MHLPP Symbols 10](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498623)

[Table 3.2 MHLPP simulation parameters 18](#_Toc500498624)

[Table 4.1 Symbols 28](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498625)

[Table 4.2 Appointment Card 29](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498626)

[Table 4.3 Important system parameters 31](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498627)

[Table 4.4 relay table entries 36](#_Toc500498628)

[Table 4.5 reply table entries of the first agent 39](#_Toc500498629)

[Table 4.6 reply table entries of the second agent 40](#_Toc500498630)

[Table 4.7 reply table entries of the last agent 42](file:///D:\rayseesea\github\Rui_Thesis\Rui_Huang_thesis.docx#_Toc500498631)

[Table 4.8 example parameters 44](#_Toc500498632)

[Table 4.9 Example event 45](#_Toc500498633)

[Table 4.10 Example initial state 45](#_Toc500498634)

[Table 4.11 User X and Y's AC lists 46](#_Toc500498635)

[Table 4.12 Elizabeth and Bob's AC lists at time 1 47](#_Toc500498636)

[Table 4.13 Elizabeth and Bob's relay table at time 1 47](#_Toc500498637)

[Table 4.14 Bob and Charlie's AC lists at time 2 48](#_Toc500498638)

[Table 4.15 Bob and Charlie's relay table at time 2 48](#_Toc500498639)

[Table 4.16 Charlie and Elizabeth's AC lists at time 3 49](#_Toc500498640)

[Table 4.17 Charlie and Elizabeth's relay table at time 3 49](#_Toc500498641)

[Table 4.18 Elizabeth and Alice's AC lists at time 4 50](#_Toc500498642)

[Table 4.19 Elizabeth and Alice's relay table at time 4 51](#_Toc500498643)

[Table 4.20 Bob and Charlie's AC lists at time 5 52](#_Toc500498644)

[Table 4.21 Bob and Charlie's relay table at time 5 52](#_Toc500498645)

[Table 4.22 Alice and Charlie's AC lists at time 6 53](#_Toc500498646)

[Table 4.23 Alice and Charlie's relay table at time 6 53](#_Toc500498647)

[Table 4.24 Charlie and David's AC lists at time 7 54](#_Toc500498648)

[Table 4.25 Charlie and David's relay table at time 7 55](#_Toc500498649)

[Table 4.26 David and Elizabeth's AC lists at time 8 56](#_Toc500498650)

[Table 4.27 David and Elizabeth's relay table at time 8 57](#_Toc500498651)

[Table 4.28 Elizabeth's reply table entry 58](#_Toc500498652)

[Table 4.29 Bob's reply table entry 59](#_Toc500498653)

[Table 4.30 Charlie's reply table entry 60](#_Toc500498654)

# 

# Introduction

Location-privacy is becoming a major concern in the Opportunistic Mobile Social Network (OMSN) which is a kind of Delay Tolerant Networks (DTNs) [1] featuring lack of continuous connectivity. More specifically, in OMSNs, it is not necessary for senders to have an end-to-end routing path to their destinations. Users make contacts when they encounter each other. LBSs are common applications in OMSNs and they are widely used in "military and government industries, emergency services and the commercial sector" [4], especially after the proliferation of localization technologies, like GPS. LBS users often send their location to LBS providers. Many people access to LBSs with their portable devices, so their locations are also bound to the devices. In this case, LBS users face a continuous risk that their location may be leaked from LBS applications. This makes people unwilling to use LBSs. Thus, protect location privacy has been a critical issue in LBSs.

## Motivation

In this thesis, we propose new protocols to protect OMSN users from leaking location-privacy, when they use LBS.

Location-Based Services (LBSs) uses the location information of users. LBS can be considered as “two types of application design: push and pull services” [4]. For example, you may receive an advertisement when you enter an area, which is a push services; if you look for the nearest restaurant, you must pull information from the network. It is obvious that you must expose a location to somewhere if you use LBS. When you are finding the nearest restaurant, you tell the LBS provider your location and your next destination where might be a restaurant nearby. If attackers can access the LBS provider’s data base, they can learn that information. Then you may receive a lot of advertisements from surrounding restaurants.

The above inference attack just bothers you with advertisements, but it becomes more dangerous when you send more queries to the LBS provider. If you use the former LBS application 10 times in a day, attackers can learn your trace from this information so that they can infer more information including your identity and home address. “Hoh used a database of week-long GPS traces from 239 drivers in the Detroit, MI area. Examining a subset of 65 drivers, their home-finding algorithm was able to find plausible home locations of about 85%, although the authors did not know the actual locations of the drivers’ homes” [5]. Therefore, the LBS provider is considered as a semi-trusted party so that users must protect their location-privacy when they use LBS applications.

The Opportunistic Mobile Social Network (OMSN) is defined “as decentralized opportunistic communication networks formed among human carried mobile devices that take advantage of mobility and social networks to create new opportunities for exchanging information and mobile ad hoc social networking” [2]. In other words, OMSN is a kind of mobile ad hoc network, and the information and technology of social network is also used in it. A typical kind of mobile device which can be carried by people, especially pedestrians, is smartphones. Since the WiFi, Bluetooth equipment on smartphones can be used for discovering other devices and direct communication between devices, users can communicate with others who are in their communication ranges without any infrastructure. Therefore, users often carry many messages and forward these messages when they encounter other users, which is so called “store-carry-forward”.

## Objectives

Disconnection, decentralization and highly delays are obvious features for OMSN, so that a decentralized strategy benefits for the location-privacy protection protocol in OMSN. Besides, decreasing interactions between users and servicers can significantly cut down the time cost. To prevent attackers from learning users’ privacy information, the protocol should obfuscate users’ location and hide their identities.

## Contribution

We propose two new decentralized location-privacy protecting protocols for OMSN, so that they do not need a three-party server which is used to obfuscate queries for users. We also create a new simulator for OMSNs called OMSN Routing protocols Simulator (ORS) to evaluate our protocols.

The first protocol which we propose is a distributed location-privacy algorithm, called Multi-Hop Location-Privacy Protection (MHLPP). It guarantees location-privacy and achieve a higher query success ratio. The introduction of social networks enables us to hide the original requester’s information behind his friends. When a user wants to send a query, he starts to look for friends based on information in his social network. He sends his query to the first encountered friend who is then responsible for forwarding the query to the intended location. This friend can also pass the query to one of his friends when they encounter. When the distance between the user carrying this query and the original requester exceeds a specified threshold, the user sends the query to the LBS server directly without having to find a friend to pass on. At that time, he also replaces the original requester’s information with its own identity and location, which enables the LBS server to receive the query without any information about the original requester. After receiving the query, the LBS server replies to the last friend (the user sending the query to the LBS) who then transmits it to the original requester.

Our second protocol is also a distributed location-privacy algorithm called Appointment Card Protocol (ACP), which aims to guarantee location-privacy and reach a higher query success ratio. The introduction of social networks enables us to hide the original requester’s information behind his friends. We introduce the Appointment Card (AC) as a kind of intermediary which records a serial of agencies, so that the query of the original requester uses the identity of the first agency in the AC, also the reply of the query can be delivered back to the original requester along the serial of agencies. The last agency called trusted agency in the AC is a user in the social tie of the original requester, in other words, the last agency is a friend (or a friend of friends) of the original requester. The trusted agency separates the strangers in the AC and the original requester, so that no stranger knows the identity of the original requester.

Our new simulator ORS is inspired by a well-known simulator the ONE [26], which is used by a lot of researchers to test their models and protocols in DTN. The major reason why we create a new simulator is that the ONE is not designed for OMSN, so that there is no social network context in it. Adding social network information into the ONE simulator must modify the basic structure of it, which might import risks for the correctness of our experiments. We also use more reliable and efficient algorithm in our simulator, so that our simulator is about 2 times faster than the ONE in the case where we test our protocols.

## Thesis Organization

This thesis is organized as follows:

Chapter 2 describes the concept of Mobile Ad hoc networks and a basic no-privacy protocol Binary Spray and Wait (BSW) [22]. We give an overview of the LBS and the existing location-privacy protocols.

Chapter 3 describes our first protocol MHLPP, which is explained in detail through architectural and mathematical instances. Comparison is made with a similar location-privacy protocol.

Chapter 4 elaborates on the second protocol ACP. We introduce a description of how it works and how it enhances the location-privacy. We also use an example to show the whole process.

Chapter 5 shows our new simulator. Also, we introduce the algorithms which are used in the simulator.

# Background and Related Work

## Mobile Ad hoc networks

The Mobile Ad hoc network (MANET) is a decentralized type of wireless network, and there is no pre-existing infrastructure managing the network. Users in the MANET can move independently and only communicate when they encounter each other. In other words, users cannot communicate if there is no user in their own communication radius. The MANET can be viewed as a kind of delay-tolerant network (DTN) which lacks continuous network connectivity, so that DTN can hardly establish instantaneous end-to-end paths. Therefore, store-and-forward is a typical strategy for users to deliver messages in DTN or MANET. Uses forward messages when they encounter others, or they just store messages.

### Delay-Tolerant Network

Regular networks, e.g. Internet, always have end-to-end paths. As shown in Figure 2.1, the node S and the node D are connected by A, B and C. Their maximum round-trip time is not excessive, and their drop probability is also small. Comparing to the regular network, there is a class of challenged networks [1] which lacks end-to-end path and suffering from high latency and long queuing times. As shown in Figure 2.2, when the node B is not working, there is no connection between the node A and the node C, so that extensive latency is inevitable if the source S sends a message to the destination D.

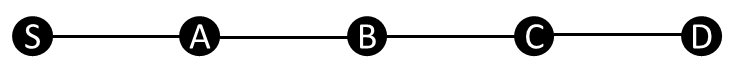


Figure 2.1 Regular Network

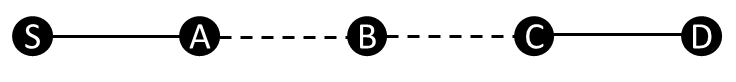


Figure 2.2 DTN

Since the challenged networks have significant differences with the regular networks, researchers introduce a new architecture Delay Tolerant Network (DTN) to deal with the unique features in the challenged works. Terrestrial Mobile Networks, Exotic Media Networks, Military Ad-Hoc Networks and Sensor/Actuator Networks are typical DTNs.

### Ad hoc Network

“Ad-hoc networks are infrastructure-less and cooperation-based networks which means that the network topologies must be decided by the network sensors themselves” [6]. In other words, the major feature of ad-hoc networks is that they have no infrastructure.

With the help from infrastructures, users in regular networks can connect to any other users. As shown in Figure 2.3, users can communicate with the ones who are not in their communication radius through the base stations (i.e. the lighting tower). As a result, users achieve a low drop probability because of their end-to-end path between each pair of users.

If we remove the base stations, users are disconnected from the ones who are not covered by their communication radius disconnect, as shown in Figure 2.4. The three left users can communicate with each other, because they are all covered by the other two user’s communication radii. The user A and the user C are not connected directly, but they need the user B’s help. That increase the instability of their connection, because the user B might leave or sleep so that the user A and the user C are disconnected. It is even worse that the left part and right part cannot communicate with each other, since there is no path between them until another user (i.e., the white user) appears in the middle to connect them. Therefore, it is harder for users to deliver messages in the ad-hoc network than that in the regular network.

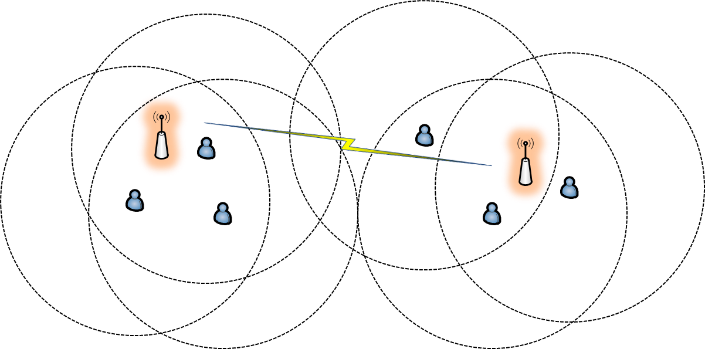


Figure 2.3 Regular Wireless Network



Figure 2.4 Ad-hoc Network

### Mobile Ad-hoc Network

The Mobile Ad-hoc Network (MANET), also called the Wireless Ad-hoc Network (WANET), is a kind of ad-hoc network, where each node forwards data for other nodes. In MANET, users are moving independently, which makes the topography changes frequently. To deliver message in MANET, uses forward messages when they encounter others, or they just store messages, as shown in Figure 2.5. The user A has a message to the user C, so he sends the message to the user B when they encounter each other. The user B stores the message and moves, until he meets the user C. When the user B and the user C make contact, the user B sends the message to the user C, so that the message from the user A is delivered. It is obvious that whether the message can be delivered depends on the movement of users. If the user B never meet the user A or the user C, the message must be dropped after the timeout.

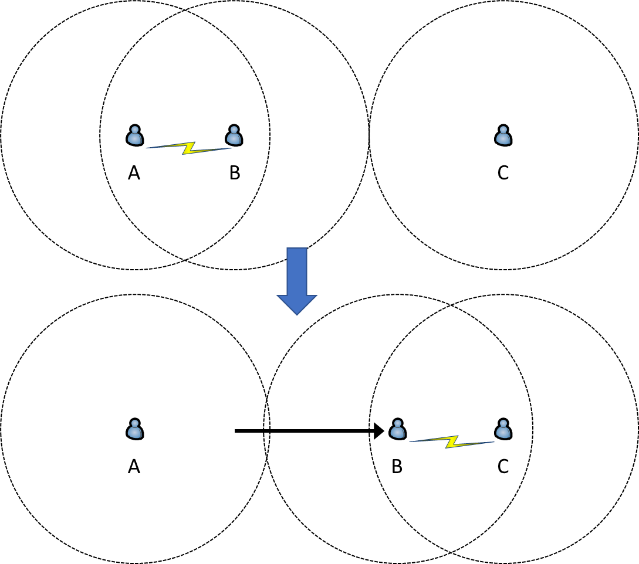


Figure 2.5 MANET

## DTN Routing Protocols

### Replica Based Protocols

The simplest strategy for DTN is that the source holds the message until he meets the destination, which is called the direct contact scheme. So, a direct path between the source and the destination is necessary for a success delivery. It is possible that the source never comes in contact with the destination so that message is not delivered. Researchers make the effort to improve the delivery performance in the DTN.

The replica based protocols works by making several replicas of the message, so that users can retransmits them upon connection establishment. Comparing to the direct contact scheme, the replica based protocols make it more possible for messages to be delivered. But they require more resources than the direct contact scheme, because they need more memory to store the replicas. Therefore, making reasonable decision for the message replication and dropping is the key to these kinds of protocols.

An extension of the direct contact scheme is that the source can make replicas and gives the replicas to the first users, so that there are at most users (i.e., the source and the users who get his replicas) participate in forwarding the message instead of the source himself only. That strategy is called the Two Hop Relay [6], because a delivered message has two hops at most. Although relays can help the source forward the message, their behaviour looks like the source in the direct contact scheme because they cannot deliver the message unless they meet the destination. In the Relay Cast scheme [7], researchers let the relays send the message to remote users, so that the replicas have more chances to reach farther area than the Two Hop Relay.

Researchers also attempt to make the relays distribute the replicas, then the kernel is who is permitted to make replicas. In the Tree based Flooding [6], only the relays in first certain hops can make replicas. In the Epidemic Routing [8], all the relays can make replicas, however, it requires large amounts of buffer space and power.

Besides, authors in [9] and [10] improve the DTN routing strategy by carefully choosing users who can receive replicas or drop replicas. In the Probabilistic Routing [9], only the users whose probability of meeting some other nodes is above a given threshold can get replicas. In the Reconfigurable Ubiquitous Networked Embedded System (RUNES) [10] strategy, users who is close to the source or the destination rarely drop the messages.

The Spray and Wait is a known protocol for DTNs. Although it is not as efficient as protocols in internet, it can work in DTNs. A message is initialized with several copies, a part of which are given to others when users encounter each other. Users keep forwarding copies until they each have only one copy of the message. When a user carrying at least one copy of the message encounters the destination, he gives all his copies to the destination to finish the delivery. The Binary Spray and Wait (BSW) is an optimized version of the Spray and Wait, which is also used in our protocols. The user who creates the message also make several copies of that message. He gives half of the copies the user whom he encounters, so that both him and the other user have a half of the copies. The users who get copies of the message continue give a half of copies they have to others until they have only one copy. The remained one copy is given to the destination only.

### Other Protocols

Except from the replica protocols, there are also other types of protocols used in DTN, but we do not explain them in detail because they are not included in the algorithms of this thesis.

#### Knowledge Based Protocols

Different from the former replica protocols which require no knowledge about the topology of the network, the users in knowledge based protocols try to evaluate their own view of the topology of the network, so that they can make better forwarding decisions. However, the topology changes so frequently that it is hard for users have a accurate topology.

#### Coding Based Protocols

Authors in [11] and [12] make approach by introducing coding techniques into routing protocols. Instead of making a few replicas, coding based protocols encrypts data to make a large number of message blocks. If the destination receives a part of the blocks, he can decrypt the message.

## Location-Based Services

The Location-Based Services (LBS) use users’ location information to provide services, which boost when mobile networks were widely deployed. For example, your smartphone can detect your coordinate and uses it in LBS applications. The point of interest and the location advertisement are familiar instances of LBS. The LBS providers collect users’ location information to provide services, which makes LBS providers a significant target of attackers. When attackers access the data bases of the LBS providers, they can learn all sensitive information in the servers. So, the LBS providers should be viewed as semi-trusted, which might expose information in front of vicious attackers.

Users face risks of information breach when they access a semi-trusted LBS provider, because anyone who has access to data in LBSs is able to steal and misuse LBS users’ location-privacy. Considering that LBSs rely on location-aware computing, it is unavoidable to leak users' location from LBSs. Therefore, balancing “these two competing aims of location privacy and location awareness” [11] is always a challenge.

## Location-privacy protocols

The basic idea of most location-privacy protocols is hiding the original requester behind a set of users, so that attackers can hardly infer the identity of the original requester. In other words, when a user sends a query, his identity cannot be inferred based on the information in the query easily. For example, the original requester can use a pseudonym instead of his own name or use an obfuscated location, so that attackers cannot tell the different between his queries and a set of other users’ queries.

The location-privacy protocols are considered as centralized and distributed ones based on whether they require any infrastructure.

### Centralized Protocols

Centralized Protocols always need some infrastructure to serve for the obfuscation process. The infrastructure could be any equipment which act as anonymization servers. When users want to send a query to the LBS server, he sends the query to the anonymization server. The anonymization server changes the identity and the location in the query, which is called anonymization or obfuscation. Then the anonymization server forwards the modified query to the LBS server, so that the attacker who attacks the LBS server can hardly learn the identity of the original requester. The LBS server can send the reply message to the anonymization server which is responsible and able to forward the reply message back to the original requester. In this way, the original requester gets the information he needs, while avoiding from being located by attackers.

The advantage of the centralized protocols is that the anonymization server has more resources and information than users in the network, which enable the server to achieve better anonymization performance and provide sufficient protection for users. For example, if the anonymization server knows locations of all users in the network, it can modify the location and identity in the queries to the most suitable, with which the LBS server can provide acceptable results while the original requester is not exposed.

There are three major defects of this kind of protocol. 1. It is hard to deploy infrastructure in some area; 2. Since the anonymization server knows users’ location, it is also a risk for users.

#### Single Server

Authors in [14] uses a central anonymity server through which the mobile users can send queries to external services. To make the communication between users and the central anonymity server trustful, users set up an encrypted connection with the anonymity server at the beginning. Users sends their encrypted queries to the anonymity server, so that the anonymity server is the only one who can learn the information in the query. The anonymity server decrypts the queries and uses a cloaking algorithm to perturb the position information in the queries. Then the anonymity server sends the modified queries to external services (e.g. LBS). This is a typical centralized protocol, which can reduce the reidentification risk for users. Its defect is that a continuous connection to the server is necessary for each user, and that condition is hard to achieve in a sparse DTN. We cannot image that a MANET user can have a stable connection with a central anonymity server, neither.

In [22], researchers employ a matchmaker which is used to match users and advertisements, then users can achieve anonymization of their identities and locations from the matchmaker. The system architecture in [22] is similar with the former [14]. Authors in [22] just focuses on the advertisement service instead of the “external services” in [14]. The role of matchmaker is matching users and advertisements. Although the functions of the matchmaker in [22] and the central anonymity server in [14] are different, the intermediate trusted three-party servers (i.e., the matchmaker and the central anonymity server) separate users and the application server (e.g., advertisement server, LBS server and so on). The architecture in [22] does not require an encryption process, so that attackers can learn the information in the queries. Although a plenty of queries arrive the matchmaker in a short time from different users also help the matchmaker mix the queries so that attackers can hardly trace the query, the burden of the users who are near the matchmaker are heavy.

In the work in [23], researches use “a trusted, third-party location anonymization engine (LAE) that acts as a middle layer between mobile users and LBS providers” [23], so that exact locations and requests from clients are replaced by a location anonymization engine before they arrive at LBS providers. Therefore, we learn that the structure of this kind of algorithms is almost like Figure 2.6. A trusted server is placed between users and the application server, which hides users’ information and offers sufficient information for the application servers to provide acceptable services.



Figure 2.6 Single Anonymization Server

#### Multiple Servers

Some early solutions, like [12] and [6], generate a specific area based on *k*-anonymity [13] for each user who needs to send queries. For example, [6] uses a central anonymity server as a mix router. As a result, it is necessary for each node to have a continuous connection to the server. That is hard to achieve in a sparse DTN. With a similar problem as in [6], [14] employs a matchmaker which is used to match users and advertisements, then users can achieve anonymization of their identities and locations from the matchmaker. However, the matchmaker is a high-risk in the network, because it collects so much private information. In the work in [15], exact locations and requests from clients are replaced by a location anonymization engine before they arrive at LBS providers. Since the anonymization engine learns all exact locations and requests, it becomes a better target for malicious attack.

### Distributed Protocols

[12] gives a rectangle as an anonymized area, in which all nodes form a group to hide the original requester. But, it requires at least *k* connected agents to complete its obfuscation process.

However, the system could also be a distributed one, like protocols [7], [8] and [9]. The obfuscation processes are performed by each separated node independently without any help from a third-party device. The utilization of label as in [18] makes it easier for nodes to mix themselves into a group, which is a significant difference from these protocols and the previous ones. Algorithms above use groups instead of an area to protect users. [8] imports 2 concepts: the obfuscation phase and the free phase. In the obfuscation phase, queries must be transmitted between friends for *k* times. When there are only a few friends in the network, it is hard for a node to find an available next hop in obfuscation phase. [8] and [9] attempt to improve [7]’s performance in the obfuscation phase, which might be a safety tradeoff, because some ineligible users in [7] are chosen as friends based on the additional standards imported by [8] and [9].

# Multi-Hop Location-Privacy Protection

## System Model

Our network architecture consists of two main entities: Users and LBS Providers (LBSPs). Due to the introduction of the social network, the users’ social information can be used in obfuscation forwarding process. Based on available information in the social network, the relationship between two users can be considered as friends or strangers. The user who makes a query to an LBSP will be called the original requester while the others are called intermediate users. LBSPs are located in fixed locations and their coordinates are known by all users when users join the network. Attackers are assumed to be able to access LBSPs, and attempt to locate original requesters. We assume that the LBSPs are semi-trusted and the strangers are un-trusted. We also assume that both entities have sufficient resources, like computational capability, storage and battery power.

Since two friends could be a pair of multi-hop neighbors, users can leverage Optimized Link State Routing Protocol [20] to seek friends continuously after entering the network, so that they can recognize each other and make contact in time. When a user carries an obfuscation phase query, he might send the query to a multi-hop friend through several strangers. In this case, a secure communication is necessary for between them, so that they must send the query encrypted to prevent strangers from learning anything about the query. Each user obtains a pair of asymmetric keys (public and secret key) before he joins the network from a certificate authority using well-regarded techniques, like in [19]. Whenever a user detects a new friend, he sends a request to the friend asking for his public key. In this way, a user can get his friends’ public key when they encounter each other. Even though several strangers can be active in the obfuscation phase, the queries can still be securely sent to the user’s friend.

The relationship strength is often “a hidden effect of nodal profile similarities” [21]. Let *SVi,j* denote a value of relationship strength which user *i* determines whether user *j* is an acceptable friend based on the relationship strength. For every pair of users (*i* and *j*), we assume that there is an . If is bigger than a specific friend threshold *Tmin*, set by the original requester, user *j* is considered as a friend of user *i*; otherwise, it will be treated as a stranger. The notations used in this paper and their meanings are shown in Table 3.1.

## Details of MHLPP

MHLPP aims to protect the original requester’s (*N*0’s) location-privacy using an obfuscation path. In other words, a query *q* which needs to be obfuscated must go through a series of friends after it leaves *N*0. The whole process includes two parts: the obfuscation phase and the free phase. In the former phase, *q* is only transmitted among friends, until it is sent to an area called “*obfuscation area*”. At the end of that phase the last friend *Nf* replaces all *N*0’s information by its own and forwards *q* with an arbitrary DTN forwarding protocol, like the one suggested in [22]. In this case, what attackers can learn from the database in LBSP is *Nf*’s information, so they can hardly infer the original requester’s identity and location based on that information. The free phase starts when the query *q* is forwarded by *Nf* and ends when it reaches the LBSP.

Table 3.1MHLPP Symbols

|  |  |
| --- | --- |
| Parameter | Meanings |
| *N0* | the original requester |
| *Ni* | if *i* > 0, it denotes the friend chosen by *Ni*-1.  If *i* = 0, it is *N0*. |
| *Nf* | the last friend who handles the obfuscation query |
| *Nd* | the destination or the LBSP |
| *Ki* | the public key of *Ni* |
| *Si* | the secret key of *Ni* |
| *q* | a query of *N0* |
| *rq* | the requirement for the query *q* |
| *msg* | a message contains *q* and *rq* |
| *Emsgi* | the encrypted *msg* using *Ki* |
| *Sid* | the original requester’s identity |
| *Did* | the destination’s identity |
| *Ls* | the location of *N*0 when it sends the query to *N*1 |
| *Rp* | the inner radius of the obfuscation area |
| *Rs* | the external radius of the obfuscation area |
| *Tmin* | the social value bound for friends |
| *Cmax* | the extra path limit in each obfuscation forward |

Because *Nf* is the only identity the LBSP knows, the LBSP has no choice other than replying to the last friend *Nf* when it receives the obfuscated query. The reply can be delivered with an arbitrary DTN routing protocol as the free phase query does. The friend *Nf* should remember who is the real destination (*N0*) of this reply, then he transmits it to *N*0. In this way, *N*0 is able to send a query *q* to an LBSP while not exposing his own information.

The obfuscation phase of a query *q* starts when the query leaves the original requester *N*0. When a user is holding an obfuscation phase query, he starts sensing connected friends continuously, which enables it to communicate with one-hop or multi-hop neighbor friends. Even though users in the mobile network use OLSR protocol [20] to detect others automatically, they do not communicate with their friend unless they have a requirement to send an obfuscation query. Also, they do not ask their friends for public keys. Therefore, carrying an obfuscation phase query requires a user to execute MHLPP algorithm.

When *N*0 finds the first available friend *N*1, he asks *N*1 for a public key *K*1, which will enable him to encrypt his query using *K*1. That prevent others, e.g. strangers, from learning information in the message *msg* that that contains both *q* and *rq*. *rq* is *N*0’s requirement for *q*, which is always sent with the query *q* and remains constant until the end of the obfuscation phase (we discuss this in part C). Friends who get the query can infer *q*’s obfuscation area based on parameters *Rp*, *Rs* and *Ls* in *rq*, which is a ring with inner radius *Rp*, external radius *Rs* and center *Ls*. Before *N*0 sends the query *q* to his friend, he initializes parameter *Ls* to his current location and encrypts *msg* using *K*1 to get *Emsg*1 which is what *N*0 sends to *N*1.

The destination of *Emsg*1 is *N*1, which is a plaintext in *Emsg*1, so that other intermediate users (strangers) can help *N*0 forward *Emsg*1 to *N*1. In this step, strangers transmit *Emsg*1 using the OLSR protocol if *N*1 is a multi-hop neighbor of *N*0. Strangers learn nothing other than the identity of *N*1, because both *q* and *rq* are encrypted. They cannot help attackers locate *N*0 because they do not know *Sid* (the identity of *N*0) and *Did* (the LBSP) included in *rq*.

When *N*1 receives *Emsg*1, he decrypts it with its secret key *S*1 to get *q* and *rq*. If *q* is already in the obfuscation area defined in *rq* (i.e., it is already in the ring), the query *q* finishes its the obfuscation phase. Then *N*1 replaces all information of *N*0 with his own. For example, the *Sid* is replaced with *N*1. If a location is necessary for the LBS, *N*1 uses his own current location and records this change in his memory before initiating the free phase. A free phase query can be then be forwarded to the destination (i.e. LBSP).

If *q* is still in the obfuscation phase (not in the ring), *N*1 performs similar actions just as *N*0 expect modifying *rq*. Another difference is that instead of finding friend randomly, *N*1 would seek for a friend who is nearer in the obfuscation area, and so will the following friend.

The detailed algorithm is explained in Algorithm 3‑1 where *Nx* is a neighbor of *Ni* who is carrying the query *q*. Function “*DealWithQuery*” is responsible for dealing with a query. For *N*0, he generates the query *q* and its requirement *rq*. If *Ni* receives *Emsgi*, he decrypts it with his own secret key *Si*. If *q* finished its obfuscation phase at *Ni*, *q* will be required to be forwarded in free phase immediately. Otherwise, *q* needs to be processed in the obfuscation process. Both *q* and *rq* are stored in *Ni* until they are sent to the next friend. *Ni* starts detecting friends continuously if and only if *Ni* carries one or more obfuscation phase queries.

When *Ni* detects a new neighbor *Nx* (one-hop or multi-hop neighbor), he follows steps in “*WhenEncounterUser*”. For an expired query, *Ni* simply drops it. If the query *q* is already inside its obfuscation area, *Ni* switches it to the free phase. If *q* stays in the obfuscation phase and *Nx* is an available friend, *Ni* encrypts both *q* and *rq* using *Nx*’s public key *Kx* to get an encrypted message *Emsgx*. Then *Ni* forwards *Emsgx* to *Nx* and stops sensing friends after *Emsgx* departs from it.

When we mention that *Ni* switches a query *q* to the free phase, *Ni* actually replaces *N*0’s information with its own one in *q* to get *q\** and records this replacement in its storage, then *Ni* uses the Spray and Wait protocol [22] to forward *q\** in plaintext. That allows *Ni* to hide *N*0’s identity and forward a reply from LBSP to *N*0.

## Requirement parameters

Algorithm 3‑1 The obfuscation phase in MHLPP

**function** DealWithQuery ()

**if** the current user is *N*0

generate (*q*, *rq*) by himself

**else if** the current user is *Ni*, *i* > 0



**end if**

**if** *q* can switch to the free phase based on *rq* then

SwitchFree (*q*)

**else**

*msg* ← (*q*, *rq*)

Store *msg*

Start sensing friends

**end if**

**end function**

**function** WhenEncounterUser (*Nx*)

*q* ← get query from *msg*, *rq* ← get requirement from *msg*

**if** *q* timesout then

remove *msg*

return

**end if**

**if** *q* is eligible to switch into the free phase then

SwitchFree (*q*)

return

**end if**

**if** *SVi, x* is bigger than *Tmin* in *rq* then

**if** it is the first hop of *q* then

assign the current location to *Ls* of *rq*

**end if**



forward *Emsgx* to *Nx*

remove *msg* from memory

stop sensing friends

**end if**

**end function**

**function** SwitchFree (*q*)

*q\**← replace *q*’s requester-information (*N*0) by *Ni*

record (*q*, *q\**)

forward *q\** with DTN protocols.

**end function**

In the obfuscation phase, *rq* is always in *msg* so that friends who get *msg* can make decisions (e.g. selections of friends) based on it.

All parameters (i.e., *Sid*, *Did*, *Rp*, *Rs*, *Ls* ,*Tmin* and *Cmax*) in *rq* are given by *N*0 before *Emsg*1 leaves *N*0. Parameters *Sid* and *Did* record the identities of *N*0 and the destination (LBSP) *Nd*, based on which last friend *Nf* is able to send the query freely to *Did* (*Nd*) and forward the reply to *Sid* (*N*0).

The obfuscation area is a ring with an inner radius *Rp* and an external radius *Rs*. As shown in Figure 3.1-a, the obfuscation area is actually the grey area “*a*”. Obfuscation area must guarantee both the original requester’s location-privacy and location awareness. In other words, the value of *Rp* should be big enough, so that there are sufficient users in the inner circle (with a radius *Rp*). At the same time, *Rs* should be small enough so that the LBSP can provide a service, acceptable to *N*0.



Figure 3.1 The selection of the next friend

A user *Nx* can be chosen by *Ni* as a friend for *q* if and only if *SVi,x* is bigger than the threshold *Tmin*. The original requester *N*0 can set various values for his queries based on their importance. If *Tmin* is large, there would be fewer friends for any users in the network which reduces the query success rate to a certain extent, as a result. We assume that the original requester can balance the level of privacy and the success ratio.

Most DTN routing protocols aim to deliver queries through the shortest path, while MHLPP pays more attention to security in its obfuscation phase. Consequently, the obfuscation process in MHLPP results in a longer path from the original requester *N*0 to the destination *Nd*. To limit the length of the path, we introduce parameter *Cmax* which is the maximum extra path (e.g., the difference of the length and the distance between the *Ni* and LBSP) we can tolerate. For any friend *Ni* who gets a *Cmax* from *rq*, if he selects *Nx* as the next friend, the extra path should not be longer than *Cmax*. Let’s denote the optimal path from user *m* to *n* by *Dis*(*m*,*n*), and the extra path from *Ni* to *Nd* through *Nx* by *Ci,x,d*. Then *Ci,x,d*  can be defined as follow.

 (1)

*Ci,x,d* must be a value smaller than *Cmax*. If *Dis*(*m*,*n*) is the straight-line distance between point *m* and *n*, then next friend *Nx* should be in an ellipse *EC* with focus points *Ni* and *Nd*. Let’s denote the coordinate of *Ni* by  and the coordinate of *Nd* by . Then, the equation of the ellipse *EC* is

 (2)

As shown in Figure 3.1-a, *Ls* is the center of the ring while *Rp* and *Rs* are the inner and external radii, respectively. The query *q* switches to the free phase when it enters the ring area.

As shown in Figure 3.1-b, *Ni* who is carrying obfuscation queries should choose his next friend *Nx* in the ellipse, which avoids the query *q* going through an unacceptably long path.

In conclusion, a query *q* starts at the center and moves inside the ring, until it reaches the obfuscation area. The point *Ni* in Figure 3.1-b should be inside the ring, and the point *Nd* might be anywhere. As a result, a user *Ni* who is carrying an obfuscation query detects a friend continuously who has a larger distance from *LS* and inside an ellipse. If there is a friend like that, *Ni* sends the query to that friend *Nx*.

## Privacy Analysis

We assume that attackers can achieve all information in LBSPs know. Obviously they can know the identity of the last friend *Nf* who replaces *N*0’s information with his own. It is possible for the attackers to locate *Nf* with little cost. For example, if *N*0 stops moving after sending the query *q*, it is reasonable for *Nf* to believe that *N*0 is in a ring centered at the location of itself with radii *Rp* and *Rs*. In other words, the distance between *Nf* and *N*0 should be in a range between *Rp* and *Rs*. If attackers find all users who satisfy this condition, the original requester might be among these users with high probability. Then, a success ratio to locate *N*0 can be measured by a conditional probability

 (3)

where  is the probability that *N*0 is in the ring (i.e., the distance between *N*0 and *Nf* is larger than *rp* and smaller than *rs*). Here, is the number of users who are in the ring. Attackers locate *N*0 successfully if and only if *N*0 is in the ring, at the same time, attackers pick the correct one from all users at that area.

In the worst case, attackers know exact values *Rp* and *Rs*. Then, the Eqn. (3) becomes

 (4)

where is the probability that *N*0 is on the ring (i.e., the distance between *N*0 and *Nf* is larger than *Rp* and smaller than *Rs*). is the number of users who are in the ring. Since parameters (e.g., *Rp* and *Rs*) in *rq* are kept secret among trusted friends in our system model, attackers can hardly get the actual values of those parameters.

## Complexity discussion

In order to guarantee secure communications among friends, encryption is introduced in our protocol. In the obfuscation phase, the query is transmitted along friends, i.e. *N*0, *N*1, *N*2, ..., *Nf*. When the query is sent from *Ni* to *Ni*+1, a pair of encryption and decryption is needed, so the number of such pairs *Ten* should be equal to *f*.

*Ten* grows with both *Tmin* (threshold used to decide friend relationship) and inner radius *Rp*. Essentially, it is the number of friends participating in transmitting a query *q* in its obfuscation phase that influences *Ten*. Given a smaller *Tmin*, a user carrying the obfuscation phase query has more chances to encounter more friends in a certain area. A larger *Rp* also leads to a bigger area inside the ring, so that there are more friends in this area. We evaluate the number of encryptions and decryptions in our simulation. Figure 3.2 shows the average number of encryptions (*Ten*) with different friend thresholds *Tmin* and inner radius *Rp*. We observe that *Ten* increases steadily as we increase *Tmin* and *Rp*.

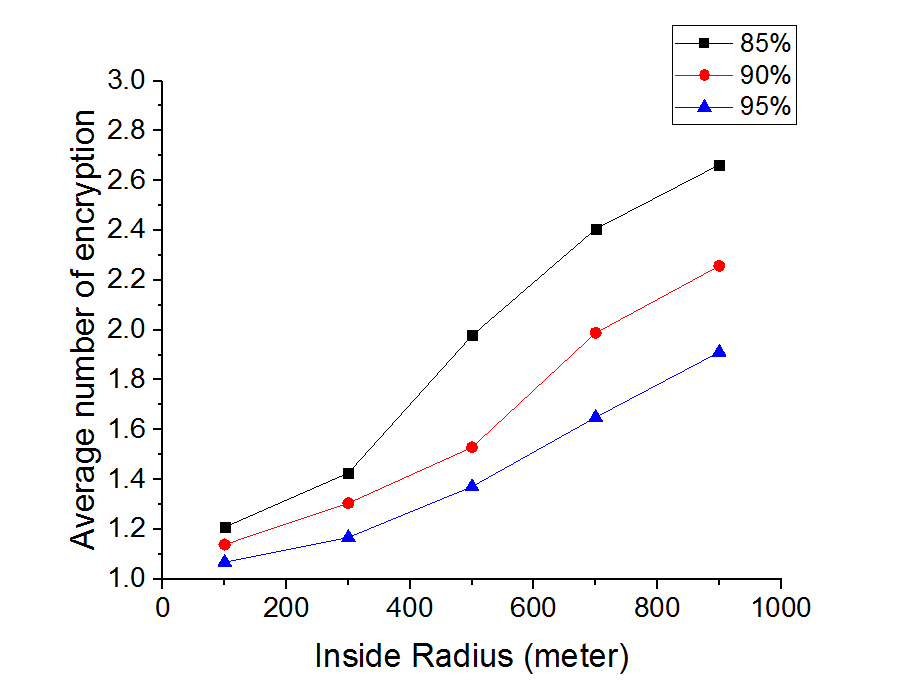


Figure 3.2 The number of encryption with various inner radius

It is evident that encryption and decryption process in MHLPP results in an extra cost in both energy and computational resources. However, the number of encryptions and decryptions is quite low (below 3) which reasonable based on the experiment.

## Performance Analysis

We use the map of Helsinki in our simulator to evaluate MHLPP. It is also compared against the known protocol, Hybrid and Social-aware Location-Privacy in Opportunistic mobile social networks (HSLPO) [5]. The simulation parameters are shown in Table II. All pedestrians and cars are users in MHLPP. These users are moving on the map along streets continuously. There is an LBSP fixed at a random location on the map. For each user, we give him random social values between 0% and 100%, each corresponding to all other users. Each value has the same probability, so we can compute the expected number of friends of an user. For example, if we are given a privacy threshold (*Tmin*) of 85%, then there might be 15% (100%-85%) users who are friends of a certain user.

As shown in Table 3.2, there are 126 users in the map. For each of them, say user *i*, we give him 126 random *SV* values, which denotes the relationship strength between him and other users, so that there are 1262 *SVs* in our simulation. The *SV*s are between 0 and 100. As a result, if the *Tmin* is equal to 85, the average number of friends of each user should be 18.9 (=).

Table 3.2 MHLPP simulation parameters

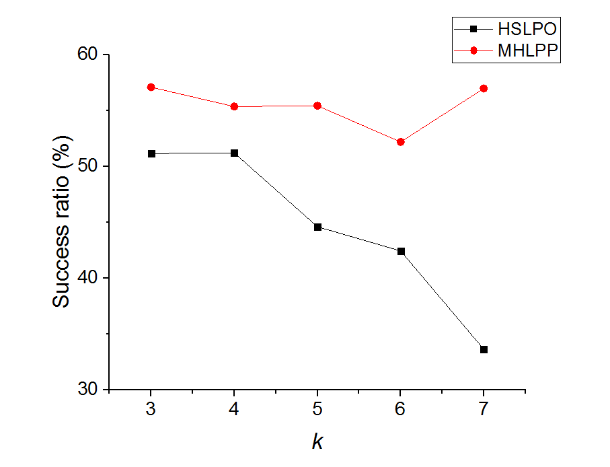
|  |  |
| --- | --- |
| Parameter | Value |
| Simulation Time | 10 minutes |
| Map Size (W x H) | 4500 m x 3400 m |
| Total number of users | 126 |
| Pedestrians/ Cars | 84/42 |
| Communication Area Radius | 10m – 90 m |
| Pedestrian Speed | 1.8-5.4 Km/h |
| Car Speed | 10-50 Km/h |

Users are placed at random locations at the beginning of each experiment. We choose another random point for each user, so that he can move back and forth along streets between that point and the point his starts at. The user speed depends on its type (pedestrians or cars) and set randomly. All queries have a 10-minute timeout. The queries which are expired before they reach the LBSP (the destination) are considered to be failed in our success ratio statistics.

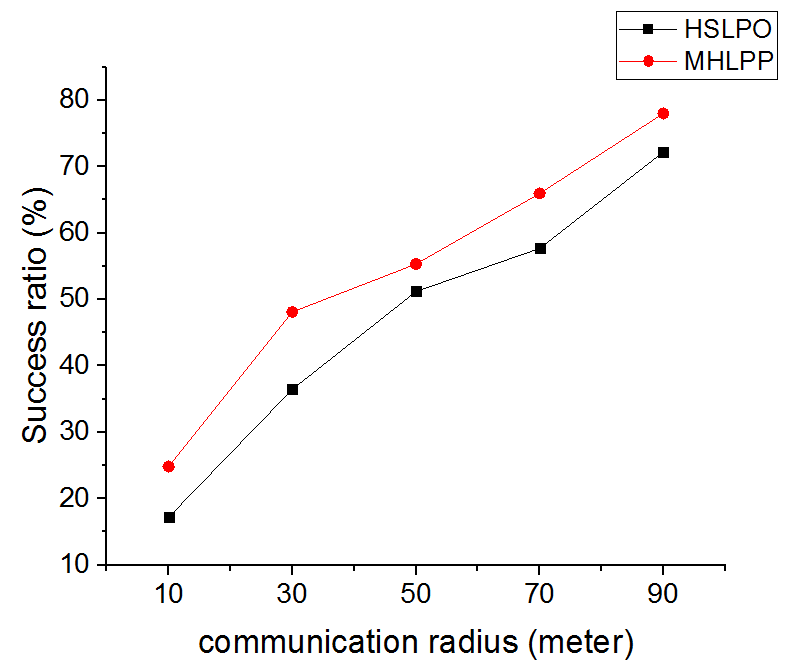
Figs. 3-5 compare performances between HSLPO and MHLPP for different values of *k*, communication radius and privacy threshold (*Tmin* in MHLPP). The *k* is the privacy-level requirement in HSLPO. Both HSLPO and MHLPP have different criteria in which a query can switch to the free phase. To make them comparable, we create a new parameter, called *obfuscation distance*. If a query leaves at location *La* and switches to the free phase at *Nf* whose location is *Lb*, then the obfuscation distance is the straight-line distance between *La* and *Lb*. We test the obfuscation distances of HSLPO with different parameters, and then we set the inner radius of MHLPP to those values. The query success ratio is the ratio of delivered queries to the total number of queries. The number of hops (*h*) is the number of intermediate users between *N*0 and the destination (LBSP). We count the number of users surrounding the last friend in a specific range, which is *k* times the communication radius. We calculate the entropy using the reciprocal of the number of surrounding users.

### Query success ratio

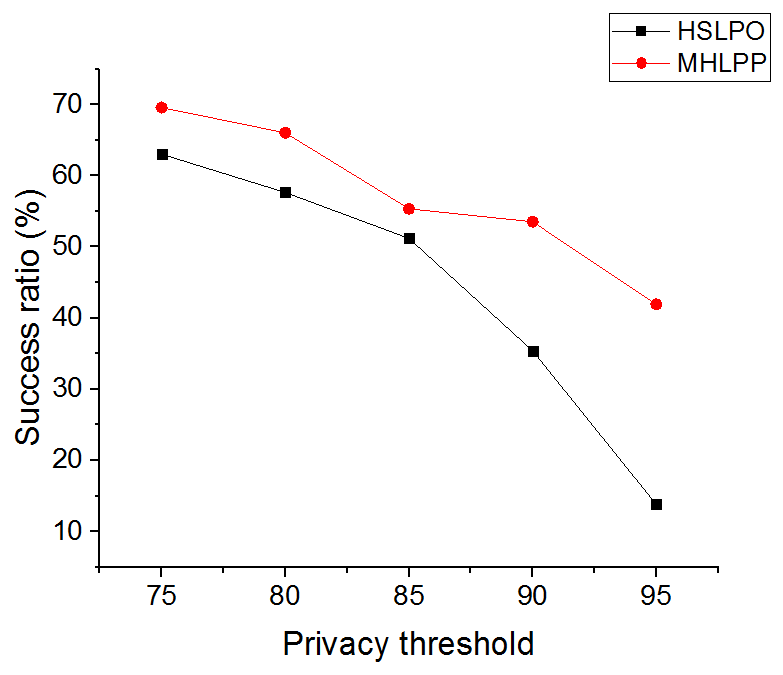
The query success ratio is the percentage of delivered queries among a number of attempts. Based on the timeout value in Table II, a query is delivered successfully, if it arrives at the LBSP (the destination) before the timeout; otherwise it fails. We use the query success ratio to evaluate the delivery performance of MHLPP.



(a) Query success ratio with various K



(b) Query success ratio with various communication radiuses



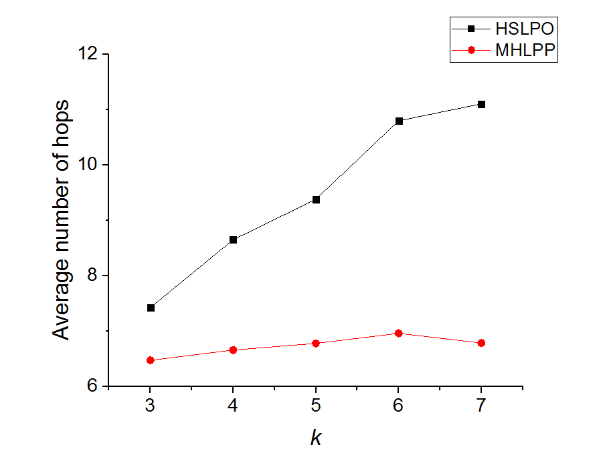
(c) Query success ratio with various privacy thresholds

Figure 3.3 The query success ratio comparison between HSLPO and MHLPP

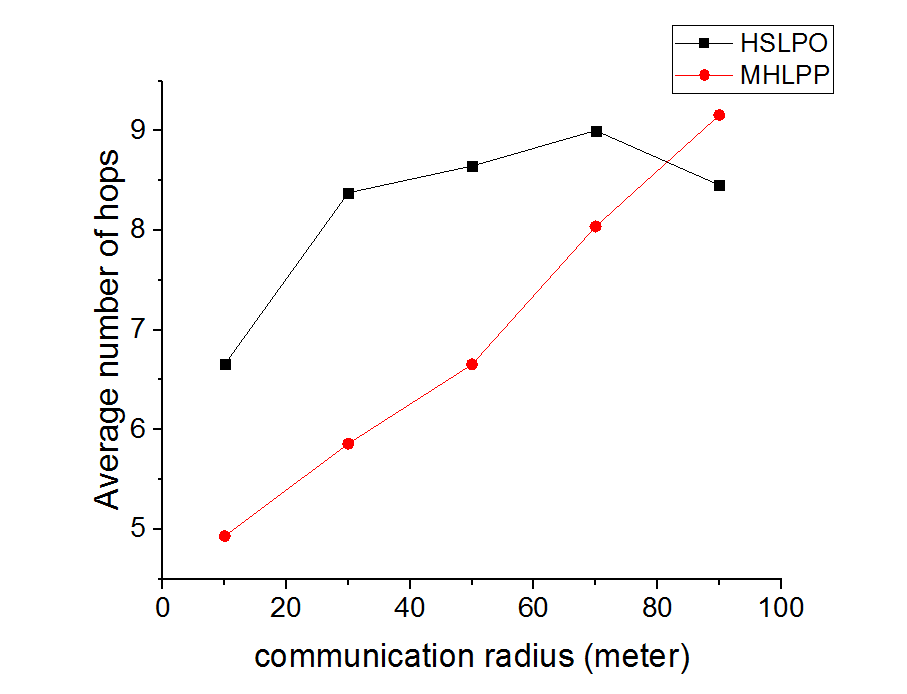
As shown in Figure 3.3-a, the success ratio in MHLPP is always higher than that in HSLPO. As the value of *k* increases, HSLPO success ratio drops sharply while MHLPP remains stable. This is because the larger *k* is, the harder it is for HSLPO to find enough friends in a limited time. The lack of friends has less impact on MHLPP. We observe that the success ratio of MHLPP rises when *k* = 7. That is because it depends on the inner radius which is equal to the obfuscation distance of HSLPO. The obfuscation distance decreases when *k* = 7, because most of the queries which complete their obfuscation phase have a short obfuscation distance. In Figure 3.3-b, both HSLPO and MSLPP values increase and have the same trend when given a larger communication radius. The reason is that the communication radius effects the free phase more than the obfuscation phase for both. As shown in Figure 3.3-c, higher privacy threshold leads to lower success ratio in two algorithms. Its impact on HSLPO is more intense than that on MHLPP, which is the most important characteristic of MHLPP. MHLPP has a better performance than HSLPO especially when there are fewer friends in the network. MHLPP can transmit messages with the help from strangers in its obfuscation phase while HSLPO cannot.

### Number of Hops

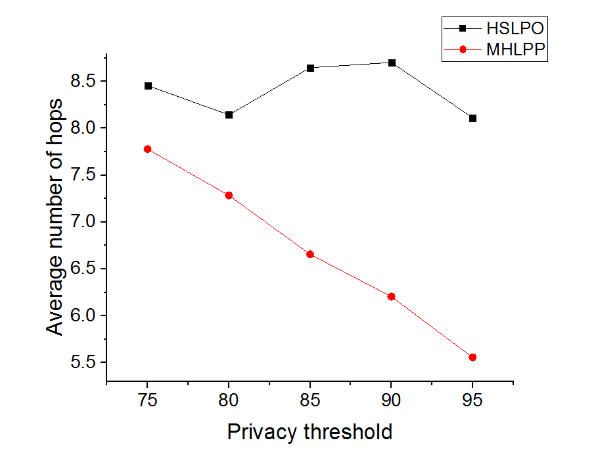
We count the number of hops it takes for queries to be delivered successfully and calculate the average. Every user who takes part in the delivery process is considered in the hop count. We introduce this criterion to measure the routing path length of the algorithm. MHLPP is more sensitive to the probability that a user encounters a friend than HSLPO is. The reason is that MHLPP aims to reach a certain distance rather than taking certain number of hops. In other words, MHLPP continues sending queries to other friends until queries enter their obfuscation area. In this process, MHLPP takes every chance to forward queries. If it is hard for MHLPP to find friends, it can also take the queries to obfuscation areas with fewer friends. Therefore, the probability that users encounter friends has less impact on the performance of MHLPP. The result of this experiment is shown in Figure 3.4.



(a) number of hops with various *k*



(b) number of hops with various communication radii



(c) number of hops with various privacy thresholds

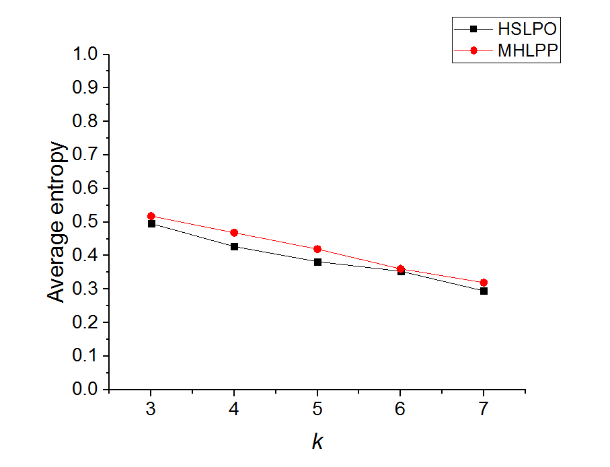
Figure 3.4 The number of hops comparison between HSLPO and MHLPP

In Figure 3.4-a, the number of hops in HSLPO is affected by parameter *k* obviously. Especially, the first *k* hops forwarders must be friends, which makes it hard for HSLPO to have queries to be forwarded successfully. Both protocols have similar number of hops in their free phases, but HSLPO has exactly *k* hops in its obfuscation phase, while MHLPP can have fewer than *k* hops. In figure 4b, the number of hops in MHLPP grows with the communication radius obviously, while it does not change a lot in HSLPO, because only successful delivery queries are counted in the statistics. Given a large communication radius, MHLPP has a much higher success ratio for delivered queries as it can connect more friends. In Figure 3.4-c, the value of MHLPP drops for higher privacy thresholds. The reason is the same as Figure 3.4-b. For HSLPO, no matter how hard it is to find a friend, it attempts to find exactly *k* friends. However, if it is too hard to find a friend, MHLPP’s friend can carry the query while moving and complete the obfuscation process. For higher privacy thresholds (i.e. resulting in fewer friends), MHLPP chooses to carry queries other than finding friends. That results in a drop in the number of hops.

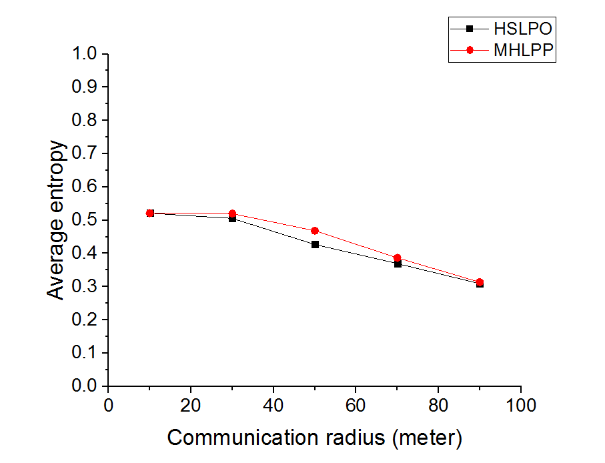
### Security

Since the principles with that two protocols protect original requesters are different, we evaluate the probability that attackers locate the original requester if the distance between him and the last friend is smaller than a some value *r*, which is equal to *k* times the communication radius. Since the last friend reveals himself to the LBSP, attackers might locate him accurately. We assume that attackers know the privacy parameter *k*. In the worst case, the distance between the original requester and the last friend is smaller than *r* when attackers start to locate the original requester. That gives attackers a chance to locate the original requester. We count all users who are inside the radius *r* of the last friend, and the original requester is one of them. For example, if there are *m* users in the area, the probability should be *1/m*. Figure 5 compares the entropy *E* of both HSLPO and MHLPP. We use the following formula for computing entropy:

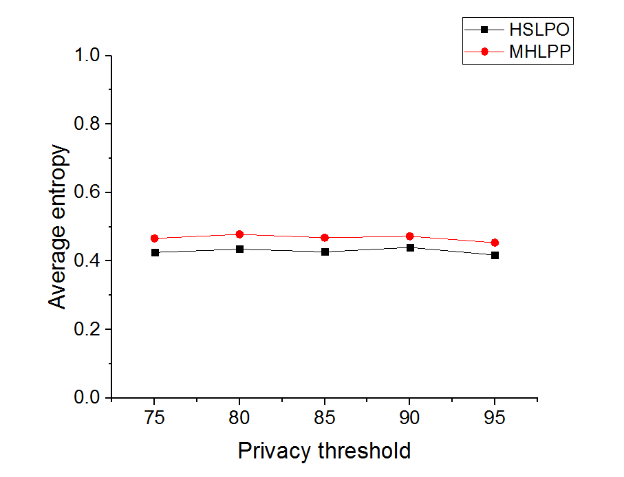
From Figure 3.5-a, we observe that MHLPP has very small (about 0.04) increase in entropy compared to HSLPO. When the original requester is in the circle centered at the last friend, HSLPO is a little more secure than MHLPP but not by very much. That is because HSLPO always switches to the free phase when the last friend encounters the previous friend, so the previous friend must in the circle. MHLPP does not have this condition. Both graphs in Figure 3.5-a and Figure 3.5-b have the same trend. The curve of MHLPP is also a little higher than that of HSLPO, while two curves almost meet when the communication radius is small or large. Given a small radius, the entropy of both protocols are small. When the radius is large, we can ignore the effect of the previous friend mentioned about for Figure 3.5-a as there are so many users in the circle. From Figure 3.5-c, since the circle neither expands or shrinks, we observe that the two curves exhibit similar behavior. The values of HSLPO are always lower than the correlated values of MHLPP. However, as we observed in the experiment, the last friend is always hundreds of meters away from the requester.



1. Locating probability entropy with various K



1. Locating probability entropy with various communication radiuses



(c) Locating probability entropy with various privacy thresholds

Figure 3.5 Locating probability entropy comparison between HSLPO and MHLPSP

# Appointment Card Protocol

## System Model

The network architecture consists of two main entities: Users and Location-Based Service Providers (LBSPs). Users are mobile and communicate with others in a certain range, e.g., the communication range of their portable devices. For a user, other ones in the social network are either strangers or his friends whom he can detect when they are in his communication range. Let denotes the relationship strength between user and , if is larger than a specific friend threshold , user is a friend of user . LBSPs are fixed and not connected in a social network, which provides Location-Based Services (LBSs) for users. We assume that the only information which is necessary for the LBS is a location from the original requester, but the original requester should still give an identity to the LBSP so that the LBSP can reply to that identity.

We consider only an external attacker capable of eavesdropping on limited traffic in the network. We assume that the attacker can access the database of LBSPs, so that he can learn everything recorded in LBSPs, including identities and locations. He launches an inference attack to each user who use the LBS and learns more privacy information based on the information including the location and the context in the queries. Therefore, the key of protecting location-privacy is degrading the relationship between his identity and the location provided by him so that the attacker can hardly infer the identity of the original requester by the known information.

We propose an Appointment Card Protocol (ACP) to protect the identity and location-privacy of the original requester by providing other users’ identity (agents), which can be any user in the network, so that ACP provides large anonymity set. The friends of the original requester separate the agents and the original requester, so that the agents have no knowledge about the original requester.

## Appointment Card Protocol Overview

We propose an Appointment Card Protocol (ACP) to protect original requesters when they are served by LBSPs. A user () generates his own Appointment Cards (ACs) containing his own identity called and a unique number called (Creator Appointment number). The appointment cards are exchanged when two users encounter each other. When the original requester sends a query, he chooses an AC and sends the query using the identity which is in the AC. A LBSP replies to when it receives the query. is the one who generates the AC and the first agent of the AC. He re-transmits it to the next agent () whom he gives the AC to, and so do the following agents until the reply reaches the last agent. The last agent is responsible to forward it to the original requester.



Figure 4.1 Whole Process Example

The Figure 4.1 is an example of the whole protocol, the explanation of the symbols in the figure is shown in Table 4.1. These symbols and the figure are used in the whole chapter to help us describe the protocol. To make it clean, we may omit some superscript and subscript from those symbols in the following sections when there is no ambiguity. The whole process can be considered as the following parts: 1). exchanging cards among all users who are called agents (i.e., 1 and 2); 2). exchanging cards among friends (i.e., 3); 3). sending query using information of appointment cards (i.e., 4); 4). forwarding the reply among agents (i.e., 5, 6 and 7); 5). relay to the original requester (i.e., 8).

Table 4.1 Symbols

|  |  |  |
| --- | --- | --- |
| *entry* | *explanation* | |
|  | A user whose identity is . | |
|  | The appointment card generated by a user . |
|  | is being forwarded by an agent . |
|  | A query whose original requester is and using |
|  | The reply of a query which uses . |
|  | is being forwarded by an agent . |
|  | The () agent of . |
|  | The parameter in . see Table 4.2 |
|  | The parameter in , which is given by an agent . see Table 4.2 |
|  | Both the and the in an AC are called the **A**ppointment **N**umber. |

## Appointment Card

Since the original requester cannot use his own identity to communicate with the LBS provider (LBSP), he must use others’ identity () to send queries, so that the LBSP can reply to the original requester through . Appointment cards make it possible that agents forward the reply to the original requester one by one. In other words, the appointment card indicates a path, through which can get its reply.

Table 4.2 Appointment Card

|  |  |  |
| --- | --- | --- |
| *entry* | *explanation* | |
|  | The identity of the creator who generates the AC. |
|  | A unique number that distinguishes an AC from other ones generated by the same creator. |
|  | The identity of an agent who gives the AC to the recent holder. (The previous hop of the AC) |
|  | A unique number that distinguishes an AC from other ones transmitted by the same agent. |
| time-out | The time when the AC expires. |
| EQ | A queue (Exchange Queue) which records users who exchange the AC in order. Its length is *EQL*. |

In Figure 4.1, the user *a*, *b* and *c* are the agents of the appointment card (i.e., , and ). These agents are strangers, so that attackers can hardly infer from the identity . At the same time, is in the original requester ’s social tie (i.e., is ’s friend or his friends’ friend, ...), and he is the only one who knows how to reach . Therefore, it is hard for attackers to infer from .

Notice that receives ’s appointment card (i.e. ) from a stranger who knows the information of and the identity of the next agent , so that it is unsafe for to use that appointment card. In other words, the appointment card cannot be used until exchanges it to another user (e.g., the user ) who trust . The appointment card is called the *ready appointment card* (*ready AC*) after it leaves the last agent (i.e., the user ), or it is called the *distributing appointment card* (*distributing AC*). It is obvious that *distributing AC*s are transmitted among agents who can be strangers, while a user can only get *ready AC*s from a friend of him.

To make users carry a similar number of ready ACs, ready ACs are also exchanged between friends. As a result, the last agent does not sure whether the user whom gets the ready AC from him is the original requester. We introduce a pseudonym mechanism, which enables the last agent to forward the reply to an unknown original requester.

All users in the network are responsible to generate their ACs respectively and they are called the creators of their own ACs. The creator writes down his own identity (), a unique Creator Appointment Number () on his AC. When users exchange ACs, they modify (the Agency ID) and (the Agency appointment number) in the AC to enable the next agent to identify whom is the previous one. Entries of AC are shown in Table 4.2.

## AC life cycle

The life cycle of an AC starts when it is generated by its creator. The first (see Table 4.3) agents add their identities into its EQ (see Table 4.2) before exchanges it, which increase the length (*EQL*) of the EQ. When the AC’s *EQL* reaches , it is eligible to be used in a query and called a *ready* AC. When an AC is used in a query, it is marked as an *used AC* by the original requester who uses the AC. No matter what state (*distributing*, *ready* or *used*) an AC is in, it can expire, as shown in the Figure 4.2. An AC starts at the *distributing* state. If the length of its *EQ* reaches *k*, it is switched to the *ready* state. It can timeout in all states. If an AC is used for once, it is switched to the *used* state. A *used* AC can also be used in other queries but cannot be given to any. All the system parameters are shown in Table 4.3. The purpose of introducing these system parameters is the following sections. After that, we introduce the whole process of the ACP.



Figure 4.2 AC's life cycle

Table 4.3 Important system parameters

|  |  |  |
| --- | --- | --- |
| Parameters | explanation | |
|  | The obfuscation distance |
|  | The friend obfuscation distance |
|  | The distributing segment |
|  | The generating period of appointment cards |
|  | The timeout for appointment cards |
|  | Avoiding time |

## System parameters

### Obfuscation distance

The obfuscation distance is the number of exchange before an AC is switched to the *ready* state. In other words, an AC must be exchanged for times before it becomes a *ready* AC.

As shown in Figure 4.3, an AC is exchanged along , , ..., . Since those agents are strangers, the only relationship between two adjacent agents is that they encounter each other somewhere. The relationship between and becomes weaker when we increase . In other words, attackers can hardly infer the identity of when he only knows the identity of , and his difficulty increases with the parameter . As a result, it is hard for attackers to infer the original requester, even though is in the social tie of the original requester. Since the reply message must go along the serial of the agents, a long obfuscation distance also lengthens the path of the reply message. Therefore, a large makes the original requester safer, while making it harder for the reply longer to be delivered.



Figure 4.3 obfuscation distance

### Friends obfuscation distance

Since is a stranger for , it is possible that is exactly the attacker. We assume that the attacker knows that is in the social tie of the original requester. The attacker can assume that is a close friend of the original requester, so the identity of gives the attacker a good tip to infer the original requester. A solution to prevent the agent from learning the original requester easily is that the last agents are friends, as shown in Figure 4.4. The is called friends obfuscation distance, and the last agents are called trusted agents.



Figure 4.4 Friends-obfuscation distance

It is true that is a friend of , , but two nonadjacent trusted agents (e.g., and ) might have a weak relationship. Since there are at least trusted agents between the stranger and the original requester, can hardly infer the original requester based on information he learns. When , all ACs must be exchanged between friends only. When , the last agent is the only one who is in the social tie of the original requester, which is the case in our example.

However, friends encounter each other rarely so that a big *m* increases the difficulty of distributing ACs. In this paper, we assign to 1.

### Distributing segment

We also use the example in Figure 4.1. If generates 3 ACs (i.e. , and ). It exchanges these ACs to , then exchanges them to and so on. At last, all of them reach the original requester . In this case, has several ACs whose routes (the list of agents) are the same, so that has no choice. In other words, since using these ACs results in the same reply routes, is restricted to choose an optimal AC based on specific situation.

If agents exchange ACs to different users, the above problem will not happen. We use a system parameter distributing segment to avoid giving all ACs which are received together to the same user. Each user has *Distributing AC List*s (*DL*s). It put received *distributing* ACs in one of those *DL*s randomly. If exchanges ACs with another user, selects one of his *DL*s, and only ACs in that *DL* will be exchanged to the other user.

For example, we assign so that each user has two *DL*s (i.e., *DL*1 and *DL*2). When receives , and from , may put and in its *DL*1, while in its *DL*2. If encounters , can give either and or only to . In this way, we separate ACs generated by the same creator.

### Generating Period

Since ACs could expire, users must generate new ACs continuously. We use GP to denote the speed of generating new ACs per user, that is, each user generates a new AC every GP second.

### AC timeout

Let denote the timeout of ACs. An AC expires after it is generated for seconds. When an AC expires, all agents delete its information from their memory.

### Avoiding time

If a user gets an AC at time , he cannot get that AC again before , where is parameter avoiding time. It is described in 4.6.2.

## Protocol Details

### Generating appointment cards

Maintaining a certain number of ACs in the network is a prerequisite for users to sending queries. Considering that ACs can expire, users must generate ACs continuously. The user who generates an AC is called the creator of the AC, at the same time, he is also the first agent of the AC. ACs are generated based on 2 principles, which are fairness and continuity.

ACs impose burdens on agents. In fact, agents are unlikely to benefit from relaying messages to the original requester. Because they need to allocate memories to save ACs’ information and cost more energy for forwarding replies. To avoid a part of users who generates more ACs than others going into overload, fairness is an essential part of the generating strategy. In other words, users should generate a similar amount of ACs.

We assume that the AC timeout (i.e. ) is 30 minutes, and every user generates 100 ACs at the beginning (e.g., the 0th minute) and generates no AC until the 30th minute. Since all agents remove the information of expired ACs from their memory, a reply can hardly be delivered when the AC it uses is timeout. As a result, it is hard for a user to select an appropriate AC for his query at the 29th minute when all ACs only have 1 () more minute, because his reply may lose even though his query is delivered successfully. Therefore, the generating strategy should be a steady and sustainable, instead of an intensive one.

In our protocol, each user generates a new AC every GP second. Since an AC has a AT-seconds timeout, there are about appointment cards hold by each user in the network. In other words, each user maintains about ACs which is generated by himself, and we achieve the fairness part. Since users generate ACs continuously, ACs have various timeout so that it is likely for a user to pick an AC who does not expire in a long time (at least longer than his queries and replies).

### Exchange distributing appointment cards

Users exchange their *distributing* ACs as frequently as possible, so that a *distributing* AC can be switched to a *ready* one quickly. Still, there are some other conditions which should be obeyed when exchanging a *distributing* AC.

We assume that two users, e.g.  and , walk together for a long time. If generates an AC (e.g., ) at and exchanges it to , then is given back to A and so on. The distributing phase of the AC only costs a few seconds, while only and are agents. That loses the meaning of exchanging ACs, because it is easy for an attacker to infer the last agent who may be or from the first agent .

We import an parameter avoiding time to optimize the agent selection strategy. If a user gets an AC at time , he cannot get that AC again before . For example, receives a certain AC from at , then sends it to a user C () who can also send it to others. If a user carrying encounters before , he cannot send to . Therefore, if the parameter is larger than or equal to the parameter *AT*, there is no repeated agent in an AC’s EQ. In other words, an AC never reaches an agent twice with an equal to *AT*. In this chapter, we assign to *AT*.

Besides, we should also avoid ACs having the same sequence of agents, as we mention in the distributing segment subsection. Now we propose the strategy of exchanging distributing ACs.

Let us take a pair of users Alice and Bob as an example. If Alice encounters another user Bob, Bob tells Alice whether he trusts her (She is viewed as a friend by Bob). Alice picks one of her distributing AC lists (e.g. *DL*1). Alice traverses all ACs in *DL*1, and distributing ACs which match all the following two conditions are exchanged to Bob.

1. If the length of the AC’s EQ (i.e. EQL) is not shorter than , then Alice must be a friend of Bob.
2. Bob does not carry the AC in the recent (time).

When Alice is sending a distributing AC to Bob, she adds her identity and recent time to the AC’s EQ. Besides, Alice must modify the AC’s to her own identity and its to a new one. The information in Table 4.4 are recorded in her memory where is called the *relay table*. We should notice that the first agent (i.e., the creator) does not have a , because he is exactly the one who generates the AC, then the should be his own identity. When Bob gets those ACs, he puts each one of the received *distributing* ACs to his *DL*s respectively and randomly. If an AC whose EQL is already equal to , the AC must be switched to a *ready* one when Bob gets it. Bob puts ready ACs in his ready AC list instead of distributing AC lists.

Table 4.4 relay table entries

|  |  |
| --- | --- |
| Name of entries | descriptions |
|  | The generated by the previous agent |
|  | The identity of the previous agent |
|  | The new (generated by himself) |
|  | The identity of the next agent |
|  | The length of the AC’s EQ (should larger than 0) |
|  | The time when the AC timeout. |

### Exchange ready appointment cards

Users asks for Ready ACs only from their friends, which ensures that the information of the ready ACs hold by a user is not exposed to strangers. The strategy of exchanging ready ACs faces a problem. That is, the number of each user’s ready ACs should be similar.

We also take friends Alice and Bob as an example. Alice has 20 ACs, while Bob has 10 ACs. When they encounter each other, it is reasonable that they both gives half of their ready ACs to the other one. As a result, they both will have half of total (15) ACs. However, that strategy does not work all the time.

The problem becomes more complex in the following condition. Alice is a friend of Bob, while Bob is not a friend of Alice. In other words, Bob trusts Alice, but Alice does not trust Bob. We assume that Alice is a trustful but suspicious girl, so that many users trust her while she trusts few users while Bob is opposite. When users encounter Alice, they ask for ready ACs from her, but Alice rarely asks for ready ACs. Therefore, Alice carries few ready ACs, while Bob carries so many ready ACs. When Alice and Bob encounter each other, it is Bob who is asking for ready ACs instead of Alice. To make the strategy fair and efficient, users must compare the number of their own ready ACs and the number of ready ACs carrying by the other user when two users are exchanging ready ACs.

When Bob encounters Alice, Bob tells Alice the number of his ready ACs () and whether he wants Alice’s ready ACs (if he trusts Alice). If Alice learns that Bob needs her ready ACs, she compares the number of her ready ACs () and . If , Alice gives no ready AC to Bob; otherwise, she gives to Bob. In this way, ready ACs do not concentrate in a group of users who trust many users.

The process of exchanging ready ACs is much simpler than that for distributing ACs. Users do not modify any information in the ready ACs including and , so that ready ACs keep static after they leave the last () agent.

### Sending queries

For an original requester who want to send queries to the LBSP, his query must be delivered to the LBSP while the LBSP cannot learn the identity of him. Besides, the reply message from the LBSP must be delivered back to the original requester. The basic idea is that the original requester sends his query using another user’s identity to the LBSP, so that LBSP can reply to that user who is responsible to forward the reply to the original requester. That user is the first agent () of the AC which is used by the queries of the original requester.

In order to enable LBSP to reply to , the original requester’s query includes an sender identity , which is equal to the in the AC. Since needs a to identify the AC used by the query (and the reply), the in the AC is also in the query. The network can deliver that query to the destination LBSP easily with any DTN protocols.

We use the example in Figure 4.1. The original requester has an AC (i.e., ) whose creator is . When uses it to send his query , the sender identity is and the of the query is the of , as shown in Figure 4.5

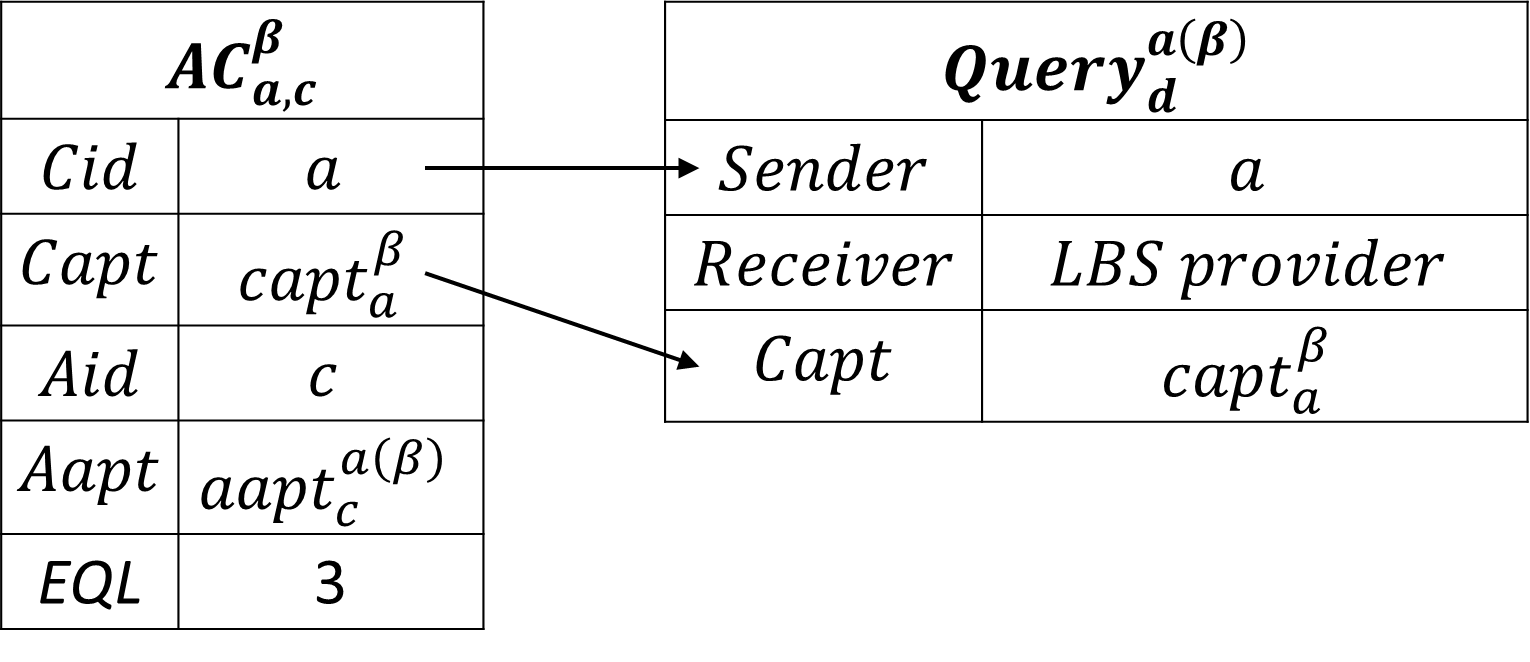


Figure 4.5 constitute query

The AC is marked as used when the query is ready to be sent, so that the AC cannot be exchanged to other users. To get the reply, the original requester also uses a pseudonym, which is described later.

### Sending replies

#### The LBSP part

When the LBSP received the query, it learns that the sender’s identity is the first agent instead of the original requester, which protects the location privacy of the original requester. The LBSP reply to the first agent using the information in the query.

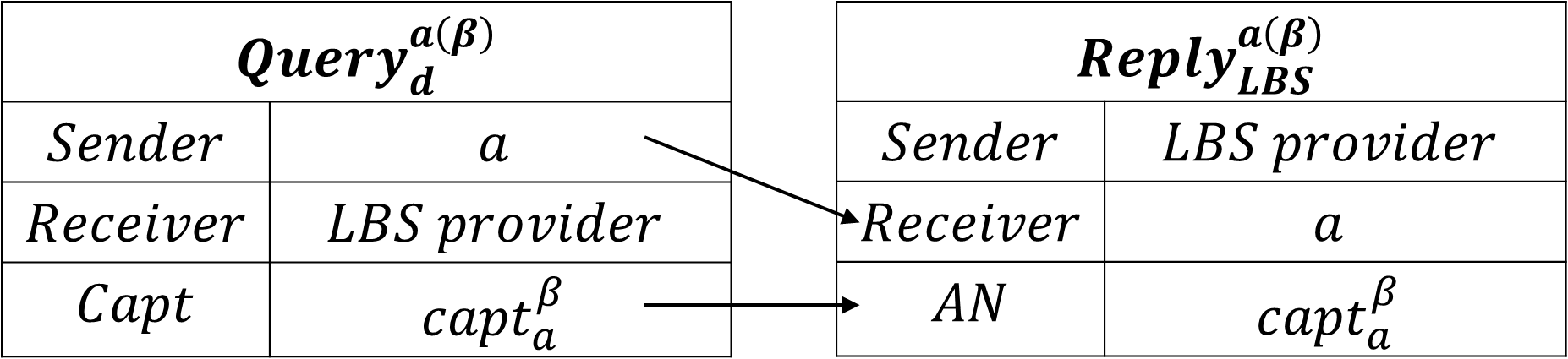


Figure 4.6 constitute replies

In Figure 4.6, the LBS provider reply to the sender , when it receives . The of the query is also included in the reply, which enables to identify the AC used in the query and the reply.

### The first agent

When the first agent (i.e. ) gets the reply from the LBSP, he learns the in the reply. He searches his reply table to get the information of the AC used in the query and reply. If the AC does not expire, there must be a corresponding entry in his reply table, as shown in Table 4.5. gets an entry where the and are equal to his identity and the of the reply. As a result, he learns the identity of the next agent (i.e. ) from the of the entry, so forwards the reply to . The of the reply is replaced with the in the entry by , which enables to identify the AC in ’s reply table.

Table 4.5 reply table entries of the first agent

|  |  |
| --- | --- |
| Name of entries | value |
|  | The user’s own identity |
|  | The of the AC used in the reply (query) |
|  | The identity of the second agent |
|  | The given to the second agent by the user. |
|  |  |
|  | The time when the AC timeout. |

For the example in Figure 4.1, the first agent is . When he receives , he learns that it is a reply from the LBS and the of the AC is . He searches his reply table for an entry whose is equal to and is equal to . As shown in Figure 4.7, the in the entry is equal to so he modifies the receiver of the reply message to . also modifies the in the reply message to which enables identifies the AC in his reply table, because the in the entry is equal to .

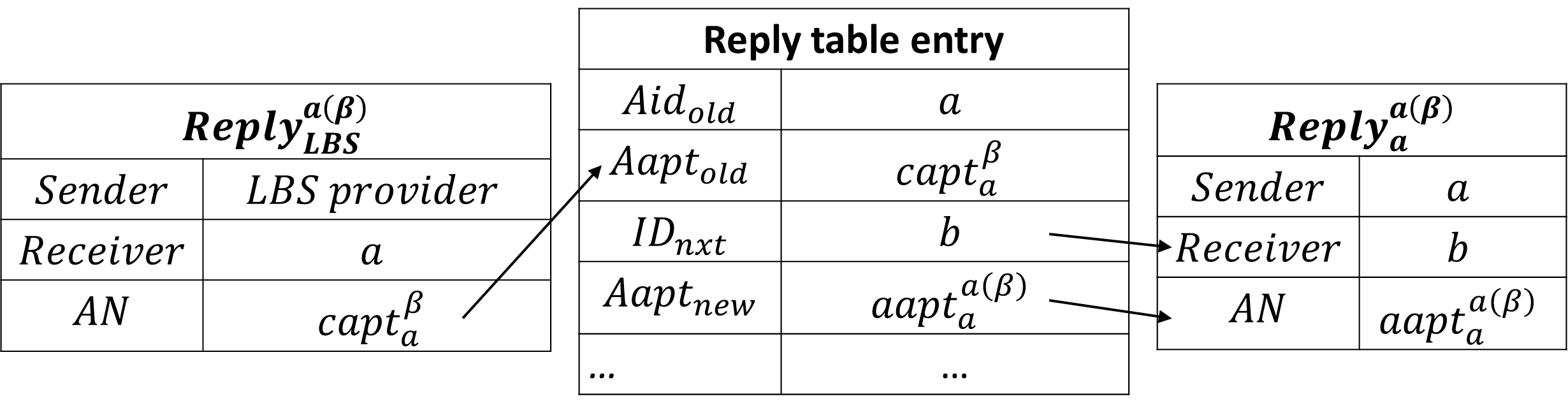


Figure 4.7 the reply of the first agent

### Intermediate agents

The process of forwarding replies in the intermediate agents (the second to the one) is similar with that of the first agent. We take the second agent as an example. When the second agent receives the reply forwarded by the first one, he learns the sender’s identity (i.e. the first agent) and the from the reply. Since he gets the AC which is used in the reply (query) from the previous agent, there must be an entry, whose is equal to the previous agent’s identity and the is the value of in the AC when he gets it, in his reply table, as shown in Table 4.6.

Table 4.6 reply table entries of the second agent

|  |  |
| --- | --- |
| Name of entries | value |
|  | The identity of the previous agent (i.e. the first agent) |
|  | The given by the previous agent. |
|  | The identity of the next agent (e.g. the third agent) |
|  | The given to the next agent by the user. |
|  | 2 |
|  | The time when the AC timeout. |

For the example in Figure 4.1, the second agent is . When he receives , he learns that it is who forwards the reply message and the of the AC is . He searches his reply table for an entry whose is equal to and is equal to . As shown in Figure 4.8, since the in the entry is equal to , he modifies the receiver of the reply message to . also modifies the in the reply message to which enables identifies the AC in his reply table, because the in the entry is equal to .

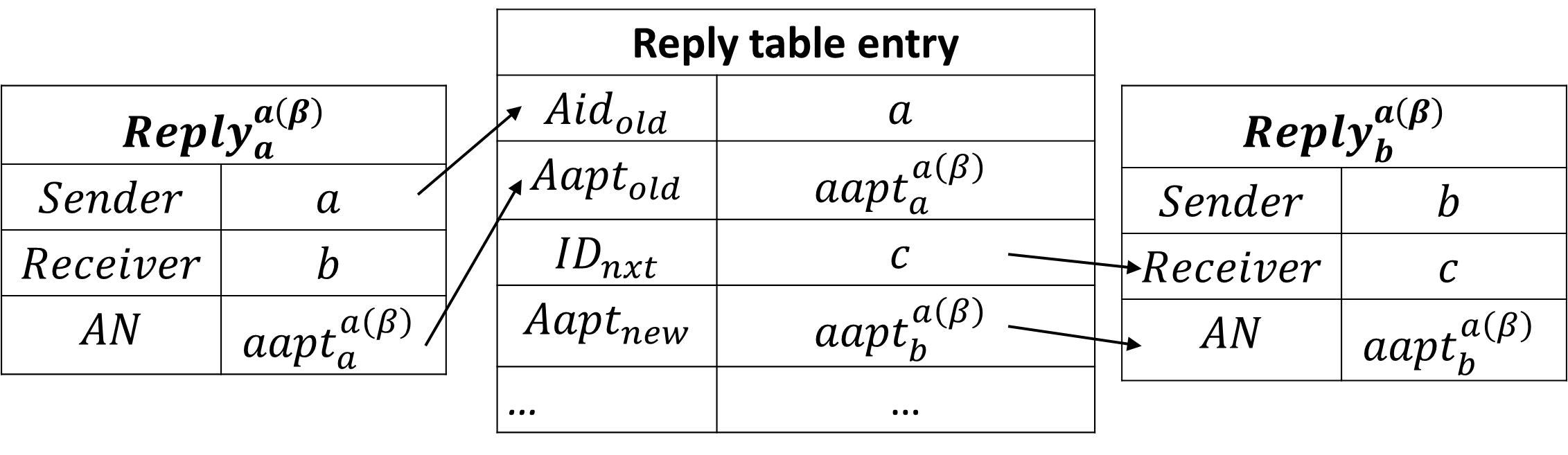


Figure 4.8 the reply of the second agent

For each agent , where , he searches his reply table for a correlative entry, when he receives a reply. The and in the entry should be equal to the sender’s identity and the in the reply. The identity of the next agent is . also assign to the in the reply to help search ’s reply table.

### The last agent

The last agent also searches for a reply table entry based on the reply, while he cannot get the identity of the next agent. The reply table entry of the last agent is shown in Table 4.7.

The last agent uses the same way to look up an entry in his reply table as the previous agents. When he finds that the is equal to , he notices that he is the last agent. Then he is responsible to forward the reply to the original requester instead of another agent. He uses his identity and in his reply table entry to generates a pseudonym , where the function is a public pseudonym generating function which everyone in the network knows, including the original requester.

When the original requester sends his query, he also gets the same pseudonym using the pseudonym generating function. Note that he can get parameters from the AC. and are equal to the and in the AC. He uses that pseudonym as his identity before he gets the reply.

Table 4.7 reply table entries of the last agent

|  |  |
| --- | --- |
| Name of entries | value |
|  | The identity of the previous agent (i.e. the first agent) |
|  | The given by the previous agent. |
|  | VOID |
|  | The given to the original requester. |
|  |  |
|  | The time when the AC timeout. |

When users deliver the reply from the last agent, they are looking for a user whose identity is that pseudonym. At last, the original requester gets the reply, because he is the only user who uses the pseudonym as his identity.

For the example in Figure 4.1, the last agent is . When he receives , he learns that it is who forwards the reply message and the of the AC is . He searches his reply table for an entry whose is equal to and is equal to . Since the in the entry is equal to 3 (i.e. ), he recognizes that he is the last agent. calculates the pseudonym then forward the reply to . The original requester gets the identity and the from the AC he uses, so that he uses the pseudonym as his identity. As a result, get the reply from . The of the reply is also assigned to to avoid identical pseudonyms. The process is shown in Figure 4.9.

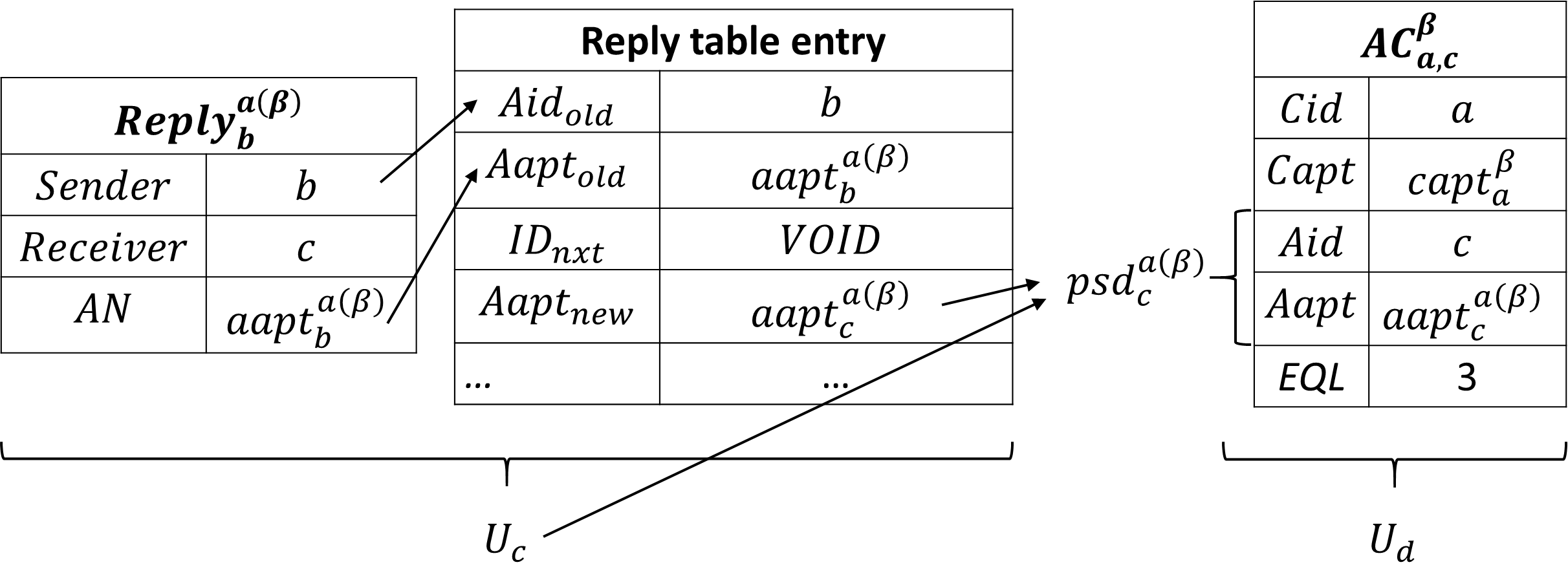


Figure 4.9 The reply of the last agent

## Appointment number

The Appointment Number (AN) including and is significant information in the AC. We explain the rules of generating them in detail and talk about the effect of ACs’ timeout mechanism in this section.

### Creator appointment number

The Creator Appointment Number () is a number which is used to identify ACs generated by the same creator. In other words, if two ACs are generated by the same creator and they do not expire, their must be different. Therefore, the first agent (i.e., the creator) cannot find two entries which have the same in his reply table, if their are both his own identity.

### Agent appointment number

The Agent appointment number () is a number used to identify appointment cards who have the same agent. Agents generate a new s for appointment cards before they exchange those cards to others. In other words, an agent gives any appointment card passing on by him a unique , which helps the next agent identify the appointment card in his reply table. Since the is unique, an agent who is not the first one cannot find two entries which have the same pair of and , neither.

### Timeout

Agents delete the entries, which contains the information of expired ACs, from his reply table. Consequently, agents cannot forward replies using ACs which expire before this very moment. That is the reason why an original requester must choose a AC which expires after his query and reply timeout. While the timeout mechanism has more advantage than disadvantage.

Since users are moving, the distance between agents and the original requester might be too large after a long time. As a result, it is hard for agents to forward the reply messages back to the original requester, so that the original requester takes a higher risk when he uses an AC which is generated for a long time. Then these kind of ACs might not be used after a period, while it cost agents a few memories to save the ACs’ information in their reply table. Therefore, all users remove the information of expired ACs to save their memories.

We should also notice that an unexpired AC and an expired one might have the same or . Because no agent keeps the record of the expired AC, so that the duplication cannot confuse agents.

## Example

In the following examples, we use 4 users (i.e., Alice, Bob, Charlie, David and Elizabeth) to show the process of the ACP. We should notice that David trusts Charlie (Charlie does not trust David), Elizabeth trusts David (David does not trust Elizabeth). Users do not trust others except the about two pairs. The parameters are shown in Table 4.8.

Table 4.8 example parameters

|  |  |  |
| --- | --- | --- |
| Parameters | explanation | |
|  | 3 |
|  | 1 |
|  | 2 |
|  | 10 minutes |
|  | 80 minutes |
|  | 80 minutes (equal to *AT*) |

We abstract the event that users encounter each other as a table, as shown in Table 4.9. The table list the time when two users encounter each other. For example, Bob and Charlie encounter each other at the third minute. After two users meet, they separate in one minute.

Table 4.9 Example event

|  |  |  |
| --- | --- | --- |
| User 1 | User 2 | Time (minute) |
| Elizabeth | Bob | 1 |
| Bob | Charlie | 2 |
| Charlie | Elizabeth | 3 |
| Elizabeth | Alice | 4 |
| Bob | Charlie | 5 |
| Alice | Charlie | 6 |
| Charlie | David | 7 |
| David | Elizabeth | 8 |

We assume that these users do not encounter any user for a long time, so that they each has 8 () ACs generated by themselves. We also assume that no AC expires in our 8-minutes example. Since the parameter *Seg* is equal to 2, each user has 2 *Distributing AC List*s (*DL*s). The initial state of our example is shown in Table 4.10. To make the example clearer, we do not place ACs in DLs randomly, ACs are always placed in the DL alternately. For example, if we put the first AC in DL1, then the second AC must be in the DL2. Users also obey the rule when they receive ACs from other users.

Table 4.10 Example initial state

|  |  |  |  |
| --- | --- | --- | --- |
| User Identity | DL1 | DL2 | RL |
| Alice (A) |  |  | ∅ |
| Bob (B) |  |  | ∅ |
| Charlie (C) |  |  | ∅ |
| David (D) |  |  | ∅ |
| Elizabeth (E) |  |  | ∅ |

The explanation of symbols used in table are shown in the former Table 4.1. To make the symbol clearer, we add the EQL of the AC on its symbol. For example, if the EQL of is equal to 2, it is marked as .

### Exchanging appointment cards

To make it easier, users do not pick DLs randomly as we described in the previous sections when they exchange ACs. Instead, when encounters , always picks his DL1 and always picks his DL2.

The tables in current section show the entire Distributing AC Lists (i.e. DL1 and DL2) and entire relay tables (RLs) of each user. We use the initial letter to denote each user in symbols. For example, D denotes David. In the reply tables, we color old entries with darker color to high light new ones. The structure of the AC-list tables is shown in Table 4.11, where we mark the picked DL with a darker color.

Table 4.11 User X and Y's AC lists

|  |  |  |  |
| --- | --- | --- | --- |
| X | DL1 | Before | X’s distributing AC list 1 before their exchange. |
| After | X’s distributing AC list 1 after their exchange. |
| DL2 | Before | X’s distributing AC list 2 before their exchange. |
| After | X’s distributing AC list 2 after their exchange. |
| RL | Before | X’s ready AC list before their exchange. |
| After | X’s ready AC list after their exchange. |
| Y | DL1 | Before | Y’s distributing AC list 1 before their exchange. |
| After | Y’s distributing AC list 1 after their exchange. |
| DL2 | Before | Y’s distributing AC list 2 before their exchange. |
| After | Y’s distributing AC list 2 after their exchange. |
| RL | Before | Y’s ready AC list before their exchange. |
| After | Y’s ready AC list after their exchange. |

The process of exchanging AC is shown step by step as following.

1. Elizabeth encounters Bob

At the first minute, Elizabeth encounters Bob. Elizabeth picks her DL1, and Bob picks his DL2. They give distributing ACs in their picked DLs to each other and update their reply tables, as shown in Table 4.12 and Table 4.13.

Table 4.12 Elizabeth and Bob's AC lists at time 1

|  |  |  |  |
| --- | --- | --- | --- |
| Elizabeth | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |
| Bob | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |

Table 4.13 Elizabeth and Bob's relay table at time 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Elizabeth | | | | | Bob | | | | |
|  |  |  |  | EQL |  |  |  |  | EQL |
| E |  | B |  | 1 | B | B |  | E |  |
| E |  | B |  | 1 | B | B |  | E |  |
| E |  | B |  | 1 | B | B |  | E |  |
| E |  | B |  | 1 | B | B |  | E |  |

1. Bob encounters Charlie

At the second minute, Bob encounters Charlie. Bob picks his DL1, and Charlie picks his DL2. They give distributing ACs in their picked DLs to each other and update their reply tables, as shown in Table 4.14 and Table 4.15.

Table 4.14 Bob and Charlie's AC lists at time 2

|  |  |  |  |
| --- | --- | --- | --- |
| Bob | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |
| Charlie | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |

Table 4.15 Bob and Charlie's relay table at time 2

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bob | | | | | Charlie | | | | |
|  |  |  |  | EQL |  |  |  |  | EQL |
| B |  | E |  | 1 | C |  | B |  | 1 |
| B |  | E |  | 1 | C |  | B |  | 1 |
| B |  | E |  | 1 | C |  | B |  | 1 |
| B |  | E |  | 1 | C |  | B |  | 1 |
| B |  | C |  | 1 |  |  |  |  |  |
| B |  | C |  | 1 |  |  |  |  |  |
| B |  | C |  | 1 |  |  |  |  |  |
| B |  | C |  | 1 |  |  |  |  |  |
| E |  | C |  | 2 |  |  |  |  |  |
| E |  | C |  | 2 |  |  |  |  |  |

1. Charlie encounters Elizabeth

At the third minute, Charlie encounters Elizabeth. Charlie picks his DL1, and Elizabeth picks her DL2. They give distributing ACs in their picked DLs to each other and update their reply tables, as shown in Table 4.16 and Table 4.17.

We should notice that Charlie does not give to Elizabeth, because Elizabeth does not trust Charlie. The EQL of the has already reached 2, which is equal to . As mentioned in 4.6.2, these kinds of ACs are only sent to users who trust the current agent. Therefore, Charlie cannot give to Elizabeth when she does not trust him.

Table 4.16 Charlie and Elizabeth's AC lists at time 3

|  |  |  |  |
| --- | --- | --- | --- |
| Charlie | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |
| Elizabeth | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |

Table 4.17 Charlie and Elizabeth's relay table at time 3

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Charlie | | | | | Elizabeth | | | | |
|  |  |  |  | EQL |  |  |  |  | EQL |
| C |  | B |  | 1 | E |  | B |  | 1 |
| C |  | B |  | 1 | E |  | B |  | 1 |
| C |  | B |  | 1 | E |  | B |  | 1 |
| C |  | B |  | 1 | E |  | B |  | 1 |
| C |  | E |  | 1 | E |  | C |  | 1 |
| C |  | E |  | 1 | E |  | C |  | 1 |
| C |  | E |  | 1 | E |  | C |  | 1 |
| C |  | E |  | 1 | E |  | C |  | 1 |
| B |  | E |  | 2 | B |  | C |  | 2 |
| B |  | E |  | 2 | B |  | C |  | 2 |

1. Elizabeth encounters Alice

At the fourth minute, Elizabeth encounters Alice. Elizabeth picks her DL1, and Alice picks her DL2. They give distributing ACs in their picked DLs to each other and update their reply tables, as shown in Table 4.18 and Table 4.19.

Elizabeth does not give to Alice for the same reason as the previous Charlie, that is, Alice does not trust Elizabeth.

Table 4.18 Elizabeth and Alice's AC lists at time 4

|  |  |  |  |
| --- | --- | --- | --- |
| Elizabeth | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |
| Alice | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |

Table 4.19 Elizabeth and Alice's relay table at time 4

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Elizabeth | | | | | Alice | | | | |
|  |  |  |  | EQL |  |  |  |  | EQL |
| E |  | B |  | 1 | A |  | E |  | 1 |
| E |  | B |  | 1 | A |  | E |  | 1 |
| E |  | B |  | 1 | A |  | E |  | 1 |
| E |  | B |  | 1 | A |  | E |  | 1 |
| E |  | C |  | 1 |  |  |  |  |  |
| E |  | C |  | 1 |  |  |  |  |  |
| E |  | C |  | 1 |  |  |  |  |  |
| E |  | C |  | 1 |  |  |  |  |  |
| B |  | C |  | 2 |  |  |  |  |  |
| B |  | C |  | 2 |  |  |  |  |  |
| B |  | A |  | 2 |  |  |  |  |  |
| B |  | A |  | 2 |  |  |  |  |  |
| C |  | A |  | 2 |  |  |  |  |  |
| C |  | A |  | 2 |  |  |  |  |  |

1. Bob encounters Charlie

At the fifth minute, Bob encounters Charlie. Bob picks his DL1, and Charlie picks his DL2. They give distributing ACs in their picked DLs to each other and update their reply tables, as shown in Table 4.20 and Table 4.21.

Charlie does not give and to Bob, because Bob does not trust Charlie. Since we set the avoiding time equal to *AT*, a distributing AC cannot be given back to previous agents. As a result, , , and are not exchanged by Charlie and Bob.

Table 4.20 Bob and Charlie's AC lists at time 5

|  |  |  |  |
| --- | --- | --- | --- |
| Bob | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |
| Charlie | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |

Table 4.21 Bob and Charlie's relay table at time 5

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bob | | | | | Charlie | | | | |
|  |  |  |  | EQL |  |  |  |  | EQL |
| B |  | E |  | 1 | C |  | B |  | 1 |
| B |  | E |  | 1 | C |  | B |  | 1 |
| B |  | E |  | 1 | C |  | B |  | 1 |
| B |  | E |  | 1 | C |  | B |  | 1 |
| B |  | C |  | 1 | C |  | E |  | 1 |
| B |  | C |  | 1 | C |  | E |  | 1 |
| B |  | C |  | 1 | C |  | E |  | 1 |
| B |  | C |  | 1 | C |  | E |  | 1 |
| E |  | C |  | 2 | B |  | E |  | 2 |
| E |  | C |  | 2 | B |  | E |  | 2 |
|  |  |  |  |  | E |  | B |  | 2 |
|  |  |  |  |  | E |  | B |  | 2 |

1. Alice encounters Charlie

At the sixth minute, Alice encounters Charlie. Alice picks her DL1, and Charlie picks his DL2. They give distributing ACs in their picked DLs to each other and update their reply tables, as shown in Table 4.22 and Table 4.23.

Table 4.22 Alice and Charlie's AC lists at time 6

|  |  |  |  |
| --- | --- | --- | --- |
| Bob | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |
| Charlie | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |

Table 4.23 Alice and Charlie's relay table at time 6

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Alice | | | | | Charlie | | | | |
|  |  |  |  | EQL |  |  |  |  | EQL |
| A |  | E |  | 1 | C |  | B |  | 1 |
| A |  | E |  | 1 | C |  | B |  | 1 |
| A |  | E |  | 1 | C |  | B |  | 1 |
| A |  | E |  | 1 | C |  | B |  | 1 |
| A |  | C |  | 1 | C |  | E |  | 1 |
| A |  | C |  | 1 | C |  | E |  | 1 |
| A |  | C |  | 1 | C |  | E |  | 1 |
| A |  | C |  | 1 | C |  | E |  | 1 |
|  |  |  |  |  | B |  | E |  | 2 |
|  |  |  |  |  | B |  | E |  | 2 |
|  |  |  |  |  | E |  | B |  | 2 |
|  |  |  |  |  | E |  | B |  | 2 |
|  |  |  |  |  | B |  | A |  | 2 |
|  |  |  |  |  | B |  | A |  | 2 |

Since Alice and Charlie do not trust each other, , , and are not exchanged by the two users.

1. Charlie encounters David

At the seventh minute, Charlie encounters David. Charlie pick his DL1, and David picks his DL2. They give distributing ACs in their picked DLs to each other and update their reply tables, as shown in Table 4.24 and Table 4.25.

Because David trust Charlie, Charlie can give and to David. When David receives and , he notices that the EQLs of the two ACs reach so that he puts these two ACs in his ready list instead of distributing lists.

Table 4.24 Charlie and David's AC lists at time 7

|  |  |  |  |
| --- | --- | --- | --- |
| Charlie | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After | ∅ |
| David | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After |  |

Table 4.25 Charlie and David's relay table at time 7

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Charlie | | | | | David | | | | |
|  |  |  |  | EQL |  |  |  |  | EQL |
| C |  | B |  | 1 | D |  | C |  | 1 |
| C |  | B |  | 1 | D |  | C |  | 1 |
| C |  | B |  | 1 | D |  | C |  | 1 |
| C |  | B |  | 1 | D |  | C |  | 1 |
| C |  | E |  | 1 |  |  |  |  |  |
| C |  | E |  | 1 |  |  |  |  |  |
| C |  | E |  | 1 |  |  |  |  |  |
| C |  | E |  | 1 |  |  |  |  |  |
| B |  | E |  | 2 |  |  |  |  |  |
| B |  | E |  | 2 |  |  |  |  |  |
| E |  | B |  | 2 |  |  |  |  |  |
| E |  | B |  | 2 |  |  |  |  |  |
| B |  | A |  | 2 |  |  |  |  |  |
| B |  | A |  | 2 |  |  |  |  |  |
| B |  | VOID |  | 3 |  |  |  |  |  |
| E |  | D |  | 2 |  |  |  |  |  |
| E |  | D |  | 2 |  |  |  |  |  |
| E |  | VOID |  | 3 |  |  |  |  |  |
| A |  | D |  | 2 |  |  |  |  |  |
| A |  | D |  | 2 |  |  |  |  |  |

1. David encounters Elizabeth

At the eighth minute, David encounters Elizabeth. David picks his DL1, and Elizabeth picks her DL2. They give distributing ACs in their picked DLs to each other and update their reply tables, as shown in Table 4.26 and Table 4.27.

After David gives Elizabeth the distributing ACs, Elizabeth gets a ready AC . Then Elizabeth has 1 ready AC, while David has 2 ready ACs, so that David has only one more ready-AC than Elizabeth. That is the reason why David does not give his own ready ACs to Elizabeth, even though Elizabeth trusts David.

We assume that Elizabeth does not get , so that she has no ready AC. Then David should give her a ready AC from his own two randomly.

Table 4.26 David and Elizabeth's AC lists at time 8

|  |  |  |  |
| --- | --- | --- | --- |
| David | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before |  |
| After |  |
| Elizabeth | DL1 | Before |  |
| After |  |
| DL2 | Before |  |
| After |  |
| RL | Before | ∅ |
| After |  |

Table 4.27 David and Elizabeth's relay table at time 8

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| David | | | | | Elizabeth | | | | |
|  |  |  |  | EQL |  |  |  |  | EQL |
| D |  | C |  | 1 | E |  | B |  | 1 |
| D |  | C |  | 1 | E |  | B |  | 1 |
| D |  | C |  | 1 | E |  | B |  | 1 |
| D |  | C |  | 1 | E |  | B |  | 1 |
| D |  | E |  | 1 | E |  | C |  | 1 |
| D |  | E |  | 1 | E |  | C |  | 1 |
| D |  | E |  | 1 | E |  | C |  | 1 |
| D |  | E |  | 1 | E |  | C |  | 1 |
| C |  | VOID |  | 3 | B |  | C |  | 2 |
|  |  |  |  |  | B |  | C |  | 2 |
|  |  |  |  |  | B |  | A |  | 2 |
|  |  |  |  |  | B |  | A |  | 2 |
|  |  |  |  |  | C |  | A |  | 2 |
|  |  |  |  |  | C |  | A |  | 2 |
|  |  |  |  |  | C |  | D |  | 2 |
|  |  |  |  |  | C |  | D |  | 2 |
|  |  |  |  |  | A |  | D |  | 2 |
|  |  |  |  |  | A |  | D |  | 2 |

### Queries and Replies

We still use the previous example and assume that the delivery of queries and replies cost little time so that no AC expires in the period.

We assume that David wants to send a query to the LBS. He chooses a AC (e.g. ) from his two ACs (i.e. and ) randomly. If we review the previous tables, we learn that is generated by Elizabeth and has 3 agents (i.e., Elizabeth, Bob and Charlie). When David uses the AC to send his query, the sender identity of the query is Elizabeth. The creator appointment number (i.e. ) is also included in the query. Then David uses the identity of the last agent (i.e. Charlie) and the agent appointment number (i.e. ) to generate a pseudonym , which is used as a temporary identity of David before he receives the reply. Besides, cannot be given to any user until it expires. The process is shown in Figure 4.10.

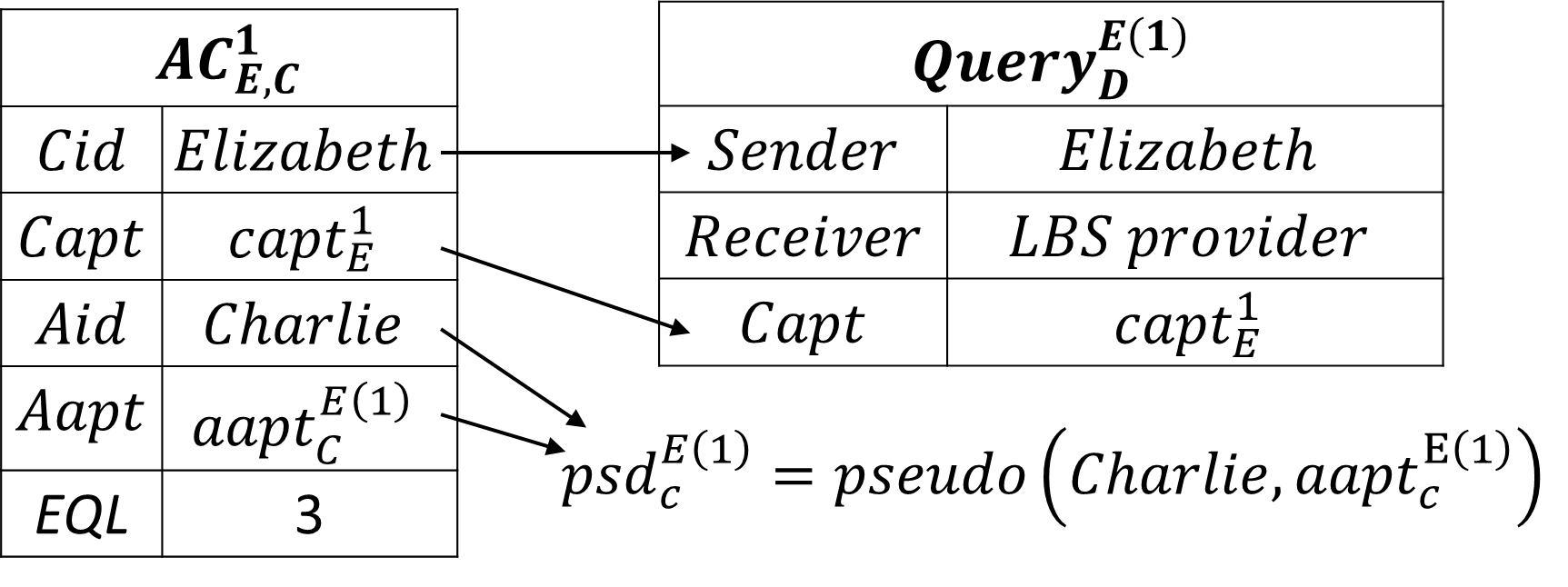


Figure 4.10 David's query

The LBS sends a reply message to Elizabeth based on the sender identity of the query. The creator appointment number (i.e. ) is also included in the reply. If an attacker learns these messages, he believes that it is Elizabeth who sends the query.

When Elizabeth receives the reply message, she searches her reply table for the AC used in the reply message, as shown in Table 4.27. Since the message is sent by the LBS, the should equal to her own identity, and the should equal to the creator appointment number. Then she must find an entry like Table 4.28, which enables Elizabeth to modify the reply message and forward to Bob, as shown in Figure 4.11.

Table 4.28 Elizabeth's reply table entry

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | EQL |
| E |  | B |  | 1 |

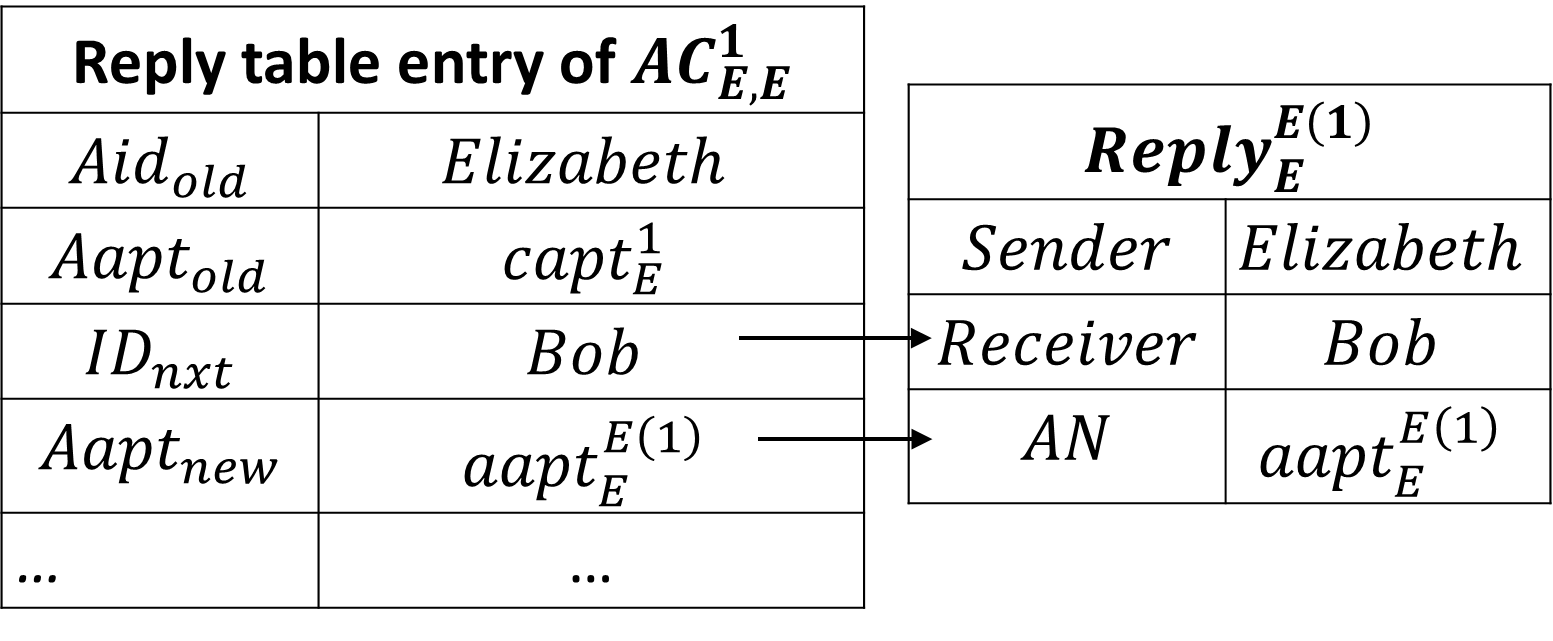


Figure 4.11 Elizabeth forwards a reply

If Bob receives the reply message successfully, he checks his reply table as shown in Table 4.21. He finds an entry whose is equal to Elizabeth and is equal to , as shown in Table 4.29. In the same way as Elizabeth, Bob forwards reply message to Charlie, as shown in Figure 4.12.

Table 4.29 Bob's reply table entry

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bob | | | | |
|  |  |  |  | EQL |
| E |  | C |  | 2 |

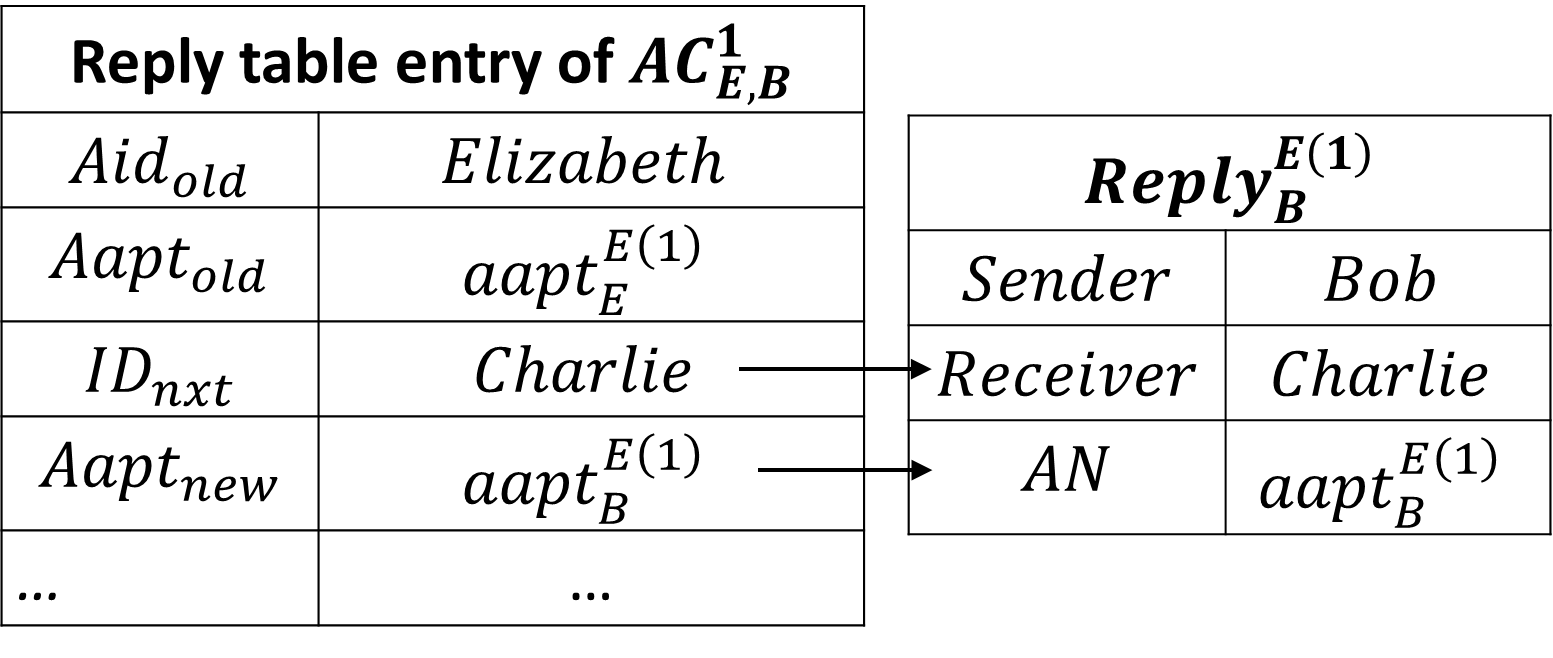


Figure 4.12 Bob forwards a reply

When the Charlie receives the reply message from Bob, Charlie learns that he is the last agent based on his reply table as shown in Table 4.25 and Table 4.30, because the EQL is equal to 3 and the is void.

Table 4.30 Charlie's reply table entry

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Charlie | | | | |
|  |  |  |  | EQL |
| B |  | VOID |  | 3 |

To forward to the unknown original requester (i.e. David), Charlie generates a pseudonym using his own identity and . We should notice that the pseudonym is exactly the temporary identity of David. Then David modifies the reply message and forwards it, as shown in Figure 4.13.

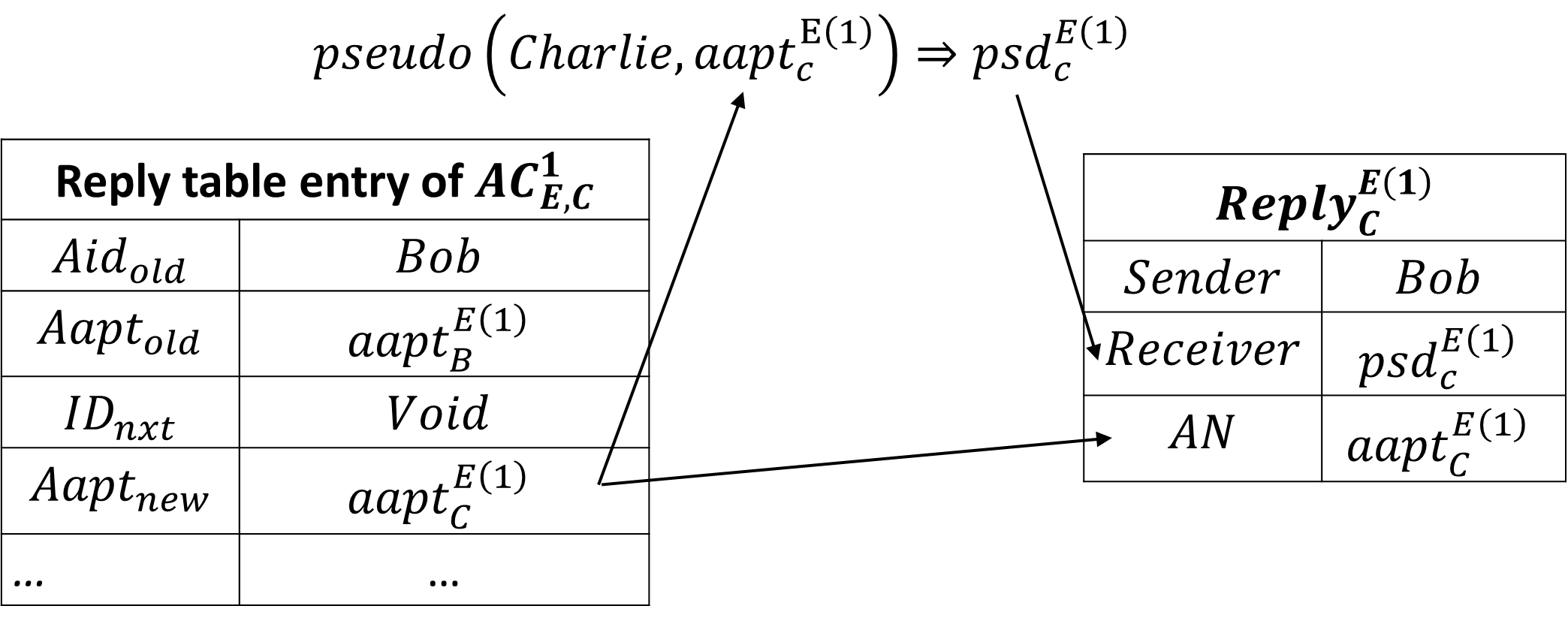


Figure 4.13 Charlie forwards a reply

When a user who carries encounters David, the user forwards the reply to him, so that David gets his reply.

## Experiment

We use the map of Helsinki in our simulator to evaluate ACP. It is also compared against Binary Spray and Wait (BSW) [22], distributed social based location privacy protocol (SLPD) [7] and our Multi-Hop Location-Privacy Protection (MHLPP). Users are moving on the map along streets continuously, while there is a LBSP fixed at a random location on the map.

For each user, we give him random social values between 0% and 100%, each corresponding to all other users. Each value has the same probability, so we can compute the expected number of friends of an user. If a user whose social value is larger than 85% is called a friend and there are *n* users in the network, there are friends.

The Shortest Path Map-Based Movement (SPMBM) [26] is used in our experiment. For each experiment, we give the simulator a random seed so that it can generate pseudo-random number based on the seed. Therefore, all the factors including users’ speed and locations are the same if two experiments have the same random seed. All those four protocols are tested using the same serial of random seed.

Before each experiment, the simulator runs for 800 seconds (simulator time). Then we pick 100 users out of 126 users randomly, and each of them send a query to the LBSP. Tests last for about 20 minutes (simulator time).

### Average query success ratio

The query success ratio is the percentage of delivered queries among a number of attempts. Since users sending 100 queries in each experiment totally, if queries are delivered to the LBSP at time , the query success ratio of time is .

As shown in Figure 4.14, we compare the average query success ratio of the four protocols with 5 kinds of communication radius (10, 30, 50, 70 and 90 meters). We observe that the ACP and the BSW get a high query success ratio, while the MHLPP and the SLPD are lower than the former two protocols. The BSW is the highest one, because it is a no-privacy protocol. The ACP is just a little lower than BSW, because the query delivery process of the ACP is almost the same as that of BSW. Since users of ACP must wait for available appointment cards, they cost more time to initial their queries. But the ACP and the BSW are in the same level, comparing to the other two protocols. The MHLPP and the SLPD need to find friend to obfuscate their queries, which baffles their delivery process.

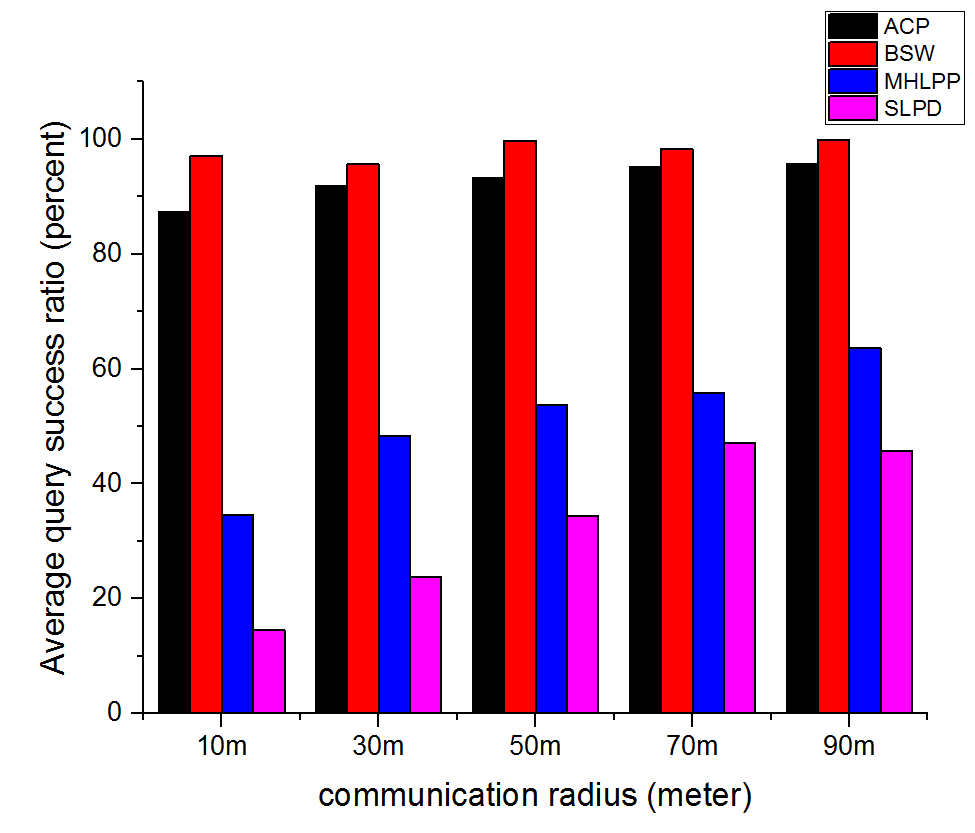


Figure 4.14 average query success ratio (10 minutes)

The experiment results when we test 20 minutes is shown in Figure 4.15. Comparing to the average query success ratio at 10 minutes, the MHLPP and the SLPD achieve much higher success ratio after 20 minutes than at 10 minutes. That is because it cost them so much time in their obfuscation phases when they need to find friends. In fact, some of queries of the MHLPP and the SLPD still do not finish their obfuscation phase at 20 minutes.

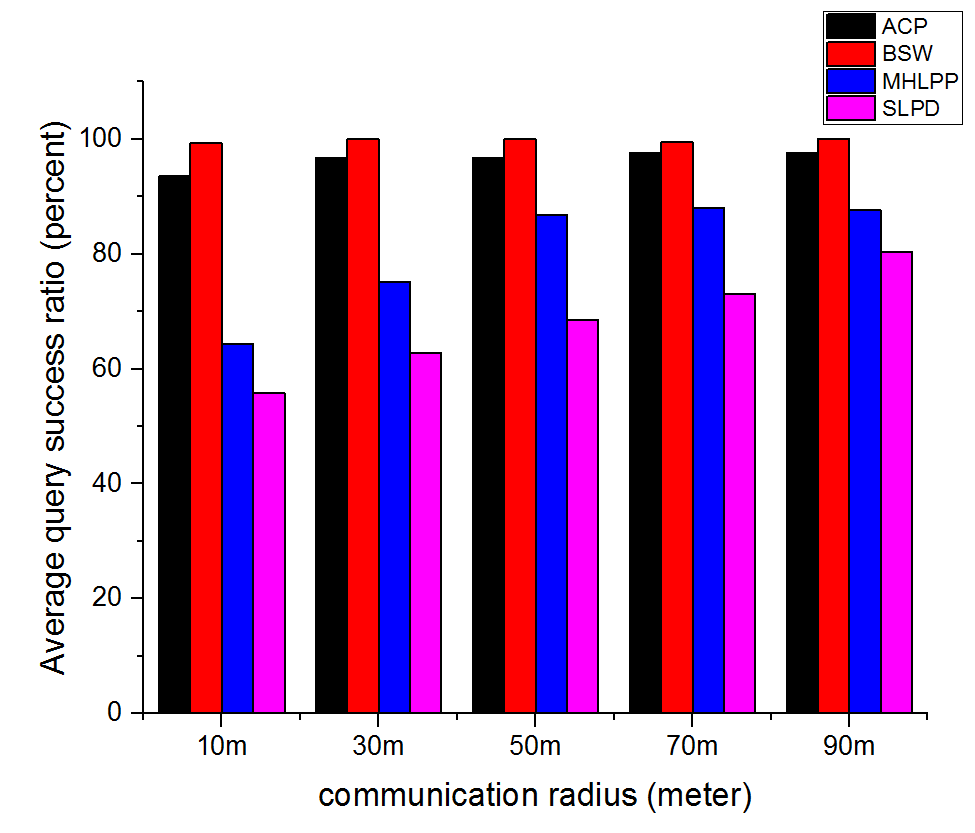


Figure 4.15 average query success ratio (20 minutes)

The communication radius can influence the success ratio. In most of cases, the success ratio rises when we increase the communication radius, and its influence is especially obvious under 50 meters. A large communication radius makes it easily for users to encounter others, which is good for them to forward queries. However, a user who is so far away from the destination does not want the intermediates of his query encounters many users nearby. Because all users who carries copies of that query are near the sender instead of the destination, which decreases their query success ratio. Therefore, when the communication radius reaches 70 meters, the success ratios almost stay at the same level.

In the Figure 4.16, we observe that the ACP and the BSW have better convergence speed than the MHLPP and the SLPD. In other words, the former two protocols have a faster speed to approach the 100% query success ratio than the later two. At the very beginning, the ACP even has a little higher success ratio than the BSW. Because the ACP users need ready appointment card, and most of the users get their first ready appointment cards at places where there are many users. Therefore, users rarely generate query near the edges of the map at the beginning, which facilitates their queries delivery process. For example, if a user generates his query at the edge of the map, the copies of his query will be sent to users who is also at the edge and it takes more time for them to deliver the query. If the user does not generate his query until he arrives a place nearer center, the copies of his query will be carried by the users in the center with higher probability.

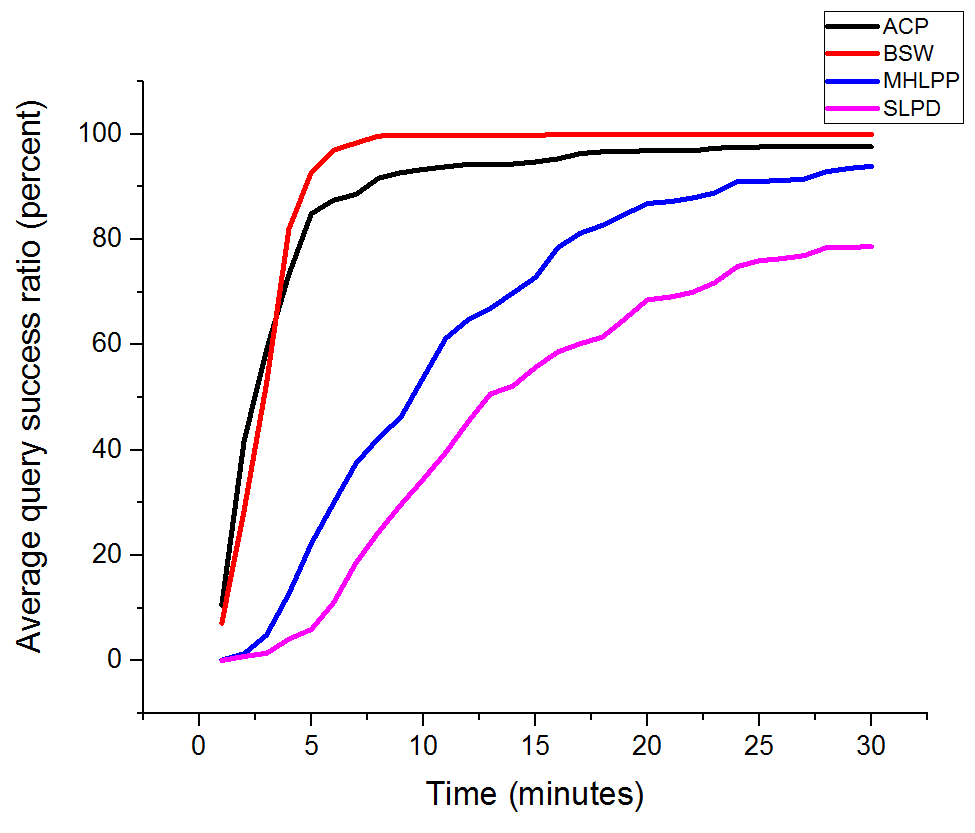


Figure 4.16 average query success ratio with 50-meters communication ratio

### Average reply success ratio

When the LBSP receives a query, it sends a reply to the requester. If the reply arrives the original requester before the test ends, we view it as success, otherwise, the reply is failed. There are two reasons for the failure of replies: 1) the query is not delivered to the LBSP successfully; 2) the query costs too much time so that the reply has no time to be delivered; 3) The route of the reply is too long. Since there are 100 queries in each experiment, the number of replies should be equal to 100.

In Figure 4.17, the BSW has a significant and reasonable higher success ratio than all other protocols, because it is a no-privacy protocol. The ACP is higher than the MHLPP and the SLPD, but its advantage is not as large as that in the query process. In fact, the reply process of the MHLPP and the SLPD are simpler than that of the ACP, but the ACP saves so much time in its query process that it earns a better reply success ratio than the other two.

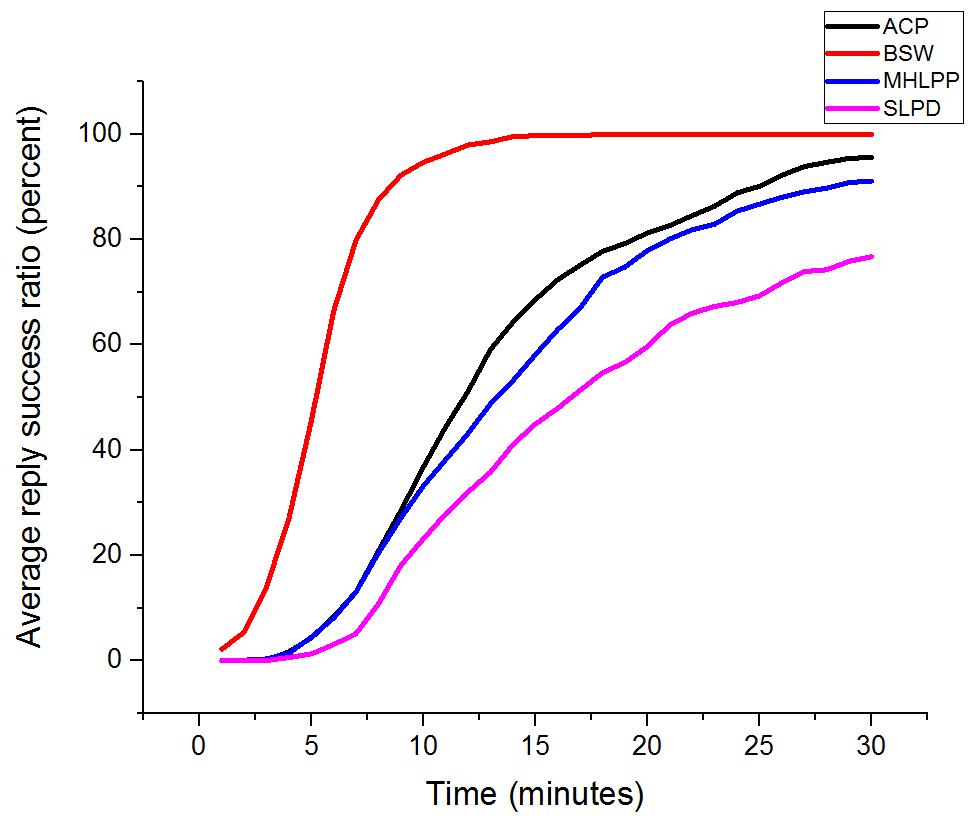


Figure 4.17 average reply success ratio with 50-meters communication ratio

### Total number of query relay

The query delivery processes of all the four protocols use the BSW protocol, and the BSW makes copies for queries and gives half of copies to any users it encounters. That is a significant cost for the network, so we use the number of query relaying (QR) to evaluate that cost. The QR is initialized as zero at the beginning of the test. When a user relay a (or several) copy (copies) of a query, we increase QR by 1. For example, in the SLPD, there are two phases: the obfuscation phase and the free phase. In the obfuscation phase, a query is forwarded among one-hop friends for times. After that, it is forwarded by the BSW protocol. The BSW protocol makes copies for the query and gives half of the copies to any encountering users. Then QR should be about . Since a user gives all its copies to the destination if he encounters the destination at once, the QR may be smaller. The smaller that number is, the smaller cost of the network is. Since there are 100 queries, we divide QR by 100 to get an average value.

In Figure 4.18, we compare QR with four protocols. We observe that all the four protocols are at a similar level, the BSW and the ACP is a little lower than the other two. For the ACP and the BSW, they deliver queries so fast that users who carries more than one copies give all their copies to the destination before they send these copies to different users separately. As a result, many copies have no chance to be forwarded, which decreases the cost. While the MHLPP and the SLPD have obfuscation phases, the queries start to be delivered freely (in a BSW way) at a random place where might be so far way from the destination, so that almost all copies can be forwarded respectively.

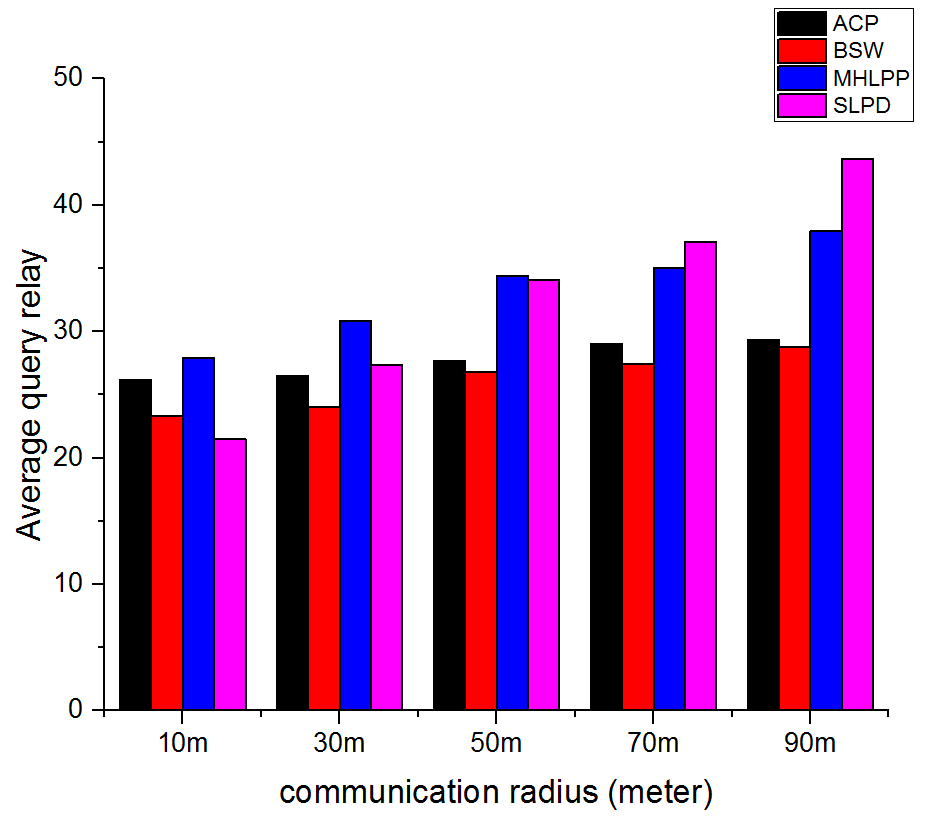


Figure 4.18 average total number of forwarding queries at 20 minutes

The communication radius effects the total number of the forwarding queries, especially for the MHLPP and the SLPD. Those two protocols can finish their obfuscation phase more quickly which a larger communication radius, so that more queries can be forwarded freely (in the BSW way), which makes their QR larger.

### Memory cost

We count the number of queries carried by each user to evaluate the memory cost of the four protocols. Several copies of a query are counted for only once.

In the Figure 4.19, we compare the number of queries per user with the four protocols at 20 minutes. We observe that the BSW is the highest in most of the cases and the ACP always stays at a similar level as the BSW. The data of other two protocols (the MHLPP and the SLPD) increase as the communication radius. The MHLPP even excesses the BSW when the communication radius is 90 meters. The reason is that quite a number of the BSW and the ACP users forward their copies to the destination so that there is no copy with them at 20 minutes, while the rest of them cannot forward their copies to the destination even given a large communication radius. While the number of the other two protocol’s free phase queries is significantly influenced by the communication radius. The more queries are in the free phase, the more copies are in the network.

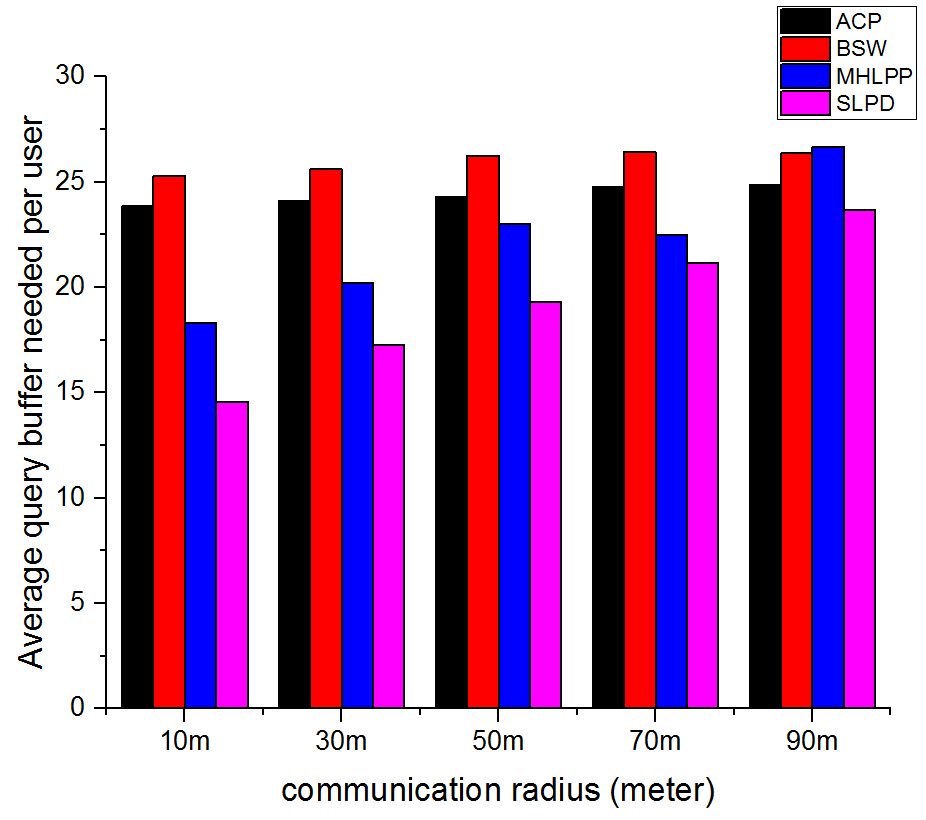
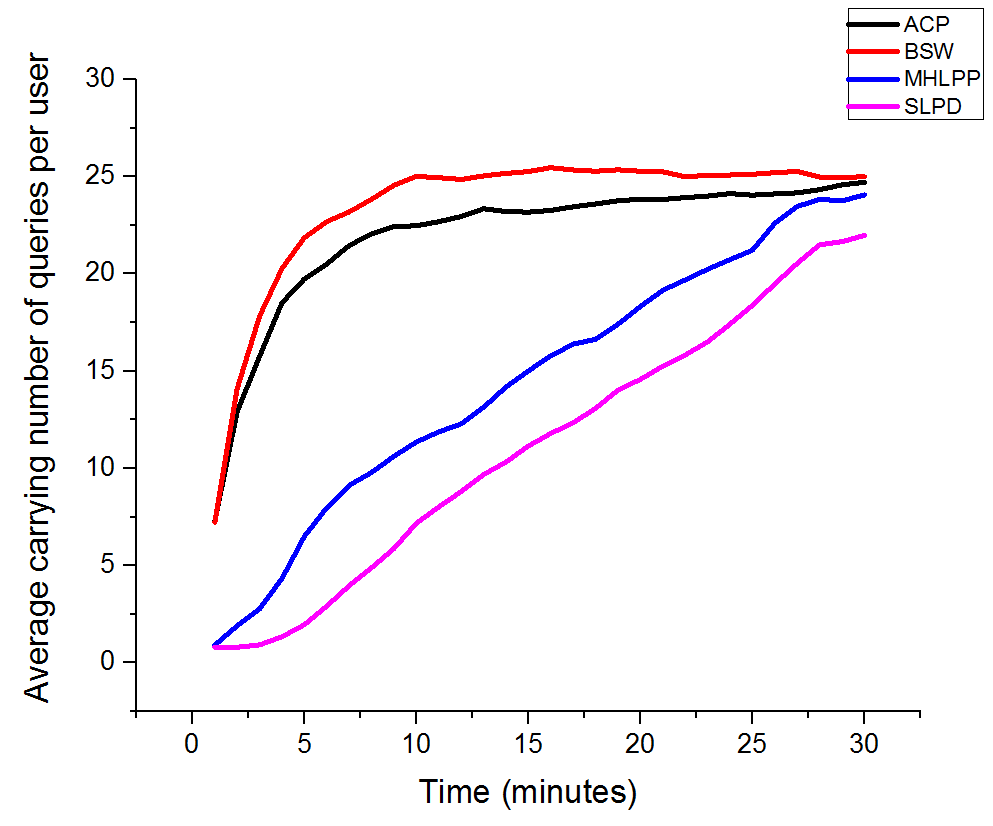
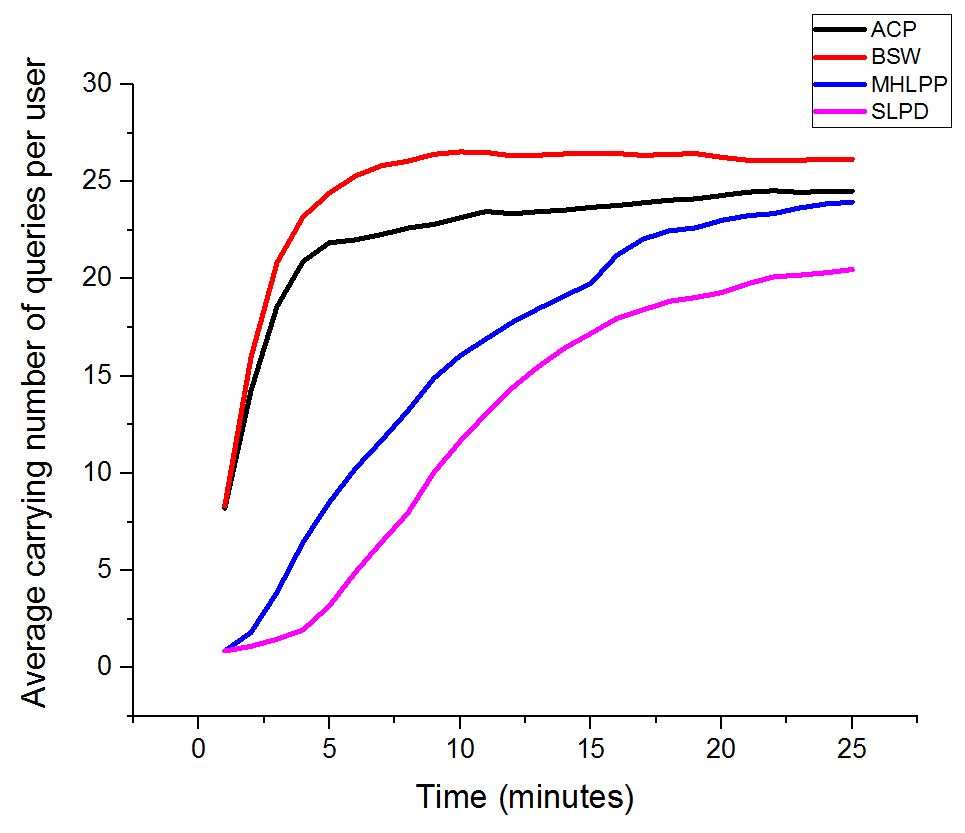


Figure 4.19 average query buffer needed at 20 minutes

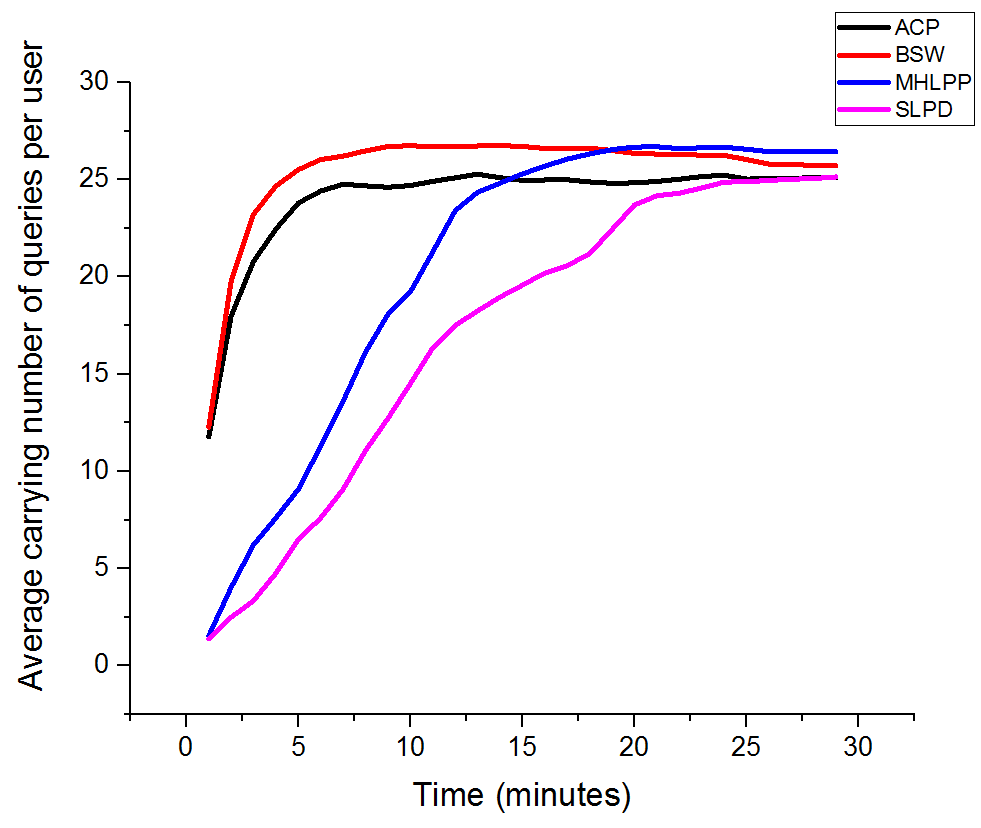
The Figure 4.20 shows the average number of queries which are carrying by users when the communication radius is 10, 50 and 90 meters. The curves of the ACP and the BSW rise sharply at the beginning and then become flat, while those of the MHLPP and the SLPD rise smoothly and continuously.



1. communication radius is equal to 10 meters



1. communication radius is equal to 30 meters



1. communication radius is equal to 90 meters

Figure 4.20 average number of carried queries per user

### Distributing appointment cards

Exchanging appointment cards is a feature of the ACP, which imports burden into the network. We count the number of exchanging appointment cards per minute to evaluate the extra cost of the ACP.

In Figure 4.22, we count the total number of exchanging ACs in the whole network. For example, if a user Alice encounters another user Bob, the total of exchanging ACs processes increases by one when Alice exchanges any ACs to Bob. We count the number of those exchanging processes occur per minute. As shown in the figure, the exchanging processes do not occur frequently, but about 2 times per minutes. Since the size of an appointment card is a small, it does not cost the network many resources. At the same time, users can get many appointment cards to help them send queries, as shown in Figure 4.21. The number of ready ACs per user is raising smoothly and steadily.

he picks an AC from his *ready-list*. He gives preference to the AC that has the earliest time-out. As a result, he sorts ACs in the *ready-list* in ascending order based on their remaining time and selects the first AC (e.g. *ac*) to send his query. That means the query is sent under the name of *ac*’s creator (i.e., *Cid*), and the location of the query is also *ac*’ location (i.e., *Clct*). Besides, *ac*’ *Capt* is also included in the query. The query can be transmitted to LBSPs by an arbitrary DTN routing protocol (e.g., Spray and Wait). When a LBSP receives the query, it provides services based on ac’s location and reply to ac’s creator (ac’s *Cid*). Users can also use a general DTN routing protocol to deliver the reply.

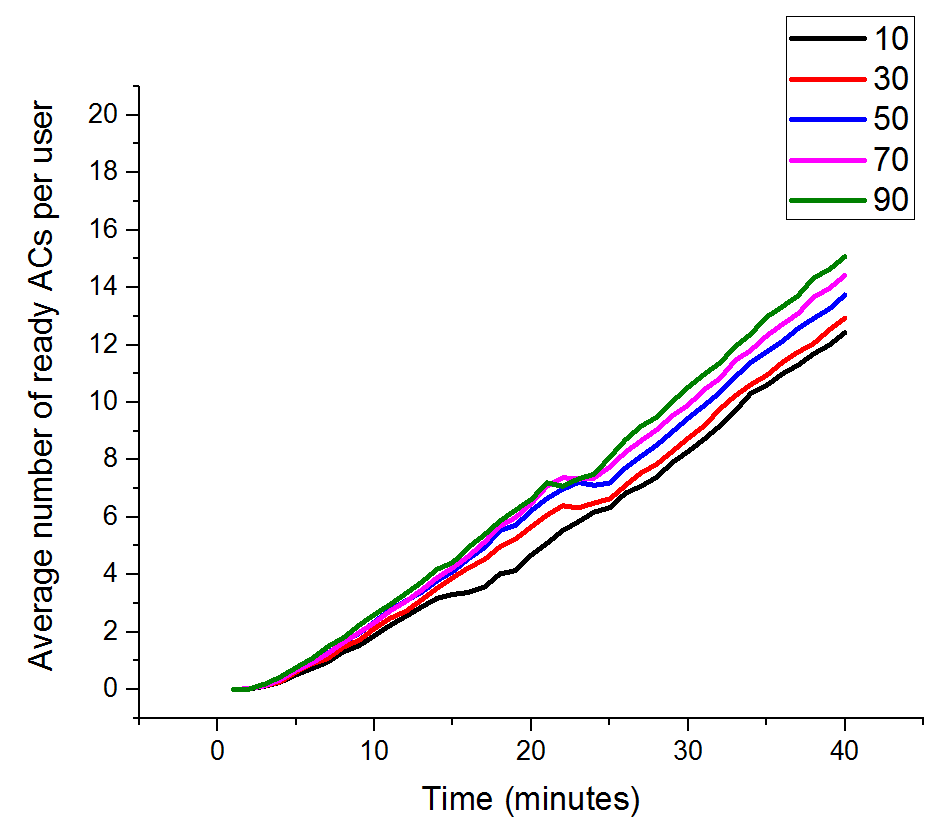


Figure 4.21 Average number of ready ACs per user

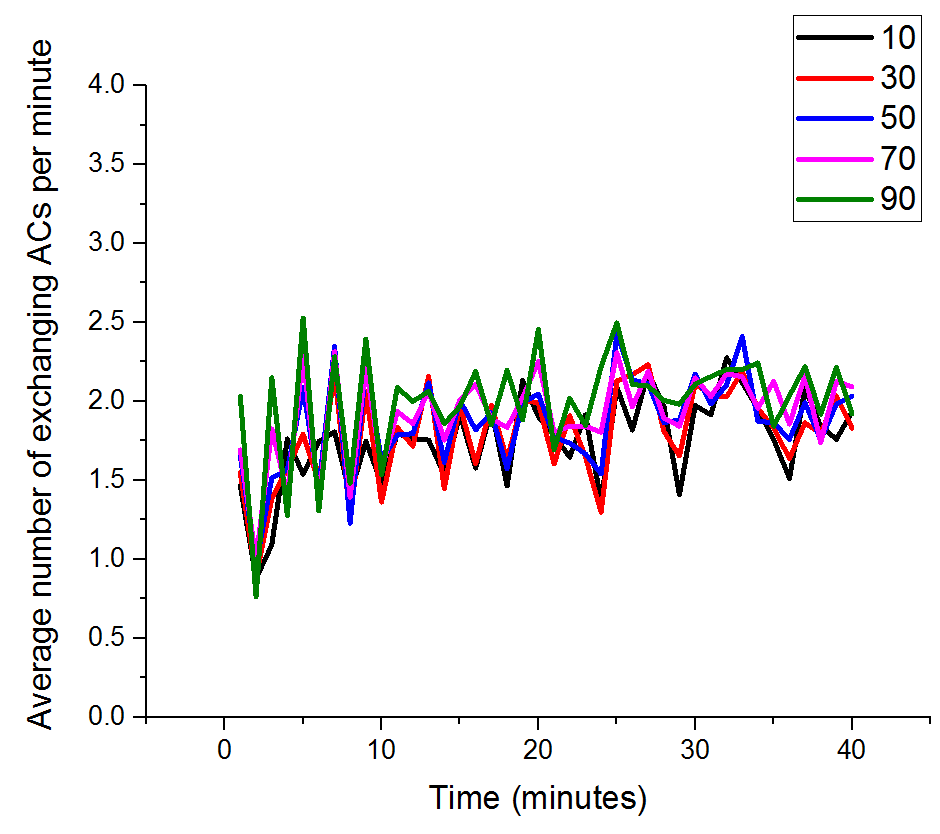


Figure 4.22 Average number of exchanging ACs per minute

# OMSN routing protocol simulator

In this chapter, we introduce two major algorithms used in our new simulator. The vertex reduce algorithm is used for simplifying the input of the all-pair shortest path algorithm, so that the simulator can initial quickly. The finding nodes algorithm is used to find users which are in the communication radius of a given user. Although a similar algorithm is also used in the ONE simulator, their implementation is not strict. That may result in wrong results. We analysis and implement the algorithm more systematically and reliably.

## Vertex reduce algorithm

Movement of network nodes is a significant factor for the performance of DTNs. To evaluate the performance of different DTN protocols, researchers presented different kinds of movement models, like in [23]. Since shortest path algorithm is an essential and time-cost part of creating many movement models, it is an important topic to optimize the time-cost for simulators calculating the all-pairs shortest path problem in the entire map.

The best known non-negative edge weight undirected map all-pairs shortest path algorithm [24] has a complexity , which is still expensive when *n* is huge. For most real-world city map, there are thousands of points in a single square kilometer. Since a tens square kilometers map is reasonable for a DTN protocol evaluation, a simulator must cut down the time-cost of the calculation of the all-pairs shortest path.

Real-world map developers often use short straight lines to present curves, like in [25], so that a several-meters curves may contain tens of points. It is obviously that we do not need to calculate shortest path for every one of these points. In this paper, we present an Vertex Reduce Algorithm (VRA) to reduce the number of points before the process of the shortest path algorithm. The basic idea is that VRA removes all points whose degrees are less than 3 from the map, while keeps the result of all-pairs shortest path algorithm correct.

The rest of this paper is organized as follows: 5.1.1 presents some basic lemmas and definitions. The process of VRA is described in 5.1.2 and 5.1.3.

### Ignorable vertex and reserved vertex

Vertices in the graph are considered as ignorable and reserved ones. In each iteration of VRA, we remove the ignorable vertex from the graph, while the reserved vertices compose a new graph.

Definition: the vertex whose degree is larger than 2 is the reserved vertex (R).

Definition: the vertex whose degree is smaller than or equal to 2 is called the ignorable vertex (G).

Definition: if two reserved vertices (e.g. ) are connected by a sequence of ignorable vertices (e.g. ), then the sequence from to (i.e., ) is a line-segment (LS).

A reserved vertex can belong to different line-segments, while an ignorable vertex belongs to a unique line-segment. Both the reserved vertex and the ignorable vertex are called vertices (V)

Definition: the shortest route between two vertices inside a line segment is called the inner shortest path (SPI), Let denote the inner shortest path between two vertices (i.e. ) inside a line segment.

Lemma1: If two vertices (i.e. , ; ) are in the same line-segment, the shortest path between them is

Since an all-pairs SPI has a complexity equal to or smaller than and the *n* here is much fewer than the number of points on the entire map, the time-cost for SPI is ignorable comparing to the time-cost of the entire map all-pairs shortest path calculation.

Lemma2: We assume that there are two different line-segments (i.e., and ). We pick a vertex from and a vertex from , then the shortest path between and is

where . Here and are in the same line-segment, so we can use lemma1 to calculate , so does . The is the only part we need to calculate using all-pair shortest path algorithms. We should notice that and could be the same vertex, but it does not make any difference to the lemma.

### Vertex reducing

VRA iteratively removes ignorable vertices from the graph until only reserved vertices remained. In each iteration, we remove all ignorable vertices from the graph but keep the edges. If there are more than one routes between a pair of reserved, we keep the shortest one and remove others.

#### Remove ignorable vertices

Since the degree of ignorable vertices is no more than 2, an ignorable vertex has only 0, 1 or 2 neighbours. In the case of 0 or 1 neighbour, we simply delete the ignorable vertex; if it has two neighbours, we connect its two neighbours before it is removed, as shown in Figure 5.1. When we remove an ignorable vertex, the weight of the line which connects its two neighbours is equal to the sum of its recent two lines’ weights (i.e., ). After we remove all ignorable vertices in a line-segment, the two reserved vertices at the ends of the line-segment are connected with a line directly, whose weight is the sum of all the intermediate ones’ (i.e., ).



Figure 5.1 Remove ignorable vertices

#### Tidy reserved vertices connections

After we remove all ignorable vertices, all the reserved vertices are connected directly. However, it is possible that there are several connections between a pair of reserved vertices. The shortest route between a pair of neighbour vertices makes other longer ones redundant obviously, so that we remove all routes except the shortest one, as shown in Figure 5.2. We assume that is the shortest line in their three connections, then and are removed.



Figure 5.2 Tidy reserved connections

#### Iterations

The whole algorithm is shown in Algorithm 5‑1. The input of the algorithm is the original entire map, which is called . In the *i*th iteration, VRA makes a copy of as . If , the is the output of the previous (i.e., ) iteration. VRA removes all ignorable vertices in the as described in 5.1.2.1 and remove all unnecessary routes between reserved vertices as described in 5.1.2.2. In other words, is smaller and smaller as *i* increases, because we always remove ignorable vertices from them. If is equal to , which means that the *i*th iteration makes no modification on the graph, the algorithm ends.

### Vertex assembling

Algorithm 5‑1

**function** reduce(*Graphi*)

Copy *Graphi* to *Graphi*+1

**for** every ignorable vertex *Gu* in *Graphi*+1 **do** remove(*Gu*) (see 5.2.1)

**for** every pair of reserved vertices **do** remove redundant connections (see 5.2.2)

**end function**

**function** VRA(*Graph*0)

i=0

**while** i=0 or there are any differences between *Graph*i-1 and *Graph*i

**do**

**call** reduce(*Graph*i)

i=i+1

**end**

**end function**

We assume that the vertex reducing process stops at the *i*th iteration. Then the is the input of an all-pairs shortest path algorithm. Let denote the shortest path from to in . We can infer the all-pairs shortest path of based on with lamme1 and lamme2. The algorithm ends when we get the result of . The algorithm is shown in Algorithm 5‑2.

We start from the and assemble all intermediate graphs, until we get the result of . For any intermediate , we have already get the all-pairs shortest path result of its reserved vertices in the previous iteration (i.e. ). Then we calculate the shortest path from any ignorable vertices to all others. When the two vertices are in the same line-segment, we use lemma1 to calculate their shortest path. The lemma 1 includes 4 parts: , , and . Since we already get the result of in the previous iteration, we just need to calculate 3 parts, which is easy and has a small scale. If the two vertices are in different line-segments, we use lemma2 to calculate the shortest path. We also already get all the parts from the previous iteration so that we just need to deal with their parts.

Algorithm 5‑2

function

for all ignorable vertices in as

for all other vertices in as

if and are in the same line-segment then

Use lemma 1 to calculate their shortest path.

else

Use lemma 2 to calculate their shortest path.

endif

end

end

end function

function

while i > 0 do

call

end

end function

## Finding nodes in a range

We assume that there are nodes in a planar map (size: ). Those nodes keep moving on the map slower than a speed threshold (). If the distance between two nodes and is smaller than a radius threshold (), they are connected. We want to calculate all-pairs connections for all nodes on the map periodically.

If we calculate these connections directly, the complexity is . We assume that and nodes are distributed on the entire map symmetrically, then the expected number of nodes which has a connection with a node () is . If we traverse all other nodes () for , it is a waste of time.

We propose an Nodes in the Range Algorithm (NRA) to optimize the calculation of all-pairs connections. we draw grids whose size is on the map. when we calculate connections of the node , only nodes in surrounding grids will be traversed.

### Static nodes

If nodes on the map are still, the radius threshold is the only factor of the selection of the grid size (). The length of the side of a grid should be equal to , as shown in Figure 5.3. For any node in the dark grid, all nodes that the distance between them is equal to or shorter than should be in the grey or dark grid. Therefore, if we want to get all connected nodes to node , we just simply need to traverse all nodes in the grey and dark grids instead of all nodes in the entire map. In this case, the expected number of nodes which we need to traverse is when we check connections for a node . If we calculate connections for all nodes, we simply need to get the distances of about pairs of nodes, instead of pairs.



Figure 5.3 grid size for still nodes

### Mobile nodes

The problem is a bit more complicated if nodes are moving. Assuming that we want to calculate all-pairs connections at , first, we draw grids at time , so that we know which grid is a node in at . Instead of , we want to get connections at . In this case, the length might not equal to the radius threshold .

Suppose that we have two nodes and , and their distance is at . Since the maximum speed of nodes is , their distance at satisfies

where . When , these two nodes are connected at , then must satisfy . If they are in the same or adjacent grids at , as shown in Figure 5.4. We have , so that : . In other words, there is no grid between two nodes at .



Figure 5.4 length of grid

### Complexities

The complexity of drawing grids is , while that of checking connections is , so the checking part is much more expensive than the drawing part. Since , the checking part achieve its best performance when , then its complexity is . That is equal to the complexity when nodes are still.

**REFERENCES**:

1. K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," in Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 27-34, 2003.
2. Pietiläinen, A.K. "Opportunistic Mobile Social Networks at Work". Ph.D. Thesis, Université Pierre et Marie Curie, Paris, 2010; <http://www.thlab.net/~apietila/pubs/thesis.pdf>.
3. Ivan Stojmenović, "Handbook of wireless networks and mobile computing,", 2002.
4. J. Schiller, A. Voisard, "Location-Based Services," Morgan Kaufmann, 2004.
5. J. Krumm, "A survey of computational location privacy," Personal and Ubiquitous Computing, Volume 13, Issue 6, pp 391–399, August 2009.
6. E. P. C. Jones and P. A. S. Ward, “Routing Strategies for Delay-Tolerant Networks”, Computer Communications, 2008.
7. U. Lee, S. Y. Oh, K. W. Lee, M. Gerla, “Relay Cast: scalable Multicast Routing in Delay Tolerant Networks”, IEEE, 2008.
8. A. Vahdat and D. Becker. “Epidemic Routing for Partially-Connected Ad Hoc Networks”, Technical Report CS-2000-06, Apr 2000.
9. A. Lindgren, A. Doria, O. Schelen, “Probabilistic Routing in Intermittently Connected Networks”, ACM SIGMOBILE Mobile Computing and Communications Review, Vol. 7, Issue 3, pp. 19-20, July 2003.
10. J. Shen, S. Moh and I. Chung, “Routing Protocols in Delay Tolerant Networks: A comparative Survey”, 23rd International Conference on Circuits/Systems, Computers and Communications, (ITC-CSCC 2008)
11. Y. Liao, K. Tan, Z. Zhang and L. Gao, “Estimation based Erasure Coding routing in Delay Tolerant Networks” Proc. of IWCMC, July 2006.
12. L. J. Chen, C. H. Yu, T. Sun, Y. C. Chen and H. H. Chu, “A Hybrid Routing Approach for Opportunistic Networks”, Proc. of ACM SIGCOMM on Challenged Networks, 2006.
13. M. M. Zanjireh and H. Larijani, "A Survey on Centralised and Distributed Clustering Routing Algorithms for WSNs," IEEE 81st Vehicular Technology Conference, 2015.
14. M. Gruteser, D. Grunwald, “Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking,” in Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, pp. 31-42, 2003.
15. S. Zakhary, M. Radenkovic, "Utilizing social links for location privacy in opportunistic delay-tolerant networks," in Proceedings of International Conference on Communications (ICC), 2012.
16. S. Zakhary, M. Radenkovic, A. Benslimane, "The quest for location-privacy in opportunistic mobile social networks," in Proceedings of IWCMC, 2013.
17. S. Zakhary, M. Radenkovic, A. Benslimane, "Efficient Location Privacy-Aware Forwarding in Opportunistic Mobile Networks," IEEE Trans. on Vehicular Technology, Vol. 63, No. 2, pp. 893 – 906, 2014.
18. E. Daly, M. Haahr, “Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs,” in Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2007.
19. M. Duckham, “Moving forward: Location privacy and location awareness,” in Proceedings of the 3rd ACM International Workshop on Security and Privacy in GIS and LBS, pp. 1-3, 2010.
20. T. Hashem, L. Kulik, “Safeguarding location privacy in wireless ad-hoc networks,” in Proceedings of UbiComp, pp. 372-390, 2007.
21. L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, Vol. 10, No. 5, pp. 557-570, 2002.
22. M. Mano, Y. Ishikawa, “Anonymizing User Location and Profile Information for Privacy-aware Mobile Services,” in Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pp 68-75, 2010.
23. T. Wang, L. Liu, “Privacy-aware mobile services over road networks,” in Proceedings of the VLDB Endowment, pp. 1042-1053, Vol. 2, 2009.
24. R. Lu, X. Lin, X. Shen, “SPRING: A Social-based Privacy-preserving Packet Forwarding Protocol for Vehicular Delay Tolerant Networks,” in Proceedings of IEEE INFOCOM, 2010.
25. J. Zhang, C. Chow, “REAL: A Reciprocal Protocol for Location Privacy in Wireless Sensor Networks,” IEEE Trans. on Dependable and Secure Computing, Vol. 12, No. 4, pp. 458 – 471, 2015.
26. P. Hui, J. Crowcroft, “How Small Labels create Big Improvements,” in Proceedings of PERCOM Workshop, pp. 65–70, 2007.
27. Diffie, W., Hellman, M., "New directions in cryptography," IEEE Trans. on Information Theory, Vol. 22, No. 6, pp. 644-654, 1976.
28. P. Jacquet, P. Muhlethaler, T. Clausen, “Optimized link state routing protocol for ad hoc networks,” in Multi Topic Conference, 2001.
29. R. Xiang, J. Neville, M. Rogati, “Modeling Relationship Strength in Online Social Networks,” in Proceedings of the 19th International Conference on World Wide Web, pp 981-990, April 26 - 30, 2010.
30. T. Spyropoulos, K. Psounis, C. S. Raghavendra, "Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks," in Proceedings of the ACM SIGCOMM Workshop on Delay-tolerant Networking, pp. 252-259, 2005.
31. F. Ekman, A. Keränen, J. Karvo, J. Ott, “Working Day Movement Model”, MobiHoc 2008.
32. A. Moffat, T. Takaoka, “all pairs shortest path algorithm with expected running time O(n2logn),” Foundations of Computer Science, 1985

1. <https://www.openstreetmap.org/>
2. Ari Keränen, Jörg Ott, Teemu Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in Proceedings of SIMUTools 2009, Rome, Italy.