City Research Online

City, University of London Institutional Repository

Citation: Arunkumar, S. (2016). Preserving Privacy in Mobile Environments. (Unpublished Doctoral thesis, City, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/19299/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

CITY UNIVERSITY
LONDON

Doctoral Thesis

Preserving Privacy in Mobile Environments

Saritha Arunkumar

A thesis submitted to

City University London, School of Engineering and Mathematical Sciences

In Fulfilment of the Requirements for the Degree

Doctor of Philosophy

in Information Engineering

November 21, 2016

Contents

List of Figures

List of Tables

List of Listings

Declaration of Authorship

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning. I hereby grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Acknowledgements

This thesis arose after years of research that has been done since I came to City University London. First and foremost, I would like to express my gratitude to my supervisor Professor Muttukrishnan Rajarajan, for his supervision, advice and guidance from the starting of my research. With his extraordinary experience and admirable insights, he has taught me about science, life, and encouragement. It is my pleasure to work with you.

I would also like to acknowledge the research collaboration work done through International Technology Alliance (ITA) project with IBM Research USA. Thanks for the invaluable advice and patience. I am grateful to my colleagues and everyone for their support and help on my research work.

I owe my final thanks to my beloved parents and my husband for their unconditional support, love, and trust. This thesis is dedicated to them.

Abstract

Technology is improving day-by-day and so is the usage of mobile devices. Every activity that would involve manual and paper transactions can now be completed in seconds using your fingertips. On one hand, life has become fairly convenient with the help of mobile devices, whereas on the other hand privacy of the data and the transactions occurring in the process have been under continuous threat. Mobile devices connect to a number of service providers for various reasons. These could include downloading data, online purchasing or could be just used to browse information which may be irrelevant at a later point. Access to critical and sensitive information may be available at a number of places. In case of a mobile device, the information may be available with the service provider. Service Provider could be in the form of any web portal. In all such scenarios, passing the information or data from the service provider into the mobile device is a major challenge, as the data/information cannot be sent in plain text format. The confidentiality and integrity of the data needs to be protected and hence, the service provider must convert the data into an encrypted format before passing it onto the mobile device, to prevent risks from sniffing and unauthorized disclosure of data. Preserving the location of the individual user of any mobile device has also been the concern for a number of researchers.

Mobile devices have become an important tool in modern communication. Mobile and other handheld devices such as ipads and tablets have over taken laptops and desktops and hence there has been an increasing research interest in this area in recent years. This includes improving the quality of communication and the overall end-to-end data security in day-to-day transactions. Mobile devices continuously connect to different service providers for day-to-day needs such as online purchases, online banking and endless surfing for information. In addition to this devices could be connecting to the service providers to receive or send sensitive information. At the Service Provider end, the data would be stored with the provider and Service Provider would only hand over the data if it confirms that the person requested it is authorized to receive the information. The exchange of data from one end of the network to the other is a major challenge due to malicious intruder mishandling of the data. Hence the confidentiality and integrity of the data needs to be protected either by transforming the sensitive information into a non-readable format or by converting into a cipher text.

Privacy has been an open problem for research as more and more information is getting leaked on a day-to-day basis. Through this thesis, I have tried to address a number of areas within the privacy realm where information and data access and sharing is a key concern along side the key aspect of location privacy. I have also tried to address the problems in the space of access control wherein I have proposed policy based languages and extensions for ensuring appropriate access control methodologies. The main goal and focus in this work has been to enforce the importance of location privacy in mobile environments and to propose solutions that resolve the problems of where and when to enforce location security. Another key goal of this work has been to create new access control and trust based solutions to ensure the right level of access to the right receiver of information. Through my research, I have explored the various privacy related attacks and suggested appropriate countermeasures for the same. In addition to proposing and showcasing solutions using policy languages for access control, I have also introduced geospatial access control solutions to ensure that the right user is accessing or requesting for the right information from the right location. This helps the appropriate and the right use of the information by the right resource. Through my thesis I have also given equal importance to the trust aspects of sharing information. I have created new trust assessment models to show how fused information can be handled and how can trust be imposed on the information provider and the information itself.

The main contribution of this thesis is to address the problems around protecting the data and individual's privacy and to propose solutions to mitigate these issues using new and novel techniques. They can be detailed as the following:

In privacy, there is always a privacy versus utility tradeoff and in order to make use of utility, trust in the location is essential. Through this research I have developed i) novel attestation models and access control methodologies including Privacy Preferences Platform (P3P) extensions, ii) Extensible Access Control Markup Language (XACML) extensions and iii) Geospatial access control through GeoXACML. iv)I have created new methodologies to enforce location privacy and shown where best to enforce privacy. v)I have also shown that global attestation is very crucial for privacy and needs accurate methods in place to attest user's location information for access. vi) Fusing of location information is very crucial as there could be a number of similar or conflicting information produced about a common source and it is very important to assess and evaluate the trust level in the information. I have proposed, developed and implemented a new trust

assessment framework. This framework looks at the incoming information and passes it on to the rule engine in the framework to make some inferences and then the trust assessment module computes the trust score based on forward chaining or background chaining scheme. The framework is used to evaluate the trust on the fused information in a streaming setup. vii) I have created new solutions to look at the similarity profiles and create identity enforcement through profiling. I have shown methods of anonymisation for location privacy and identity privacy.

Abbreviations

AA	Attribute Authority
AP	Access Points
APPLAUS	A Privacy-Preserving Location proof Updating System
ARL	Army Research Laboratory
BRS	Beta Reputation System
CDA	Cheating Detection Authority
CDR	Call Detail Records
CE	Controlled English
CLDC	Connected Limited Device Configuration
CNL	Controlled Natural Language
CR	Cloaked Region
DAML	DARPA Agent Markup Language
DoS	Demial of Service
DRM	Digital Rights Management
GeoXACML	Geospatial eXtensible Access Markup Language
GP	General Practitioner
GPS	Global Positioning System
GSMA	Global System for Mobile communication Association
ID	Identifier
IED	Improvised Explosive Device
IMEI	International Mobile Equipment Identity number
IMPI	IP Multimedia Private Identity
IMSI	International Mobile Subscriber Identity
ITA	International Technology Alliance
J2EE	Java 2 platform Enterprise Edition

J2ME	Java 2 Micro Edition
JDK	Java Development Kit
KAoS	Knowledgeable Agent-Oriented System
LA	Location Anonymiser
LBSA	Location Based Social Application
LSP	Location Service Provider
MAC	Media Access Control
MIDP	Mobile Information Device Profile
MILC	Mobile Instant Locator with Chatting
MPS	Meta Policy Server
OASIS	Organization for the Advancement of Structured Information Standards
OGC	Open Geospatial Consortium
OIL	Ontology Interchange Language
OODA	Observe Orient Decide Act
OWL	Web Ontology Language
P3P	Privacy Preferences Platform
PAP	Policy Adminstration Point
PDA	Personal Digital Assistant
PDP	Policy Decision Point
PEP	Policy Enforcement Point
SL	Subjective Logic
SOAP	Simple Object Access Protocol
SSN	Social Security Number
SSO	Single Sign-On
STAMP	Spatial-Temporal provenance Assurance with Mutual Proofs
SWRL	Semantic Web Rule Language
SYNCOIN	Synthetic Counter Insurgency
UDDI	Universal Description ,Discovery and Integration
UK MoD	UK Ministry of Defence
WSDL	Web Services Description Language
XACML	eXtensible Access Markup Language
XML	eXensible Markup Language

Publications

The results of the research described in this thesis have been published in the following papers and journals:

1. S. Arunkumar, A Raghavendra, D Weerasinghe, D Patel, M Rajarajan, "Policy extension for Data Access Control", *In Proceedings of the 6th IEEE Workshop on Secure Network Protocols (NPSec)*, pp 55 -60, October 2010.

2. S.Arunkumar, M Rajarajan, "Healthcare Data Access Control using XACML for Handheld Devices", *In Proceedings of Developments in E-systems Engineering (DESE) 2010*, pp 35 - 38, September 2010.

3. S.Arunkumar, M. Srivatsa, D. Braines, M Sensoy, "Assessing Trust over Uncertain Rules and Streaming Data", *In Proceedings of the 16th IEEE International Conference on Information Fusion (FUSION)*, pp. 922-929, July 2013.

4. S.Arunkumar, M. Srivatsa, M Rajarajan, "Location Security - Where to Enforce?", *IEEE Military Communications Conference (MILCOM)*, pp. 1651-1656, October 2014.

5. S.Arunkumar, M. Srivatsa, M Rajarajan, M Sensoy, B Soyluoglu, "GeoSpatial Access Control for Mobile Devices", *International conference of IEEE Region10 on Internet of Things*, pp. 86-89, May 2015.

6. S.Arunkumar, M. Srivatsa, M Rajarajan, M Sensoy, "Global Attestation of location in mobile devices", *Military Communications Conference (MILCOM)*, pp. 1612-1617, October 2015.

7. S.Arunkumar, M. Srivatsa, M Rajarajan, "A review paper on preserving privacy in mobile environments", *Elsevier Journal of Network and Computer Applications* 53, 74-90, July 2015.

8. S.Arunkumar, M. Srivatsa, M Rajarajan, M Sensoy, "Reasoning with streamed uncertain information from unreliable sources", *Elsevier Journal on Expert Systems with Applications*, 42(22), 8381-8392, May 2015.

Chapter 1

Introduction

1.1 Research area and questions

Ubiquitous environments are environments where one has access to devices and computers with internet wherever and whenever they need it. They are the next generation of environments, and have been discussed in a number of research papers [2], [3] and [4]. Mobile devices, Personal Digital Assistants (PDAs), and other electronic devices are used for various transactions in the ubiquitous environment and interact with each other. Mobile devices store various sensitive data in a ubiquitous environment and having a controlled access to these data becomes crucial for a secure ubiquitous environment. There are applications that can be installed on a mobile device for everything and anything and devices connect to these application providers for various reasons such as downloading data, online purchasing or could be just used to browse information which may be irrelevant at a later stage. The importance of maintaining the privacy of the information requester and the information, is critical as all the transactions are happening online and this could motivate the malicious users to perform malicious activities. In case of a mobile device, the information may be available with the application provider of the application, which the device is accessing and this could be in the form of any web portal, database or any other equivalent form. The confidentiality and integrity of the data needs to be protected and hence, the application provider must convert the data into an encrypted format before passing it onto the mobile device, to prevent risks from sniffing and unauthorised disclosure of data. Individuals trying to access sensitive information through mobile devices are risking their personal data, identity and location

at all times. Preserving the location of the individual user of any mobile device has been the concern for a number of researchers and people working in the area [3], [5] and [6]. Privacy related efforts have been made in the past [3]. Research has been carried out around privacy awareness systems that allow certain privileges to data collectors [7]. Karyda and Gritzalis in [4] listed some of the challenges in this area and the research directions for the future. In general, a malicious user is assumed to be very clever in manipulating communications over the open network and his manipulation techniques are unpredictable because they are unspecified. Also because malicious user can represent a coalition of bad guys, he may simultaneously control a number of network nodes which are geographically far apart. In anticipation of such a powerful adversary over such a vulnerable environment, Dolev and Yao in [8] proposed a threat model which has been widely accepted as the standard threat model for cryptographic methods.

In addition to the day-to-day needs of the devices to connect to the application providers, devices could be connecting to the Application providers to receive or send sensitive information. At the Application provider end, the data would be stored with the provider and the Application provider would only hand over the data if it confirms that the person requesting is authorised to receive the information. The exchange of data from one end of the network to the other is a major challenge due to malicious intruder mishandling of the data. Hence the confidentiality and integrity of the data needs to be protected either by transforming the sensitive information into a non-readable format or by converting into a cipher text. Some of the access control methodologies like the standard role based access control methods existed, but the thnigs that were used for standard desktop solutions were not the best suitable for mobile devices as well. Mobile devices have limited bandwidth and limited computational power. Under these restrictions, the problems of access control and information to the right resource for right reasons needs to be addressed. Existing solutions at the time of the research did not cater to handle the limitations on the handheld devices. The problems with the preserving privacy are applicable to all sectors including healthcare, government, communications and so on. Some of the key areas and questions still unanswered in the area of preserving privacy are around preserving the location information when using location based services. How can one preserve the identity and location information and still get access to the information requested for ? Another open question is around the authentictity of the location information provided, when information is released based

on the accuracy of the location. To add to this, how can one trust the source or provider of the information ? These questions have been barely looked at in the research area and hence I have addressed these in detail and provided solution to it.

Here is an introduction to a very practical and realistic problem with Healthcare services today in order to set the scene. From the Figure 1.1, one can see that there are 2 main players in the scenario one is the permanent General Practitioner (GP) and the other is temporary practitioner in the absence of the permanent doctor. Assuming that the trust is established between the mobile device, the Mobile Identity Provider (IDP) and the Health service repository, the next big question is the transfer of the information related to health. The information needs to be passed on from the health repository on to the GP's handheld device. This is very crucial and is a big security issue as the data that is to be transferred is highly sensitive information and hence needs to be received by the right person for the right reasons. In privacy there is always a privacy versus utility tradeoff. To make use of some kind of information, there has to be some level of trust in the location and hence attestation of location information needs to be performed.

Access control of the data and information in the mobile environments becomes a very big area of research. This thesis highlights the solution implemented to achieve the access control of data and information in mobile environments.

Trust was assumed to exist in the above scenario. However it is important to understand how trust can be assessed when information is collected from various sources and this highlights another challenging research problem. When there is streaming information that comes in with additional location information, fusing those location information and deriving the right level of trust if crucial.

Another use case to show the use of location privacy in a military environment is presented here. By showing a military scenario, it is indicative that this solution is applicable to mobile environments in many sectors. The scenario is as follows: Assume that one has a number of reports that comes in from various sources reporting of an Improvised explosive device (IED) explosion in certain location l at time t. How can one ensure the level of trust on the information about the event occurring in that location? This needs some kind of trust assessment framework to assess the level of trust on fusing location information.

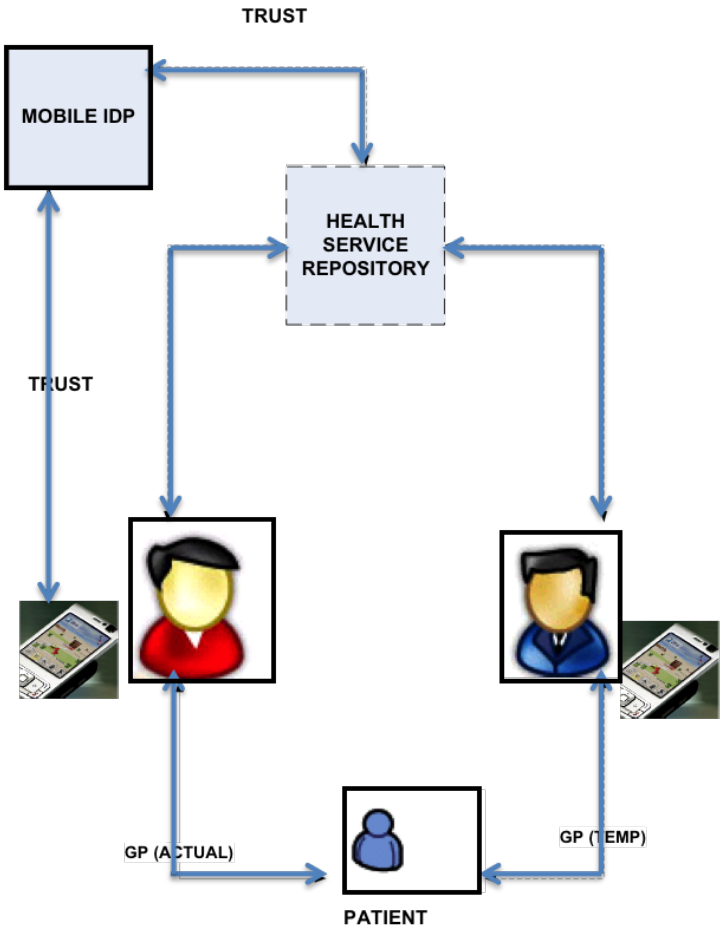

FIGURE 1.1: HealthCare Privacy Scenario.

Here is another use case in a military scenario that shows the importance of addressing the "where to enforce location privacy" in mobile environments. Figure 1.2 shows a tactical network where you can see device and the core of the network and the question I am trying to address is "where to enforce location privacy?"

Mobile environments are always prone to various security vulnerabilities. A number of papers have been written to highlight the various threats and problems due to the large amount of transactions occurring in the mobile environments [9], [10]. A very common attack on the mobile environment is the man-in-the-middle attack. Every bit of data that comes into the mobile device and goes out of the mobile device can be assumed to be sniffed by a malicious user. The information can be assumed to be sniffed by the man-in-the-middle and manipulated in order to retrieve the sensitive information. Protecting the information that is being exchanged between the mobile devices is a

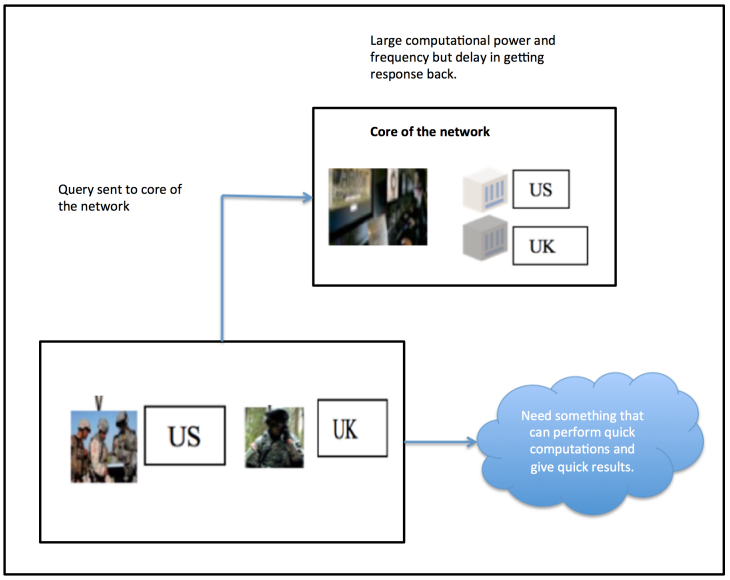

FIGURE 1.2: A tactical network scenario – highlighting the need for some capability other than at the core of the network.

major challenge. The attack discussed above covers a number of attacks including man-in-the-middle, sniffing and privacy related attacks. Another attack that is described by some of the researchers is based on the cross service attack on the mobile devices. Mulliner *et al.* have described the details of the cross service attack in [11]. Cross service attacks can occur while browsing from mobile handset sitting in a shop with wireless connectivity. The malicious user would be monitoring the new connections to the wireless network and using a known exploit and he gains access to the phone. [11] describes in detail the proof-of-concept to show the attack and also talks about the way in which the vulnerability can be exploited. With the increasing availability of mobile devices, there is a growing demand for location-based applications. In response to such a user demand, various location based services have been emerging recently [12], [13].

In the mobile environment, it is quite common to have a man-in-the-middle trying to sniff at the information being passed between the mobile device and the application providers [11]. Therefore it is crucial to have data access control mechanisms in place. It would be interesting to highlight the importance of European data protection guidelines that have recently undergone revisions to include the privacy of individuals data and personally identifiable information. Some of the notable changes include explicit consent from the user when data is being handled. More transparency about how the data is handled is another important change that has been added. The reform also includes the mandate

for complete accountability and responsibility of the service provider when personal data is being processed [14].

1.2 Research objectives

Information and data from/to mobile devices can be misused and mishandled by intruders or malicious users. In order to make sure that the right individual is accessing the data, it is very important to have appropriate access control rules and policies in place. Hence, it was very important for this research to consider some of the key methodologies that could be used to ensure access control of data and information.

- Consider the example from Figure 1.1 that shows a typical scenario in a health service. Providing appropriate access control for the data is a very big issue in the health sector on its own and it is definitely the same issue in any other areas of infrastructure. Hence data access control is one of the main research problems today.

- In a military environment, it is very critical to preserve the location information of the individuals out in the field but it is equally important to get access to the data and information that they are requesting. As with any privacy solution, there is always a privacy Vs utility tradeoff and when in need of information there has to be a certain level of trust of the location. The objective here is to have some level of attestation in order to ensure the right levels of access controls for information.

- Figure 1.2 shows a military use case where privacy is very important and critical for all the right reasons. The core of the network has lot of computational power and bandwidth but then information needs to be received along with the location privacy maintained, it is not the best of the places to have the solution. The objective here is to come up with something else that can do some quick computations in the limited time and can provide accurate results to the requestor but without revealing the sensitive identity and location information. Maintaining location privacy in a military setup is very important for any successful operation and when certain information is very sensitive there is a tradeoff of providing the valid location which can only be verified through location attestation.

- When there are multiple locations being reported, it is very difficult to fuse these information and derive a consensus from the data. The main objective with this is to have a mechanism that could help trust decisions automatically.

- When location services are requested on the mobile devices, the location information can be faked and can be misrepresented for meeting specific requirements. Gathering evidences independently becomes important where each entity reports contacts with other entities in specific locations at specific times. Some of the reports may be not true and in order to scrutinise these, global consistency check would be required. With a global consistency check I can ensure if a particular entity is being dishonest about the location information.

  Hence, the objectives of this thesis are to propose solutions based on attestation of location, access control through policy extensions and geographical co-ordinates, trust assessment through mathematical models and location privacy through edge servers and anonymization techniques.

1.3 Research assumption

To shape the research, some assumptions were made giving some starting points and directions to the work:

- Location is a very important factor in the mobile environments.

- Small scale mobility tracesets are human to human encounters and relatively large scale datasets are many to many encounters but using taxicab datasets I have got a mixture of multiple people taking the same route and hence got more data in a larger scale both in area and time.

- By using Subjective logic, I assume that trust has beta distribution and the best model of trust is represented in the form of a triple (b,d,u) which is *belief, disbelief, uncertainty.*

- With the Geospatial access control through GeoXACML, the main goal is not to provide mandatory access control. The main criteria is that if the user satisfies the location constraint, information will be provided hence providing discretionary

access control. For mandatory access control, well known solutions like Digital Right Management (DRM) exist.

- Global attestation of location works assuming that the majority of users are honest users and have persistent IDs like Media Access Control (MAC) address and so on. Hence ID scheme is a requirement for this solution to work. It is also important to have a minority count on the collusive users.

1.4 Contributions

This research is aimed to provide a solution for preserving privacy in mobile environments and covering the aspects of how to enforce location privacy, where to enforce location privacy and how to ensure trust when location information is fused. The overall solution allows the ease of use of mobile devices with appropriate privacy controls in place to ensure that the right individual, in the right place and in a controlled environment, is accessing the data and information.

Contributions of this research work include the following:

- I have proposed and developed a solution to address the question of "How to enforce location security" in detail.

  Privacy comes with a privacy Vs utility trade-off. When a certain utility is required, it is essential to ensure a certain level of trust on the location. Through this research I have proposed a novel method using global attestation scheme and developed and tested it . Global attestation of location ensures the accuracy of the location by performing global consistency checks. I have used the EigenTrust [15] and PeerTrust [16] check methodologies to calculate global and personalised trust matrix while confirming attestation.

- I have proposed and developed new methods of access control for maintaining privacy.

  In order to address this, I introduce P3P based access control and then extend the P3P policies to ensure that the data access control is achieved. The dependencies are mainly on the P3P policy server and the design and architecture ensures that the server is configured appropriately for the level of security required.

A solution using XACML based access control has also been introduced through this research. With the use of policy decision point, policy enforcement point and policy administration point, an XACML based access control solution ensures the information and data is accessed by the right sources. I have extended XACML and developed a new GeoXACML implementation to ensure that access control is achieved along with the location verification.

- I have designed a novel trust assessment model to validate fused location information.

  When there is a set of location information that is provided through streamed information, it is difficult to derive at a consensus on the location and this could be resolved through some level of trust assessment. In this research, a new trust assessment model using subjective logic operators was designed to derive various outcomes.

- I have designed a novel solution to address the "Where to enforce location privacy" problem.

  This research describes the drawbacks with implementing the solution for location privacy at the device and the core of the network. It gives a detailed analysis of the best place to enforce location security using the edge based solution and implementing anonymization methods for the same.

1.5 Outline

The thesis is organised in 6 chapters as follows.

Chapter 2 presents an analysis of the existing approaches dealing with privacy related attacks, methods of preserving privacy in mobile environments, policy based access control, location and identity based mobile security, geospatial access control and trust in the mobile environments. The focus is mainly on the different attacks related to privacy and then the various methods used in order to preserve location, identity, the drawbacks with the existing solutions, along side data access control and trust in the overall mobile security space.

In chapter 3, I highlight the importance of access control for ensuring privacy in mobile devices and propose a mechanism for data access control using P3P policy extension and XACML. It details each of the policy languages and the process in which access control can be achieved by using P3P and XACML. The chapter further describes how geospatial access control can be achieved by creating a new implementation of GeoXACML. The chapter also covers the details about the implementation and evaluation of each of the methods.

In Chapter 4, I address the "How to enforce location privacy" problem. Through this chapter, I will highlight the privacy Vs utility trade-offs and show the importance of ensuring trust in location. I then present a new model called the global attestation model which ensures the location accuracy through global consistency checks and trust matrix scores.

Chapter 5 describes various aspects involved when location information is fused. The involvement of trust in the mobile environments is discussed in detail and it covers various use cases and system highlights. These showcase the significance of using trust assessment with opinion to ensure that the fused and conflicting information is resolved. It details the various steps in the opinion assessment framework and introduces Subjective Logic which is the method used to compute the trust. The chapter shows the detailed architecture of the trust framework with implementation and evaluation results. It shows how to ensure trust when encountered with fused and conflicting information.

Chapter 6 describes the "Where to enforce location privacy" problem and provides solution to show that location privacy can best be enforced at the edge. It shows the drawbacks in implementing location privacy on the device and at the core of the network and highlights why it is important to enforce location privacy on the edge of the network. Anonymization methods of performing location privacy on the edge of the network is also described in this chapter. The chapter covers the detailed design of the architecture used for implementation and the evaluation steps with results.

Finally, in Chapter 6, I present the conclusions. In this chapter, I highlight the contributions of the research, the benefits and implications of this and some topics for future work.

Chapter 2

Background and related work

2.1 Overview

This chapter presents an analysis of the existing works in the field of preserving privacy in mobile environments. It shows the various privacy attacks and the solutions that have been designed for preserving location privacy in mobile environment. I further highlight the drawbacks in each of the existing solutions. This chapter also includes review work on policy based access control for mobile, location and identity security, trust and geospatial access control.

2.2 Background

Here you can see definitions of the key words used throughout this thesis. These definitions will help the reader better understand the usage of the words in relation to the work in privacy.

2.2.1 Access Control

Access control is a method by which a system grants or denies the right to see certain information, do certain actions. Access control can be achieved using different mechanisms. Access control has played an important roles since the time of computer evolution. But with the use of smartphones the importance of access control has got

complicated. Computers and servers have lot of bandwidth and computational power hence can deal with lots of computations before getting access to information. In the case of mobile device, this is slightly different. The device has limited bandwidth and computational power, it has limited storage capability and hence using all these constraints access control checks have to be performed before giving access to information on the mobile device.

2.2.2 Privacy

Privacy is a mechanism where an individual or group of people can restrict providing information about them to the outside world. It is also a state where individuals do not want to share personal information with public. Privacy is very important in the digital world as information shared in the public domain can be misused.

2.2.3 Policy languages

Consumers are concerned about losing their privacy while conducting business using computers. Studies reveal that if consumers are not comfortable with the organizational procedure for handling their personal information, they take their business elsewhere [17]-[18]. With the growth of Internet usage and an increase in online business, consumers expect high levels of online privacy [19]. In fact consumers frequently mention lack of trust as one of the reasons for not purchasing from the Internet [20]. From the organization's perspective, the need to protect consumer privacy is a growing concern [21]. If organizations do not follow efficient privacy practices, consumers move away from the organizations and legal consequences arise. Protecting or not protecting the privacy of consumers can impact the business growth and revenue of the organizations. Thus, the formalization of an organization's promises into privacy policy is an essential part of the organization's customer relationship practices. Organizations express their internal privacy practices through privacy promises as statements in the privacy policies. Consumers are able to analyze the organization's commitment towards protecting consumers' privacy through these privacy policies. Unfortunately, few consumers take the time to read the complete privacy policies of all organizations with whom they interact online. Studies have shown that the privacy policies on the Internet are long, legalistic

and difficult to understand [22], [23]. Also system administrators find it difficult to maintain the internal access control information manually [24]. Researchers have attempted to solve this problem by developing privacy policy languages that help represent the human readable privacy policies and access control into machine-readable format [25]-[26]. These privacy policy languages enable software agents to understand privacy policies better and help users to make informed decisions. These languages also help systems to make decision regarding an user's access to any data in the organization. The capability of these languages depends on various features; one important feature is expressiveness of the language. As a result, understanding the expressiveness of the privacy policy languages becomes essential. Different types of languages are available to represent the human-readable policies into machine-readable format. Some languages are designed to help the organizations to express their privacy policies in machine-readable format and some of the languages help users to define their privacy preferences, which can be used to make decisions for the user. Each language has its own syntax and mechanism for implementation. There is no one standard metric available by which one would analyze these languages.

2.2.4 Security Capsule

The security capsule is a software application for mobile devices and it contains security services which are used to protect the sensitive information in the mobile device. It is a key component in the communication module of the mobile device used to interact with the identity provider and the service provider. The very existence of the security capsule in the mobile device is through the registration process. The registration process has two steps namely, registration with identity provider and registration with service provider. The first step starts with the security capsule being downloaded into the mobile device from the identity provider. The mobile user verifies the authentication of the Identity provider for transmitting the security capsule and the integrity of the downloaded security capsule. In the second step, the security capsule registers with the service provider for services. The user and service provider share a unique identification. Therefore the user registration request will be uniquely identified by the service provider. If an identity doesn't exist then the Identity provider generates an identity for the mobile user. The mobile device authenticates with the identity provider and both parties share

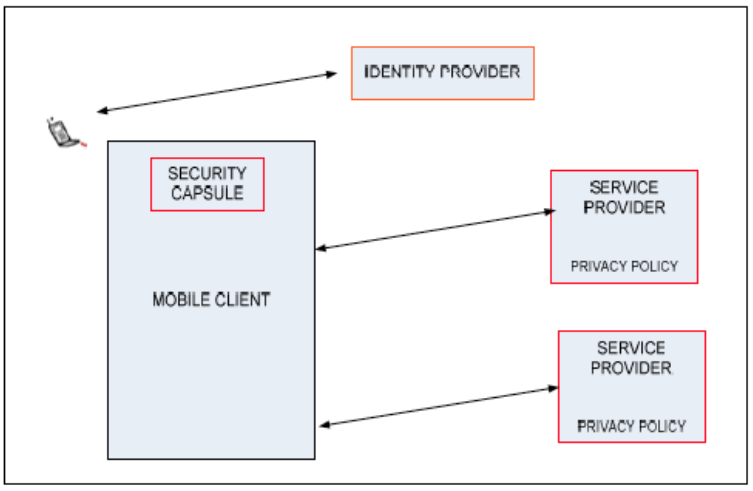

FIGURE 2.1: Security capsule within a mobile device.

a secure communication channel. The security capsule architecture and the functionality are described in detail in [27].

Figure 2.1 shows the mobile device with the security capsule obtained through registration process and the interactions between the service providers and the mobile client. The primary challenge in a security capsule is to provide controlled and appropriate data access control to the right user. This is based on the real time key that is received from the service provider. The service provider sends the data/information requested by the mobile device in an encrypted format. The real time key is used to encrypt the data and the mobile device requires this real time key to access the data. In order to receive the real time key, the mobile client needs to first provide the appropriate user preferences based on the P3P policy of the service provider. The P3P policy is published in the form of Web Services Description Language (WSDL) in the service provider. P3P Policy of the mobile device needs to be stored in the device itself. There has to be a specific way in which the P3P policy is stored in the mobile client. P3P file is being stored on the mobile device is in the form of a XML file. XACML policy file can be expressed in XML format.

2.2.5 P3P

The World Wide Web Consortium (W3C) P3P1.0 Specification [28] enables web sites to communicate their privacy practices in a standard format that can be retrieved automatically and interpreted by user agents. P3P user agents will allow users to be

informed of site practices (in both machine and human-readable formats) and to automate decision-making based on these practices when appropriate. Thus users need not read the privacy policies at every site they visit. Although P3P provides a technical mechanism for ensuring that users can be informed about privacy policies before they release personal information, it does not provide an automated mechanism for making sure sites act according to their policies. Products implementing this specification may provide some assistance in that regard, but that is up to specific implementations and outside the scope of this specification. However, P3P is complementary to laws and self-regulatory programs that can provide enforcement mechanisms. In addition, P3P does not include mechanisms for transferring data or for securing personal data in transit or storage. P3P may be built into tools designed to facilitate data transfer. These tools should include appropriate security safeguards. The standard P3P policy can be described in the Extensible Markup Language (XML) format and is explained with example [28]. It is important to note here that P3P relies mainly on trust. P3P policy has not been implemented in a handheld device such as mobile phone, PDA, etc. This work introduces P3P into the mobile space and hence proposes a new architecture for data access control using P3P policy in a mobile device. Trust between the service provider and the user's mobile device can be established by using the Security capsule.

2.2.6 XACML

XACML (eXtensible Access Control Markup Language), was formed by the OASIS (Organization for the Advancement of Structured Information Standards) standards consortium. XACML is a simple, flexible way to express and enforce access control policies in a variety of environments, using a single language. The XACML language in effect protects content from unauthorized use in enterprise data exchanges. XACML is mainly derived around and written in, XML, which is understood in most global environments. OASIS, which drives the development, convergence, and adoption of e-business standards, has ratified XACML. XACML gives an extensive and powerful set of features to the developers. It allows an organization to create and deploy authorization policies to match its mix of assets and business use-cases, then plug in additional policies as the business and its standards evolve. XACML helps in resolving issues related to security applications and there have been a number of papers published in order to prove the same. Q Xuebing et al. [29] detailed in their paper how XACML can be used to

solve some of the issues with mobile environment. This work introduces XACML in the mobile environment and hence proposes a new architecture for data access control.

XACML defines three top-level policy elements: < Rulle >, < Policy > and < PolicySet > [7]. The < Rule > element contains a Boolean expression that can be evaluated in isolation, but that is not intended to be accessed in isolation by a PDP (Policy Decision Point). So, it is not intended to form the basis of an authorization decision by itself. It is intended to exist in isolation only within an XACML PAP (Policy Administration Point), where it may form the basic unit of management, and be re-used in multiple policies. The < Policy > element contains a set of < Rule > elements and a specified procedure for combining the results of their evaluation. It is the basic unit of policy used by the PDP, and so it is intended to form the basis of an authorization decision. The < PolicySet > element contains a set of < Policy > or other < PolicySet > elements and a specified procedure for combining the results of their evaluation. It is the standard means for combining separate policies into a single combined policy.

2.2.7 Trust

There are various definitions of trust. McKnight and Chervany [30] defines trust as follows: Trust is the extent to which one party is willing to depend on somebody, or something, in a given situation with a feeling of relative security, even though negative consequences are possible.

2.2.8 Fusion

Fusion is the process of integrating lot of data and information and making sense from the information. It is also a process where it is represented in a consistent manner and for useful representation. Combined process of information from the source of same object or scene to obtain more complex, reliable and accurate information. Combination of data from multiple sources and gather that information into discrete, actionable items in order to achieve inferences, which will be more efficient and narrowly tailored than if they were achieved by means of disparate sources.

2.2.9 Subjective Logic

Subjective logic is a belief reasoning calculus that is compatible with, and an extension of probabilistic logic. It can for example be used for modeling trust networks, for modelling Bayesian networks, for Inteligence Analysis and logical argumentation. In general, subjective logic is suitable for modeling and analysing situations involving uncertainty, incomplete knowledge and different world views.

Beliefs are represented as opinions. A binomial opinion applies to a single proposition, and a multinomial opinion applies to a whole frame of multiple propositions (state space). Binomial opinions are equivalent to Beta probability density functions, and multinomial opinions are equivalent to the more general Dirichlet probability density functions. This makes subjective logic suitable for reasoning with evidence represented as Beta or Dirichlet probability density functions.

2.2.10 Security Vs Privacy

There is a clear difference between Security and Privacy and it is important to highlight that difference in the thesis here. Security is mainly confidentiality, integrity and availability of the data. Security puts in the appropriate processes in place to ensure that data is accurate and reliable. Privacy is the appropriate use of the data. The data should be used according to the agreed purposes and hence this is the key differentiator between privacy and security.

2.2.11 Crowdsourcing

Crowdsourcing information systems can be classified in many different ways. One of the classifications can be the nature of collaboration: explicit or implicit. In explicit collaboration systems (e.g.,Wikipedia or Linux), users collaborate explicitly to build information artifacts. On the other hand, implicit collaboration systems let users collaborate implicitly to solve a problem for the system owners.

2.3 Privacy related attacks and countermeasures

There has been a number of privacy related attacks that have come into existence today. One of the attacks is a sensor sniffing attack in which it assumes that the threat model is where the attackers are able to install malicious software onto the devices. This can be done by exploiting the software vulnerabilities or by tricking to install untrusted code. It is also assumed that the attacker has no physical access to the device but can receive the sensor date through voice or data channels. More details about this can be found in the work by L Cai *et al.*[31]. A number of viruses have been created to exploit the vulnerabilities that exist on today's mobile devices. One of the viruses, which originated in Spain, sends text messages to random mobile phone numbers [9]. As mobile phones become more and more intelligent the attacks against them will keep increasing. A number of vulnerabilities have been exploited using the Bluetooth capability of mobile devices leading to exposure of personal data [9]. Another potential attack that has been in existence is stealing user's personal data and downloading it without the consent of the mobile owner [9].

As widely discussed in [32], Information sharing is key to the operational efficiency of a coalition network. From an information provider's perspective, successful decision-making at the consumer using the shared data, indirectly results in utility for the provider and offers incentive for sharing. However, the act of sharing presents risks to strategic assets. The risk arises from the possibility that the data being shared may reveal more information to its recipient than was intended. Policies instituted at the producer to manage this risk often negatively affects the quality of decision-making at the consumer by introducing additional sources of uncertainty, thereby reducing the provider's utility.

One of the ways in which the balancing between risk and utility is achieved is through deliberate obfuscation of data. Obfuscation is a process through which the quality of the shared information is degraded in a controlled manner to protect against sensitive inferences regarding strategic assets [33]. This allows the data to retain utility while lowering the associated risk.

A report highlighted that Google's Android phones are vulnerable to privacy attacks [34]. The vulnerability results from the use of unencrypted wireless networks like Wi-Fi

to log into various Google services such as contacts, calendar and services like Picasa. When users request a digital certificate to sign into these services without re-typing the login information, Google's servers relay an authentication token back to the user's phone. This allows the user to be able to be logged into the accounts for 2 weeks without having to re-login. This sounds like a matter of convenience to the user but it has turned out to be a security flaw due to the fact that the authentication token is sent out in plain text. Malicious users can track the unsecured network and capture the authentication token thus allowing access to various services leading to a total breach of privacy. Another news has shown that some of Europe's biggest mobile phone companies signed up to new privacy guidelines published by the Global System for Mobile communication Association (GSMA) [35]. Signature based methods can be easily circumvented using code obfuscation necessitating a new signature for each malware variant [36], forcing the anti-malware client to regularly update its signature database. To tackle wide variety of new malware, a comprehensive evaluation framework incorporating robust static and dynamic methods can be proposed on Android platform. Manual analysis has become in-feasible due to the exponential increase in the number of unknown malware samples.

2.4 Classification of preserving privacy in mobile environments

The below architecture shows the complete classification of the different techniques used to preserve the privacy in mobile environments. It defines the problems involved as well as the techniques proposed to overcome these shortcomings. The privacy techniques are classified under two main headings: (1) *Data privacy* and (2) *Contextual privacy.*

Data privacy mainly involves the data that is being transmitted to and from the mobile device. This data could be in the form of a message, text or information. The data could be sensitive information or it could even be a confirmation on some booking that was done for an online shopping. Figure 2.2 shows the privacy classifications and within the data privacy section, it shows the 2 main areas of problems, i.e. the mobile query and the mobile resources. The mobile query could request the service providers for information that could be sensitive in nature. Hence this has always been a problem to understand and hence preserve the privacy of the information. In addition to

this, the data confidentiality is guaranteed through authentication. The other area of classification of privacy is based on the contextual privacy.

Contextual privacy can be further divided into two areas namely location privacy and identity privacy. When mobile users request for static resources or mobile resources, pseudo–identifiers are sent and location is anonymized. The data that is being transmitted is protected against third party malicious users. Although the information can be assumed at all times to be hijacked by malicious users, malicious users protect the data against unauthorised access. This is achieved by using data access control mechanisms such as P3P policy extension and XACML policies. These are described in detail in the later sections. Location privacy mainly deals with the location of the requester. In mobile environments, users are frequently requested for their location information when they try to access a new online service. For example, when a user requests for nearby restaurant information from a location-based server, the location based server needs to know the location of the user and hence the location information is normally requested. However, in most of the cases, the user doesn't want to disclose the location information to arbitrary location based service providers. This can be achieved by a number of different mechanisms described later on. To briefly name the mechanisms here as shown in Figure 2.2, let's start with k-anonymity. In this method, user's location information is updated with pseudo-IDs and then the generalised location information is sent to the location based service provider. Due to some groups being created that fail to provide overall anonymity, another mechanism called s-proximity has been implemented [37]. This mechanism creates a larger number of anonymous user profiles to ensure that the location based service provider cannot identify the location of the requester. Another location privacy mechanism that is described in here is Casper [38]. Casper is a combination of location anonymizer and privacy aware query processor. Few other mechanisms like the encrypted data store [39], key agreement [40], privacy tools [41], In-device spatial cloaking assisted by cloud [42] are also part of the location privacy and are described in detail in the future sections. Contextual privacy has another classification namely Identity privacy. Identity privacy mainly talks about the user/mobile server requester who issues the requests. In order to preserve the identity of the user who issues the requests, a number of mechanisms have been explored. They are mainly user profile pseudo-identifier conversion, privacy aware query processor and authentication based methods. Each one of them is detailed in further sections.

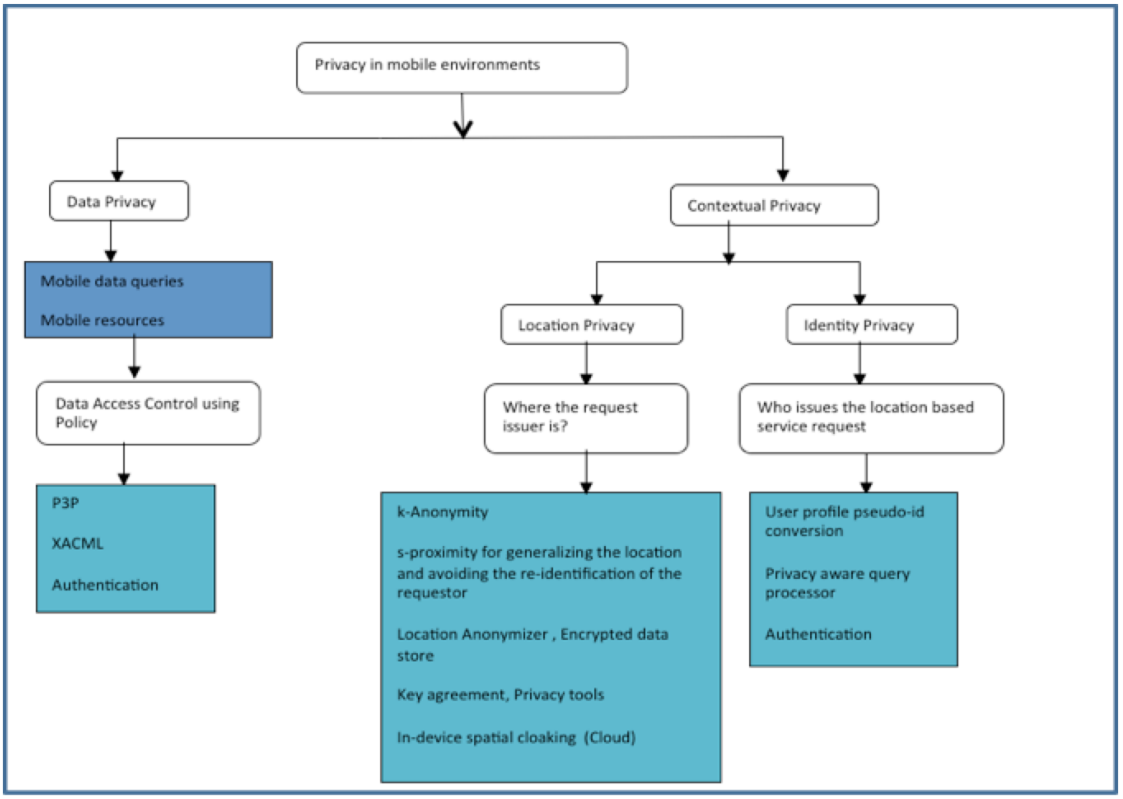

FIGURE 2.2: Classification of privacy preservation mechanisms in mobile environments.

2.4.1 Profile anonymization model

Preserving privacy using anonymization has been discussed in a number of research papers [37], [43], [44], and [45]. The authors in [46] have looked at the k-anonymity in order to generalise the location. The user of a mobile device usually requests information for 2 main types of resources namely static resources and mobile resources. In case of static resources, pseudo-identifiers are sent and the location is anonymized. In the case of mobile resources, IDs are updated with pseudo-ids and then the generalised location and profile are sent back to the requester.

Although the profile anonymization model works well using the k-anonymity, there have been a number of attacks that can be performed on the k-anonymity model that has led to the identification of the query issuer in the location based services. To overcome some of the shortcomings in the current k-anonymity model the s-proximity model was proposed. The next section discusses the advantages of the s-proximity compared to k-anonymity model.

2.4.2 Identity inference protection using s-proximity in location based services

The k-anonymity model as described in the previous section tries to hide the location of the query issuer who tried to request for location based information from the Location Service Provider (LSP). In [37], the author shows that the k-anonymity is not enough as it can be easily prone to attacks thus resulting in the re-identification of the query requester. Two main attacks that have been depicted in the work are heterogeneity attack and conformity attacks. K-anonymity can create groups that fail to provide the overall anonymity due to lack of sufficient match among members with respect to some sensitive user attribute. The communication between the query requester and the LSP is as follows: Initially, the user sends a location-based query to the Location Anonymizer (LA), which then replaces the exact location with a Cloaked Region (CR). It is then passed on to the LSP. The attacks prove that in this process, by some combined work by the LSP's or an LSP can individually break down the anonymity set and prove the identification of the specific query requester in cases where the query is specific or not too generic. Hence, [37] proposes a solution that generalises the query and hence makes it difficult for the LSP to identify the actual query requester. This is achieved in the s-proximity model. The work suggests that both k-anonymity and s-proximity are needed to anonymize the query requester's identity in a location-based service. In the s-proximity model, the LA is replaced by context aware LA with further modules such as query generalization, query analyser and partitioning agent. With the detailed implementation of these modules privacy of the user is preserved and hence the privacy preservation is achieved in location-based environments.

2.4.3 Casper: Query processing without compromising privacy

Originally, Walid and his co-authors present a new privacy-aware query processing framework, Capser*, in which mobile and stationary users can obtain snapshot and/or continuous location-based services without revealing their private location information. In particular, they propose a privacy-aware query processor embedded inside a location-based database server to deal with snapshot and continuous queries based on the knowledge of the user's cloaked location rather than the exact location. Their proposed privacy-aware

query processor is completely independent of how they compute the user's cloaked location. In other words, any existing location anonymization algorithms that blur the user's private location into cloaked rectilinear areas can be employed to protect the user's location privacy. The new Casper: a new framework in which mobile and stationary users can entertain location-based services without revealing their location information. The method addresses the issue of user having to give away the location information while requesting for any location-based services through a location based database server. Casper involves two main components namely, location anonymizer and privacy aware query processor. The work [38] describes in detail how exactly the two main components perform with regard to the four novel areas of scalability, quality, efficiency and flexibility. Casper functions mainly in the following manner. When the mobile user sends the location information along with the query request for a particular location based service, the location anonymizer picks it up and blurs the location information to a spatial region along with the query and passes it to the location based database server. The privacy aware query processor is built into the location based database server and it looks at the request and returns a set of answers that matches the mobile users query. The Figure 2.3 shows the mobile device making a request to the location based service provider. This is passed through the anonymizer and into the location based database server. The anonymizer does its task and the privacy aware query processor performs its function and the most relevant out of the four data and query would be passed on to the location based service providers.

2.4.4 Casper model

The authors in [38] also point out to three novel types of data and query that Casper handles. According to them all the traditional anonymizers can only work on the public query over public data. In [38], the authors propose three novel areas of transactions namely, private query over public data, public query over private data and private query over private data. A detailed analysis of the three methods is shown and the authors assess its performance and scalability.

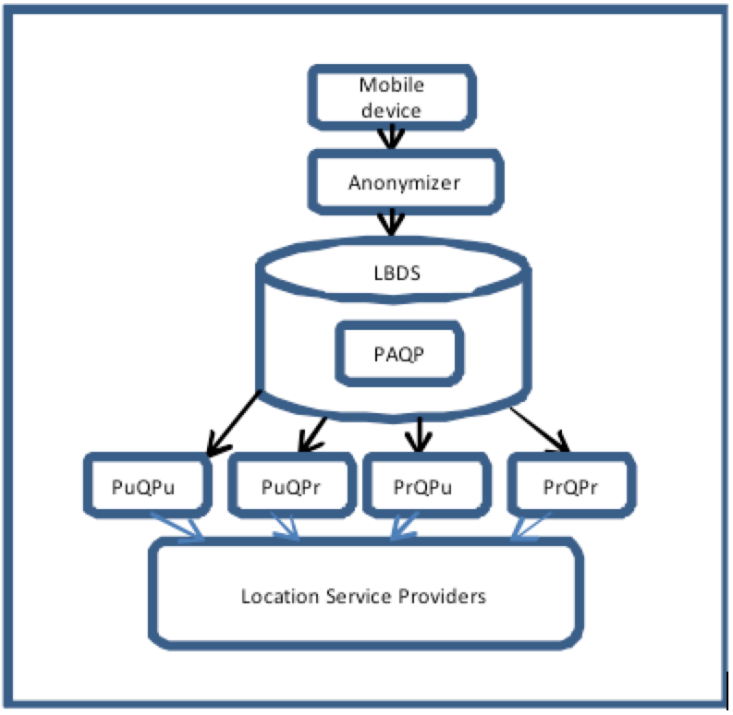

FIGURE 2.3: Casper model.

Casper functionality with private query over public data (PrQPu), public query over private data (PuQPr), and private query over private data (PrQPr) can be shown as:

$(q_{pr}, d_{pu})(q_{pu}, d_{pr}), (q_{pr}, d_{pr})$

$(M_{lr}) \rightarrow$ mobile request and location

$A \rightarrow$ Anonymisation

$(C_{lr}) \rightarrow$ Cloaking region

$A(M_{lr}(Q, D))$

$Q, D \rightarrow (q_{pr}, d_{pu})(q_{pu}, d_{pr}), (q_{pr}, d_{pr})(q_{pu}, d_{pu})$

$A(M_{lr}(Q, D)) = \{C_{lr}(Q, D)\}$

$\{C_{lr}, (Q, D)\} =$

$\{C_{lr}(q_{pr}, d_{pu})\}\{C_{lr}(q_{pu}, d_{pr})\}$

$\{C_{lr}(q_{pr}, d_{pr})\}\{C_{lr}(q_{pu}, d_{pu})\}$

According to the authors, by using the Casper's novel solution, the location information

will never be compromised. They also address another level of anonymizer called the adaptive location anonymizer, which works, similar to the original location anonymizer with some differences. Further details can be found in [38].

2.4.5 Encrypted data store to preserve privacy

Location based social applications (LBSA) are used considerably in today's smartphones. Smartphones using these applications send location information to untrusted third party servers. In [39] the authors argue that the LBSAs should adapt an approach where the untrusted third-party servers are treated simply as encrypted data stores, and the application functionality be moved to the client devices. The location coordinates are encrypted, when shared, and can be decrypted only by the users that the data is intended for. This approach significantly improves user location privacy. The authors also argue that this approach not only improves privacy, but also is flexible enough to support a wide variety of location-based applications used today. Location information can be easily accessed by the third party servers and hence can be passed on to other sources due to various reasons as mentioned in [39]. In [39], the authors propose a design for building LBSAs that provides a low-cost, practical, and deployable alternative to existing design while providing strong user location privacy. The key insight behind this design is to treat the server as a simple encrypted data store, and move the application functionality to the client's smartphone. All the location information shared is encrypted and the lack of plain location information on the storage server improves user privacy. This approach easily works on today's smartphones because the servers running LBSAs today provide their service by running simple operations such as certain database or hash table lookups, performing simple computations on the location data, and sending the results to be displayed on the clients terminals. For example, in a nearby restaurant review application, the server takes the user location, finds restaurants that are in the vicinity of the user's location, queries the reviews of these restaurants, and sends the results back to the users for display. In the proposed approach, the data storage and lookup operations happen on encrypted data but still remain on the storage server. The clients receive the encrypted results, decrypt and display the results to the users. The clients only incur an additional cost of decrypting the received content, and perform simple calculations on the decrypted data It is important to notice that the encrypted

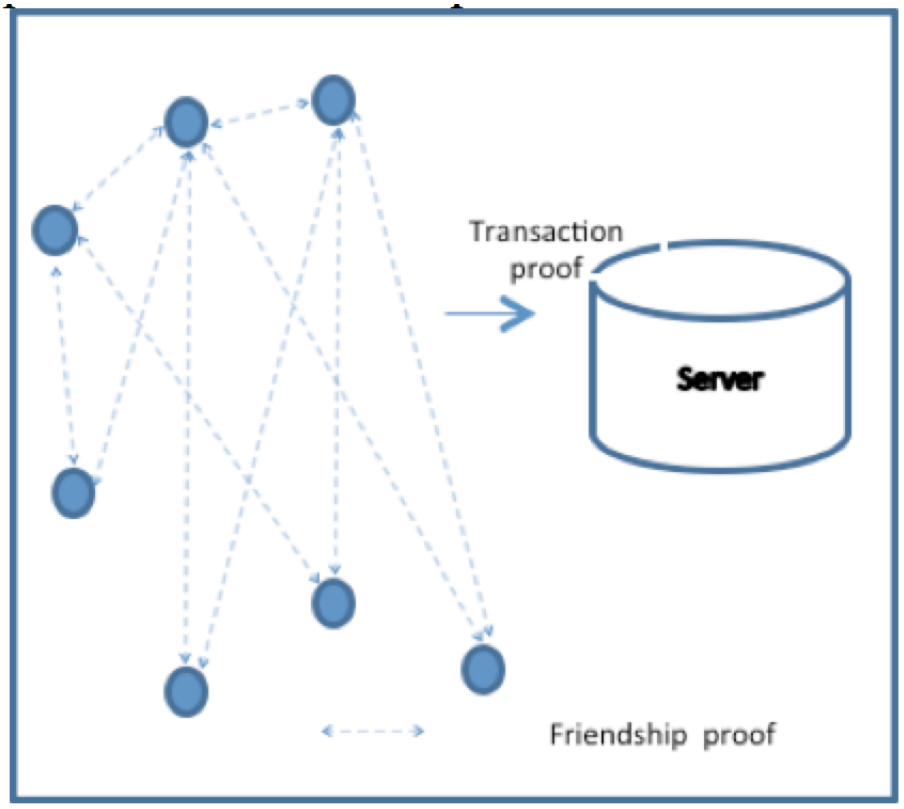

FIGURE 2.4: Encrypted data store model.

data store. This encrypted data store obfuscates the location information and this is another way of location anonymisation that is used for location privacy.

Figure 2.4 shows that friends exchange friendship proofs and store them in their devices and then users generate and store the transaction proofs in the server and their friends later on retrieve this. By using lightweight cryptographic schemes such as encryption, decryption with real time keys, the authors claim that they can easily move the functionality to the smartphones and provide services while preserving privacy. The work discusses two proofs namely, friendship proof and transaction proof.

Friendship proofs cryptographically attest the social connection (or friendship) between two users, and similarly, transaction proofs cryptographically attest certain data generated by a user. Using these proofs, any user in the network can verify if it is a friend, and if so decrypt the data generated. But no other user other than a friend will be able to see the contents. Finally, the interface exposed by the storage server is narrow enough that one can reason about the privacy guarantees, and yet they are flexible enough to build several LBSAs. As a result, a single storage server can support many different LBSAs.

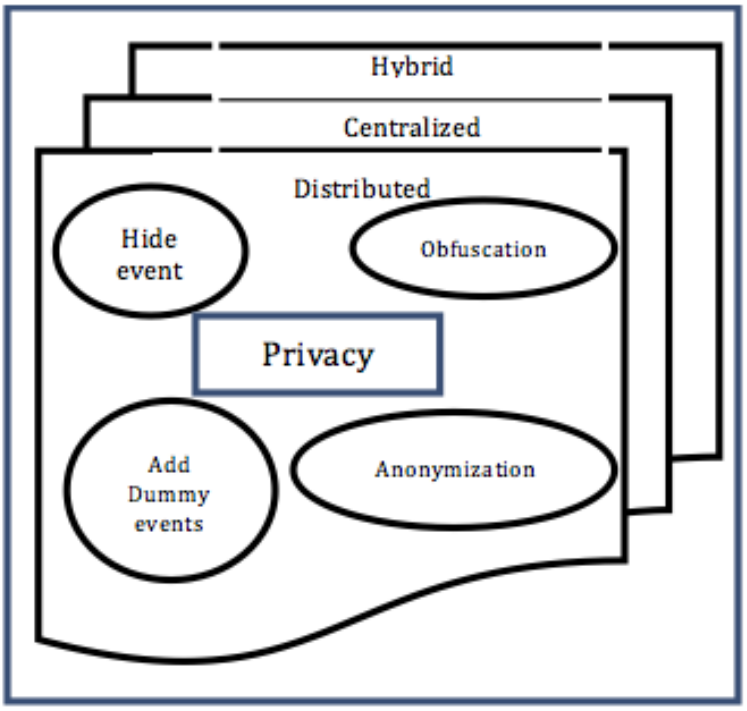

FIGURE 2.5: Location privacy with privacy tools.

2.4.6 Unified framework for location privacy

According to work [41] there are three entities that play a role in preserving location privacy: users, applications, and privacy tools. Each entity controls the amount of shared information and thus affects user privacy. Users and applications might intentionally (e.g., by being cautious about sharing unnecessary information) or unintentionally (e.g., by sharing incorrect information) reduce the amount of information revealed. Privacy policies influence the way applications can share information with different entities, and they are applied to the application based on the users' decisions. Various privacy tools [41], also use sophisticated algorithms to guarantee users' privacy. In order to capture the effect of the three entities in preserving location privacy of users, in [41], they abstract the entities and model a location-privacy preserving mechanism as a single unit that separates actual events of the users and the adversary. The work [41] defines a location-privacy preserving mechanism as a transformation function that modifies the users' actual events before they can become observable by any observer. The work discusses the privacy tools in detail.

Privacy tools work in three architectures: (i) Distributed (user-side): They can work in a distributed way by being implemented on individual mobile devices, where each

device transforms its events and modifies what an observer can see about the user's spatio-temporal state. This can be done either with the help of information that a device gets from other devices or exclusively with the information that the user has. (ii) Centralised (server-side): They can work in a centralised manner by using a trusted central server that acts as a privacy preserving proxy and modifies users' messages (correspond to events in our model) before being observable by an untrusted entity. (iii) Hybrid: They can be a hybrid of both distributed and centralised architectures. The four main functions in the location privacy preserving mechanism include hiding events, adding dummy events, obfuscation and anonymization.

2.4.7 Authentication and key agreement for location privacy

A. Loukas *et al.* discussed in [40] about mobile instant locator with chatting capability along with preserving privacy and security. Mobile instant locator with chatting (MILC) was developed for usage within a closed community and hence worked very well in the University scenario described in the work. The research in [40] also highlights that with its popularity grew its demand and since it also incorporated privacy preserving techniques it was very attractive to other communities too. MILC works towards making the communication confidential and maintaining the privacy of the user. According to [40] the MILC server is developed in Java. The client server communications are handled using the RSA 1024 bit asymmetric keys. Client gets successfully authenticated with the server and from then onwards every communication between the two ends is secured by using a symmetric session key created at the server end. The research in [40] proves that supporting pseudonymity and location privacy can preserve the end-user privacy. The option of presenting or disclosing the location is left to the choice of the user. If the user decides not to disclose the location, user's privacy is maintained. Pseudonymity is provided per session. When the user connects to the MILC server is offered with the option of choosing a different pseudonym for the current session. In [40], authors have also compared MILC with three other applications [47], [48], [49] and show comparison based on security requirements. The comparison is based on the following six basic criteria: mutual authentication, confidentiality, integrity, pseudonymity, resistance to DoS (Denial of Service) caused by insiders and location privacy. The comparative view of all the applications considering the above mentioned seven basic criteria show that the applications support user authentication and pseudonymity. MILC additionally provides

mutual client–server authentication. Moreover, the pseudonym of a MILC user cannot be associated with the permanent identity in any way. Excluding MILC, BuddyMob is the only one supporting location privacy, but this applies for guest users only.

2.4.8 In-device spatial cloaking assisted by cloud

A number of privacy mechanisms proposed mostly deal with single point of service. When there is a single point of service, things are bound to go wrong somehow somewhere. Song and Sean [42] talks about the cloud services available that makes it so much more versatile in terms of the services being available in the cloud. The authors describe how the location based services that are requested by the mobile device are delivered to them by means of using spatial cloaking that is assisted by cloud capabilities. There are clients in the mobile devices that would be responsible for generating the cloaking region. The main difference of the In-device spatial cloaking solution in comparison to the Casper solution is that here it is the device generating the cloaked region and hence the work strongly portrays that using the in-device cloaking privacy can be preserved. The in-device spatial cloaking solution involves a location trusted server, which takes the information from the mobile device strips that information and carries only the spatial cloaked information and the service request and passes it on to the service provider.

The works described above shows the various methods used in research for preserving privacy in mobile environments. The details for each of the methods have been included in their respective sections above with references to their work. The main reason to showcase all the work above is to highlight the existing research work in the area and the drawbacks in the existing solutions leading to my thesis on preserving privacy in the mobile environments.

2.5 Policy based access control

Rei is a policy specification language defined in Web Ontology Language (OWL)-Lite [50]. It is a highly expressive and extensible declarative policy specification language well suited for describing security policies in pervasive environments. Rei includes constructs for expressing rights, prohibitions, obligations and dispensations. It also includes constructs for setting positive or negative modality preferences and allows for stating

priority between policies. It models speech acts like delegation, revocation, request and cancellation. This allows policies to be expressed in a less exhaustive way and also allows for distributed policy management. In [51] and [52], the authors show how policies can be used for guiding the behavior of entities within a domain. The advantage of using policies lies in being able to modify the security functionality without having to change the implementations of the entities themselves. It defines a policy as a set of rules describing concepts like permission, prohibition, obligation and dispensation over possible actions in the environment with respect to the requester, the action and the context. For example, a privacy policy about not disclosing an Social Security Number (SSN) would be a prohibition over taking any action that results in the SSN being disclosed. Rei allows the inclusion of Prolog-like variables that extend the expressivity of Web Ontology Language (OWL). These variables allow relations like uncle of, same age as, and different group from that are not directly possible in OWL .

It also models speech acts for remote policy management like delegation and revocation that affect permissions and prohibitions, and request and cancel that affect obligations and dispensations. Another set of specifications included in Rei are those for meta policies. These are used to resolve any conflict that may arise. For example, if a user is both permitted to and prohibited from performing a certain action, then the meta policies are used to decide whether the permission or the prohibition holds. Rei is capable of describing deontic concepts over entities and actions based on their properties. Policies can be written in Rei that are based on properties of entities and other domain conditions. Actions can be generically described specifying a subject “X” a target “Y” and imposing constraints on both “X” and “Y” to satisfy certain properties. e.g. in plain English, the policy would describe “Action A with target Y can be granted to subject X, provided X satisfies certain properties and Y satisfies certain properties”. Consider the following scenario: The “A” lab policy states that devices owned by “A” lab in possession of people affiliated with “A” lab, are allowed to use the capabilities of these devices inside the “A” lab, but not if they leave the lab. The Rei ontology can be augmented with a suitable domain specific ontology which provides a sufficient vocabulary to describe the security policy. For more expressive policies Darpa Agent Markup Language (DAML)+ Ontology Interchange Language (OIL), OWL-Lite can also be used.

Thus a policy written in Rei is able to specify in abstract terms the safe or acceptable

use policy for the set of trusted devices. The policy enforcer along with the context manager ensures that the appropriate policy is enforced and updated periodically.

Knowledgeable Agent-Oriented System (KAoS) is a policy language based in OWL and is similar to Rei as it can be used to develop positive and negative authorization and obligation policies over actions. KAoS policies are descriptions of actions that are permitted (or not) or obligated (or not) limiting its expressivity as policies are restricted to OWL. However, KAoS allows the classification of policy statements enabling conflicts to be discovered from the rules themselves. The Rei engine includes run-time conflict resolution but cannot predetermine conflicts. Another advantage of KAoS is that, if policy descriptions stay within OWL-Lite or OWL-DL (Definition List), then the computation is decidable and has well understood complexity results. In terms of speech acts, however, KAoS only models delegations, whereas, Rei includes an integrated approach to speech acts for policy management, which is useful in autonomic, distributed systems. Rei also provides specifications and tools for policy analysis and consistency checking that KAoS does not. Semantic Web Rule Language (SWRL) is a rule language for describing Horn like rules in OWL [53]. There are very few tools for it like Hoolet [54] that are developed. There has been some work done on attribute based access control [55] by Vincent and his colleagues. Sun-Moon Jo has worked on research around access control [56] system for telemedicine secure XML documents.

2.5.1 P3P based access control

Number of research work has been carried out in the area of P3P and Web Services [57] - [58]. M. Zuidweg *et al.* [57] described P3P in a web-services based context-aware application platform. They proposed the requirements for applying a P3P based privacy control mechanism in context aware Web Architecture for Service Platforms (WASP). G Myles *et al.* worked on preserving privacy in environments with location based applications [59]. They describe the initial stages of an extensible system for enabling privacy in environments that support location-based applications. They show [59] a privacy system especially to protect information related to personal location. There has been some work done in the area of personalised applications in ubiquitous environments. Brar and Kay described the underlying concepts of privacy and security in ubiquitous personalised applications [3]. A study of the exact requirements of security

and privacy rights in ubiquitous environments has been conducted by M Fahrmair *et al.* [60]. The conclusion of the work resulted in describing a system that enables description and enforcement of limitations for the user of confidential/private data. Trust can be used to provide fine grained control over the use of personal information resulting in managing privacy. A trust based approach to control privacy exposure in ubiquitous computing environment is proposed by P D Giang *et al.* [61].

2.5.2 XACML based access control

There has been some new security architecture for the integration of mobile agent and WebServices technology [62]. This architecture provides a new authentication scheme for Web service provider to verify the mobile agent owner's identity by employing an identity-based signature protocol without using the username/password pair, which is infeasible for mobile agent. Furthermore, the server only needs one key to encrypt one service that can be available to a group of users. The complete composite access control model for mobile agents is described in [62] where they have presented two key aspects of the role-based access control for mobile agents: a) authorization infrastructure, and b) the structure of role-based access control policies. The policies map agent role to user role and provide a composite policy for PDP (Policy Decision Point) decisions regarding access control applied to mobile agents. Currently policies are limited to a single domain. Future research is planned to be pursued to identify role-based XACML policies for mobile agents in a federation. The challenges and requirements existing in mobile environments relates to the subject ID. The Subject ID of the requester plays a critical role in both the authentication and authorization stages for any access control system. Unfortunately, a subject ID issued by one domain will probably not be recognised by another; that is to say, a mobile user requester cannot be authorised by a local domain until he/she obtains a valid subject ID from the local domain. SSO (Single sign-on)-based Identity Management seems to solve the problem, however it doesn't work very well for temporary subject IDs that could last as little as oneday. The traditional XACML polices, used for user access control in distributed environments, can be used for mobile agents' access control [63]. Such polices are used to manage delegation of access rights from users to agents while at the same time following the core principles of the XACML standard. In [63], the authors propose a combination of policies that map

users to their mobile agents and make access control decisions for mobile agents by evaluating complex policy sets. An XACML-based architecture is proposed in [29] to tackle the problems of compromise to the requester's data confidentiality and integrity, and the issue of applicability of reputation data. A Subject ID mapping service is the foundation of the architecture, upon which a Meta Policy Server (MPS) is designed to locate the policies for a requester and provide guidelines for overall security management, while reverse authorization is used to guarantee the requester's privacy. In addition, a private reputation attribute authority (AA) handles reputation data applicability problem. A security handshake protocol for secure communication between the MPS and subject attribute authorities is also an important part of the solution.

The above sections on policy based access control cover the details of the various policy languages that have been in use. This chapter also details the access control methodologies that have been considered using P3P and XACML.

2.6 Geospatial access control

In the previous section, I had covered policy based access control and introduced the review on XACML based access control models. This section looks at the work that has been done along the lines of geospatial access control for mobile devices. It also looks at the importance of location attestation and the various approaches that research has explored in order to achieve attestation.

2.6.1 Geospatial access control for mobile devices

Jansen *et al.* [64], [65] describe an implementation of assigning and enforcing policies on handheld devices using Java smartcards. The organization policy is distributed via tamperproof smartcards. All the devices are smartcard enabled. The policy to be enforced is read from the smartcard, which requires authentication by a username and pin. After authentication a card monitor continuously monitors the existence of the smartcard. If the smartcard is removed the device reverts to the default policy of the device. In [65], [66] the authors describe the use of a policy specification language, a policy distribution mechanism and certificate representation. An XACML-based architecture is proposed [29] to tackle the problems of compromise to the requesters data confidentiality and

integrity, and the issue of applicability of reputation data. The traditional XACML polices, used for user access control in distributed environments, can be used for mobile agents access control [63]. Such policies are used to manage delegation of access rights from users to agents while at the same time following the core principles of the XACML standard. In [63], the authors propose a combination of policies that map users to their mobile agents and make access control decisions for mobile agents by evaluating complex policy sets. The work in [67] deals with the use of P3P policy by extending it for data access control and use of XACML policy [68] in the mobile device for data access control. Xuebing *et al.* [29] detailed in their work how XACML can be used to solve some of the issues with mobile environment.

GeoXACML [69] is a geo-specific extension to XACML 2.0 and it is standardised by Open GeoSpatial Consortium (OGC). GeoXACML supports the declaration and enforcement of geospecific access rights. It also makes sure that it controls access to services, data and other information in a service oriented type architecture. GeoXACML [70] has been standardised but there has not been any open standard implementation of this to control the access rights to resources. My work is the first attempt to a fully functional implementation of GeoXACML for access control as per the authors understanding. Roese and his team worked on research for location based access control in data network [71]. Abedi and team had done some work on tracking spatio-temporal movement of human [72] in terms of space utilisation.There has been some work that has been done on multi-granularity spatial-temporal access control model [73].

2.6.2 Location attestation

Saroiu *et al.* [74] have described location proofs as a new mechanism that enables the existence of mobile applications that needs proof of the user's location. The wireless access point to mobile devices handles it. The solution is mainly based on users and wireless access points (APs) exchanging their signed public keys to create time-stamped location proofs. The research describes an implementation of the location proofs. The work shows six potential applications that would be used by an infrastructure that provides location proofs. It also showcases a protocol that is demonstrable over WiFi and characterises security properties of the design. The research details the difficulties that come from collusion attacks such as when sharing devices with one another. VeriPlace

[75] is a location proof architecture that takes care of the challenges involving user trying to fool the systems by receiving location proofs for locations where they are not located. These solutions take care of the user's privacy and are capable of detecting cheating.

VeriPlace used cryptographic techniques in order to achieve system security. VeriPlace needs three types of trusted entities that are run by different parties to avoid collusion. To protect users privacy, each trusted entity is aware if either a users identity or the location, but not both of them. VeriPlace uses Cheating Detection Authority (CDA) to check if any cheating has occured. Using the encrypted access point information, the CDA decrypts it and checks whether any two APs are far from each other, if yes it indicates cheat. Zhu *et al.* [76] proposes a Privacy-Preserving Location proof Updating System (APPLAUS) in which co-located mobile devices, which are bluetooth, enabled generate location proofs and then transfers the changes to a location proof server. In order to protect source location privacy pseudonyms, which are changed periodically, are used. The solution also shows a model in which the users can assess their location privacy levels and when and whether to receive the location proof requests. The work also shows a way to secure from colluding attacks by presenting betweenness ranking based and correlation clustering basd approaches for outlier detection. Implementation and deployment of APPLAUS is easily done in bluetooth enabled devices and doesn't require a lot of computation cost.

Spatial-Temporal provenance Assurance with Mutual Proofs (STAMP) scheme [77] is a solution that gives security and privacy assurance to mobile users proofs for their past location visits. STAMP is based on mobile devices in vicinity

to mutually generate location proofs. Some of the key features of STAMP have been to maintain the integrity and non-transferability of the location proofs and location privacy of the users. The implementation also shows that it requires very low computational costs to execute this on the mobile devices. Hasan *et al.* [78] analysed the secure location provenance problem and introduced methods for making location proofs resistant against collusion. The proofs lets the user prove the location in different granularity to a number of auditors. The research shows two schemes using hash chains and bloom filters. Some of the experimentation results from the proof of concept shows that these schemes are realistic in today's mobile environments.

Another interesting research [79] shows that they have designed two privacy-preserving alibi schemes: one for corroborators who have no personal privacy issues and another one for corroborators who want to keep control over the disclosure of their identities. The research also demonstrates that schemes are implementable and usable in today's mobile devices.

I have reviewed the geospatial access control models for mobile devices in the above sections and also reviewed the methods used to achieve location attestation. It is interesting to see that location attestation has been addressed from a local attestation perspective which has issues when it faces consistency check. In my work in further chapters, I will be showing the benefits of using global attestation of location for mobile devices.

2.7 Trust in mobile environments

This section highlights the research contributions in the trust environment. I have also tried to highlight the key differences in the existing research and my contributions to the area of trust.

2.7.1 Trust based solutions

There has been research in areas of trust based data management. Patwardhan *et al.* [80] propose a trust-based data management framework for enabling individual devices to harness the potential power of distributed computation, storage, and sensory resources available in pervasive computing environments. They take a holistic approach that considers trust, security, and privacy issues of data management in these environments. Lim *et al.* [81] has explored data stream management and exploited the notion of confidence policy while taking into account trustworthiness of data items in data management and query processing. They propose a provenance-based framework that enforces confidence policies in the evolution of continuous queries over streaming data. Based on the notion of physical and logical networks, they introduce the notions of the physical and logical provenances for data items, respectively. The authors also introduce a cyclic framework of computing actual trust scores of data items and network nodes based on the value and provenance similarities embedded in data items. Unlike these approaches, in my work,

I use unstructured reports from unreliable information sources and derived structured knowledge based on Controlled English(CE -natural language processing mechanism) using International Technology Alliance (ITA), which is a military project for the military of United States of America and United Kingdom) assets. ITA is a collaborative research alliance between United Kingdom Ministry of Defence (UK MoD) and US Army Research Laboratory (ARL) and a consortium of leading industry and academic partners. Some of the assets that has resulted due to this research collaboration includes Controlled English [82] and Information Fabric [83]. Controlled English (CE) is an ITA asset that has been extended and modified by the ITA programme. Mott defines the syntax with examples of CE formatted text and shows how to map to predicate logic [82] . CE is a type of Controlled Natural Language (CNL) and is mainly used to be readable by any one who knows English. It represents information in a very structured form using certain syntax. This syntax can be parsed and the content can be understood by a computer. Hence, CE enables the communication between computer and humans. Role of transfer based and performance based cues on initial trust in mobile shopping services was an area explored in the recent paper [84]. It is cross environment perspective and an interesting angle to the trust problem. The role of security, design and content factors on customer trust in mobile banking has been explored in the recent paper [85]. Schwitter has described the use of controlled natural language as a high level specification language for modelling commonsense reasoning problems [86]. The work shows how defaults and exceptions can be incorporated into a current controlled natural language and what is required to signify and motive with them in a not so straight way.

In my work which I explain in further chapters, I use a trust model that is based on trust evidence. The retrieval or derivation of trust evidence is out of the scope. Usually, there are two types of evidence: direct and indirect. In the literature, there are several trust approaches where direct evidence is combined with indirect evidence to model trust in information sources. Direct evidence is based on personal observations, while indirect evidence is received from others. Jøsang and Ismail proposed the beta reputation system (BRS) [87]. It estimates the trustworthiness of an information source using beta probability density functions. For this purpose, aggregation of direct evidence and indirect evidence from information sources are used as the parameters of beta distributions. Evidence shared by sources are equivalent to binary opinions in Subjective Logic [1]. Whitby *et al.* extended BRS to handle misleading indirect evidence from malicious

agents using a majority-based algorithm [88].

Teacy *et al.* proposed TRAVOS [89], which is similar to BRS, but it uses personal observations about information sources to filter misleading indirect evidence. Subjective logic has been extended to include belief updates from partially visible evidence [90]. Lance *et al.* show that assets of the partial observable update as a function of the state likelihood and demonstrate the use of these likelihoods for a trust estimation application.

The value of the partial observable updates is shown through different imitations including the trust estimation case. Yu and Singh proposed a trust approach that handles misleading indirect evidence using a version of weighted majority algorithm [91]. In their algorithm, weights are assigned to information sources. These weights are initiated as 1:0 and can be considered as the trustworthiness of the corresponding sources. The algorithm makes predictions about trust related propositions (e.g, 1 is trustworthy) based on the weighted sum of indirect evidence (i.e., ratings) provided by those sources. The authors proposed to tune the weights after an unsuccessful prediction so that the weights assigned to the unreliable sources are decreased. They assume that the ratings from dishonest sources may conflict with the personal observations. By decreasing the weights of these sources over time, misleading evidence is filtered.

In chapter 5, I describe conflicts between binomial opinions and analyse evidence about information sources to resolve conflicts between opinions before performing fusion. Conflicts in knowledge lead to inconsistencies that hamper the reasoning over the knowledge. Therefore, before using such knowledge bases, their conflicts should be resolved. Gobeck and Halaschek in [92] present a belief revision algorithm for ontologies, which is based on trust degrees of information sources to remove conflicting statements from a knowledge base. However, as the authors point out, the proposed algorithm is not guaranteed to be optimal. Dong *et al.* [93] propose to resolve conflicts in information from multiple sources by a voting mechanism. Double counting in votes is avoided by considering the information dependencies among sources. The dependences are derived from Bayesian analysis. When considering multiple sources espousing multiple claims, it is possible to estimate their reliability through corroboration without direct and/or indirect evidence. For example, fact finding algorithms aim to identify the truth given conflicting claims. Yin *et al.* proposed TruthFinder [94] that utilises an iterative approach to estimate trustworthiness of information sources and information they provide. Their approach

based on the assumption that a source is trustworthy if it provides many pieces of true information, and a piece of information is likely to be true if it is provided by many trustworthy sources. Therefore, very similar to BRS, TrustFinder also assumes that the information provided by the majority is trustworthy.

The above section has described trust in mobile environments and how my approach defers from what research has done so far. I highlighted the importance of fused opinions and need for trust when there are continuous streaming of information.

2.8 Location based mobile security

In this section, I describe the various methods that have been used to look at location and identity privacy related issues and solutions proposed by various research work in the past.

2.8.1 Location privacy

A number of papers related to mobile environments and its vulnerabilities have been published in the recent past. Sniffing attacks have been talked about in [31]. They explore the vulnerability where attackers snoop on users by sniffing on their mobile phone sensors, such as the microphone, camera, and GPS receiver. [95] discusses about Soundcomber, which is a stealthy Trojan with innocuous permissions that can sense the context of its audible surroundings to target and extract a very small amount of high-value data. As sensor-rich smartphones become more ubiquitous, sensory malware has the potential to breach the privacy of individuals at mass scales. There have been a number of different papers concentrating on the different vulnerabilities of mobile devices and how the operating system in the device allows users to control access to sensitive information including location, camera images, and contacts. In [96] the authors have introduced TaintDroid, which operates as an efficient system wide information flow-tracking tool. This tool has the capability of tracking multiple sources of sensitive data. The authors also studied the behaviour of thirty popular third party applications chosen at random from Android marketplace and concluded that two-thirds of those applications display suspicious handling of sensitive data. A paper is dedicated to the mobile phone vulnerabilities, which talks about the different malwares that are targeted on the mobile

devices [9]. The work details on how some of the malwares can be implemented easily in order to make the mobile phones vulnerable to attacks. Preventing the cell phones from malicious users or infiltrators is very important and there have been a number of research papers concentrating on the same. In [97], VirusMeter is detailed which detects existence of malware with abnormal power consumption. VirusMeter relies on a concise lightweight user-centric power model and aims to detect mobile malware in two modes: While the real-time detection mode provides immediate detection, running VirusMeter under the battery-charging mode can further improve the detection accuracy without concerns about resource consumption. Using real-world malware the authors have experimentally shown that VirusMeter can effectively and efficiently detect their existence. In [98] the authors adapted a special and feasible method, blind signature, to generate an authorised anonymous ID that replaces the real ID of an authorised mobile device. They presented a two-phase protocol to address location privacy, however, did not consider that the randomness introduced during the blinding phase can be removed easily. They also prove that the administrator can link real ID with authorised anonymous ID. In addition to this they propose an improved registration and re-confusion protocol using the same cryptographic technique, blind signature based on bilinear pairings. A considerable amount of research work has been carried out in the area of location-based applications. In [10], authors propose a security model for location based services using outsourced databases and demonstrate how one can use distributed storage and international mobile subscriber identity (IMSI) as user identification to secure the location data. In [99], the authors investigated the problem of protecting location privacy of mobile users in the setting of ubiquitous computing. They find it challenging, as there are various requests that are forced by the organisation and the users. In [99], [100] the authors proposed an authorised-anonymous-ID based scheme, which is used to replace the real ID of an authorised mobile device. With authorised-anonymous-IDs, they also designed an architecture that is able to provide the mobile users with complete control over their location privacy and still allowing the organisation to authenticate the mobile users. In [101], the authors have designed novel protocols to provide location-based services, which do not require a user to trust a third party. They also analysed a class of location-based services that do not directly transfer user locations. L Barkhuus *et al.*, discusses users' concerns about the location-based services that would disclose their location and in turn user's privacy [102]. In this work, the authors have presented two types of location-based services, location-tracking and position-aware services. They

have shown a case study that examines user's concern for privacy in relation to location-based services and compared people's perceived usefulness of the two types of services. The work concludes that the concerns are more when third parties are tracking a user's location. Location based services with privacy as the main concern has been described in [12], [13], [103], [5], [6], [104] and [105]. In [12], authors have refined the mix zone model, describing a quantifiable metric of location privacy from the point of view of the attacker. In [106], the authors discuss the issues in the location-aware mobile devices in context by addressing the basic technology issues involved. They also discuss issues that are possible and not possible in the future. Further they outline privacy issues that arise from the conjunction of technical feasibility and government/marketplace activities that might use location information. In this work a representative sample of important issues is enumerated and discussed. Regulation is then discussed as a broader term covering the various entities and agencies that might structure and regulate the use of location information and provide the appropriate levels of privacy protection to constituents while promoting appropriate advances in new products and services. Other challenges such as user privacy are also important in ubiquitous environments. Privacy related efforts have been made in the past [3]. Research has been carried out around privacy awareness systems that allow certain privileges to data collectors [7]. Karyda and Gritzalis [4] listed some of the challenges in this area and the future research directions.

This section has provided a summary of the various issues that exist within location privacy aspects of mobile environments. Details show the possible attacks related to privacy and various mechanisms used to preserve end-user location privacy and access control. It gives a comparison of the different privacy preserving methods in mobile environments and their drawbacks to provide guidance to the readers. I also identify the open research challenges in the area of privacy and highlight the problems with the prior art.

Chapter 3

Privacy through access control and attestation

3.1 Overview

Preserving the location of the individual user of any fixed device has also been the concern for a number of researchers recently [103], [5] and [6]. Other challenges such as user privacy are also important in ubiquitous environments. Privacy related efforts have been made in the past [3]. Research has been carried out around privacy awareness systems that allow certain privileges to data collectors [7]. Karyda and Gritzalis [4] listed some of the challenges in this area and the research directions for the future.

Mobile devices have been used in the modern world to access information, exchange information and to store information. With new applications and new solutions coming to existence every day, handheld devices are being used more and more to completely take over the functionality of a wallet, laptop, computer, briefcase and books. With the increasing demand on the mobile devices usage, the number of threats and issues related to security with regards to the handheld devices are doubled. Information exchange is a key requirement for every sector and handheld devices are being used for the very same purpose more frequently than ever before. It is very important to address the question "How to enforce location privacy?" In the literature review section, I have shown various methods that have been used to ensure location privacy but with drawbacks. In this

chapter, I would like to show the methods that I propose to ensure that privacy is achieved.

I also address the issue around "How can one ensure that the data being accessed is being accessed by the right user of the data". How can one ensure that the data requestor is the genuine requestor of the data and has the right to access the data being requested. This is a very interesting problem which highlights the very existence of access control mechanisms. The process of making sure that data is accessed by the right source for the right requirement is called access control. Making use of access control in order to ensure that the right user is getting hold of the right data has always been a concern in all major areas of business. Using one of the usecases described in Introduction i.e. Healthcare data involves a lot of personal data of the end users and hence it is critical to ensure that the data is not misused and is not mishandled. It is equally important to ensure that data is being accessed by the right user and does not fall into the hands of a malicious user who would mishandle the data. Researchers in the past have introduced solutions around access control for healthcare with P3P and XACML. By making use of P3P and XACML policy, and through policy negotiation, relevant data is provided to the requestor based on the information provided in the P3P and XACML policy. The solution has taken care of the access control of the data but something very important has been forgotten in this solution. The data can be requested from anywhere and there is no way to stop an outsider from requesting the data. This introduces the importance of contextual information for the requestor of the data. I have considered the location information of the data requestor before granting access to the data. This has been achieved by making use of geospatial attributes of the handheld device from where the request is made. Once the geographical coordinates are verified, the policy is then applied on the request and then access to the data is eventually granted. This solution has been implemented by using GeoXACML policy. This is the first of a kind implementation of GeoXACML for access control in mobile healthcare.

As far as privacy is concerned, there is always a privacy Vs utility trade-off. This is very well known and in order to access sensitive information, one has to ensure a specific level of trust on the location. In order to achieve this, location attestation is the way forward and in this chapter, I will provide a new solution using global attestation of location. This chapter shows a detailed implementation of the solution that I designed through the various stages in my thesis. This chapter also shows a detailed evaluation of

the various mechanisms used to achieve privacy in mobile environments. It details the evaluations for policy based access control and global location attestation. It explains that scenario used to evaluate the setup and shows the results.

3.2 P3P extension for access control

Privacy of the information being transmitted is very important so is the location. The location information also needs to be protected. A number of researchers have been investigating these issues recently [5], [6], and [107]. P3P provides a flexible and powerful mechanism to extend its syntax and semantics using one element: EXTENSION. This element is used to indicate portions of the policy reference file which belong to an extension. The meaning of the data within the EXTENSION element is defined by the extension itself. < EXTENSION > For example, if www.catalog.example.com would like to add to P3P a feature to indicate that a certain set of data elements were only to be collected from users living in the United States, Canada, or Mexico, [28] it could add a mandatory extension similar to the one shown below:

< DATA-GROUP >

...

<EXTENSION optional="no" >

<COLLECTION-GEOGRAPHY type="include"

xmlns="http://www.catalog.example.com/P3P/region">

<USA/> <Canada/> <Mexico/ >

</COLLECTION-GEOGRAPHY>

</EXTENSION>

</DATA-GROUP>

There are some extensions that needs to be carried out to P3P in order to make it work with the handheld devices. The first step is to extend the P3P policy to be used within the security capsule. The second step is to implement the P3P policy within the

handheld device. The next step is to integrate the P3P policy in the sensors within the ubiquitous environment. The extensions that will be used in the P3P policy within the security capsule of the mobile device include the identities that are used in the capsule. These include IP Multimedia Private Identity (IMPI) number and International Mobile Equipment Identity (IMEI) number. IMPI is the mobile operator assigned identity for the mobile user. IMEI is the unique identity for the mobile device and this is issued by the mobile device manufacturer. These extensions will be embedded in the P3P Policy within the <EXTENSION> tag of the policy. Web sites can publish their policies in Websites, headers or in the source file format. For handheld devices, there needs to be a mechanism to publish the policies. In order to store the P3P policy in the mobile device, it is saved as an XML file and stored in the device

3.3 P3P extension application

The proposed architecture deals with the implementation of a policy in the mobile environment. P3P policy is extended in this approach in order to give access to the data requested by the mobile device from the service provider. For requesting sensitive information (i.e. healthcare data, financial data or personal data) from the service provider, the mobile device needs to establish trust with the service provider.

3.3.1 Architecture

This trust is established by the security capsule during the registration process with the identity and service providers. In response to the request from the mobile device, the service provider sends the sensitive data to the device in encrypted format. In order to have access to the data, the mobile client needs to request the service provider for the real time key. The service provider used the real time key to encrypt the sensitive data requested by the mobile device. The proposed architecture demonstrates the process involved in retrieving the real time key from the service provider. The architecture includes the following steps to access the data from the service provider. The first step starts with the mobile client requesting for a real time key from the service provider. As a response to this request the service provider will send a challenge request along with the policy of the service provider.

3.3.2 Protocol implementation

The security capsule in the mobile client will process the challenge request and sends back the challenge response along with the policy user preferences from the mobile client [27]. The policy file sent by the service provider in case of P3P will be parsed in the mobile client and the identity, which is in the extension data type of the parsed file, will be retrieved. This identity information, which is known to the mobile device, is then hashed and will be sent to the service provider. The identity provider and service provider initially decide on certain computational techniques and negotiate the computations to be performed on the hash. In the service provider the hash value undergoes these computations and the resulting value is used as a key to encrypt the real time key. The service provider will then encrypt the real time key with this new computed value and sends back the encrypted key to the mobile client. The mobile client, which already has the hash value of the identity, will perform the same negotiated computations decided between the identity and service provider. The computed value is used as the key to decrypt the encrypted key resulting in the real time key. The real time key is then used to decrypt the encrypted data. This process is depicted in Figure 3.1 below. The steps as shown in the figure are:

1. Mobile device makes the real time key request to the WebServices.
2. WebServices sends the challenge request and the policy and asks for user preferences.
3. Mobile client sends the challenge response and hash of the identity retrieved from the policy file.
4. WebServices encrypts the real time key with the modified hash (hash undergoes some computations decided during registration) as the encryption key and sends the real-time key response.
5. The encrypted real-time key is decrypted in the mobile client by using the modified hash as the decryption key.

The real time key is then used to decrypt the encrypted data. In case, if the identity information of P3P file do not exist in the mobile client, then the service provider will send a request asking for additional information from the mobile client. The mobile client needs to send responses to the request and eventually the service provider decides whether the information is sufficient in order to send the real time key. P3P policy is stored in a XML file format in the mobile client. Secondly, the service provider does not have a standard format for storing the P3P policy file. It is possible to publish the P3P

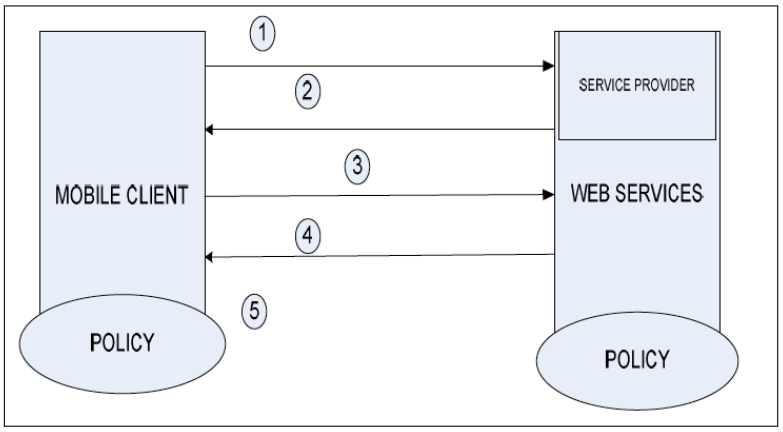

FIGURE 3.1: Steps involved in the proposed architecture for privacy.

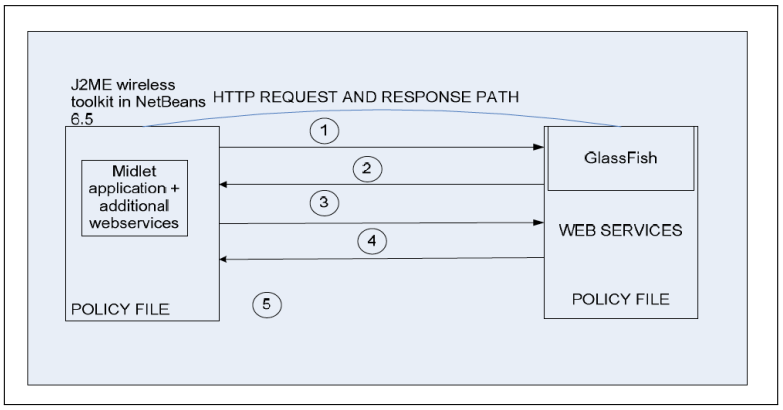

FIGURE 3.2: Steps involved in the proposed architecture for privacy.

policy in WSDL or by using Universal Description, Discovery and Integration (UDDI). P3P policy can be published in a WSDL file in a Web Services environment. In case of XACML policy, in response to the initial request, the service provider will send a challenge request and a request created by PEP for XACML policy from the mobile device. The mobile client will send the XACML policy with the relevant details in it. WebServices will then pass the request through the PDP which will look at the request and decide whether the request is eligible to be granted access to the information. Based on the decision made by the PDP, WebServices encrypts the real time key and sends it as a response to the mobile device. The key is then decrypted in the mobile device and the original information is retrieved. A proof of concept is developed to validate the proposed architecture. The prototype is shown in Figure 3.2 and is implemented using open source software tools and development kits. The business logic of the implementation is developed in Java 2 Platform Enterprise Edition (J2EE) environment and is deployed in GlassFish V2 Web Services. It uses the Java Development Kit (JDK) 1.6. The mobile client is developed using the Java 2 micro edition (J2ME) wireless toolkit

in NetBeans by creating a Midlet application. The toolkit includes Mobile Information Device Profile (MIDP) libraries. MIDP is a key element of J2ME. When combined with the Connected Limited Device Configuration (CLDC) MIDP provides a standard Java runtime environment for mobile phones and mainstream PDAs. All the communications between the mobile device and the Web Services are performed using the Simple Object Access Protocol (SOAP) protocol over HTTP. The main functions or methods are written in the Web Services and these methods are invoked accordingly. The Web Services is deployed and appropriate stubs are created for it. A WSDL is created for each of the Web Service methods and this is published in order to test and invoke the appropriate methods. WSDL is an XML based language for describing Web Services. Details are available in the specification document [108]. This WSDL document can also be used to perform quick tests on each Web Service method and to verify the results. SOAP monitor can be used to monitor all the traffic between the mobile client and the Web Services. A number of SOAP monitoring tools are available to do the same. The communications are secured using well-known Java crypto and security libraries. The prototype proves the implementation of secure mechanism for the access of data sent by the service provider. The model validates the mechanism of extending the P3P policy for successful data access control. Other possible implementation methods are described in [109].

3.4 Evaluation setup for policy based access control

A number of P3P policy requests have been created and sent to the server to check for the policy extension functionality in order to achieve access control. Various P3P policy requests were sent to check the use cases including various parameters such as personal information from the requestor, location details, query and other mandatory attributes to allow the policy file to pass the negotiation check.

3.4.1 Results for P3P access control

With the P3P policy extensions and with various use cases that allowed to check for the policy negotiation in order to achieve access control, I have proposed the architecture for controlling the data access from service provider into mobile devices. The main novelty

of this is the introduction of P3P policy into the mobile device environment. P3P policy has widely been used for web sites but has never been extended into the mobile area. This work has made an attempt to implement and prove the same. The P3P policy is extended in order to be used for validating the user preferences and for using the identity to be part of the key to the actual data. The proposed architecture is validated through the implementation of a prototype mobile testbed implementing the extended P3P polices. Since XACML is the latest technology for mobile environments and it overcomes a number of limitations of P3P policy with the help of the policy decision and policy enhancement capabilities. The XACML components PEP and PDP enhance the implementation capabilities and provide more control over the policies in the mobile environments. The XACML implementation is also being tested and is expected to see better results than P3P. The proposed architecture makes way for the ubiquitous environments to be safe in controlling access to any kind of sensitive information. With the introduction of P3P and XACML policies into the devices in ubiquitous environments, privacy can be preserved and security will not be compromised.

3.5 XACML based access control

In this section, I get into the details of using XACML for accessing the data in the mobile device. In a typical health services scenario, the authorised individual/doctor will request the patient's information from the central health repository. This data will be sent to the mobile device or handset of the doctor in an encrypted format. The data is encrypted using the real time key which is available at the central health repository also known as Service Provider in WebServices terminology. Although the doctor will not be able to read any content that is been sent, the information is still transferred to the doctor's handset. This work proposes an architecture that performs certain policy decision checks using XACML and computations in order to get to the real time key which is being used to encrypt the content sent to the device. By using the real time key, the data in the handset can then be decrypted and the doctor will be able to read the original contents of the patient's information he requested.

3.6 Proposed solution

The proposed architecture deals with the implementation of XACML in the mobile environment. As shown in Figure 3.3, XACML policy is used to give access to the data requested by the mobile device from the Central health repository (service provider). This work assumes that the initial trust between the mobile device and the Central health repository (service provider) is established.

3.6.1 Architecture

The message flow begins with the doctor requesting the patient's information from the Central health repository (service provider). When the service provider receives the request it will send the data by encrypting it using the real time key. The proposed architecture depicts the processes involved in retrieving the real time key and the steps used to access the data sent by the service provider. The first step in this architecture begins with the handset making a request to the central health repository requesting for a real time key. In response to this request, the service provider will send a challenge request and XACML policy with the key policy decision checks. On receiving these, the mobile device will parse through the XACML policy and provide a response in the form of a hash value of the values requested by the XACML request policy from the service provider. The XACML policy on the mobile device is stored as an XML file. The Central health repository receives the challenge response and the hash value. An initial set of secret computation techniques is decided and both the service provider and the mobile device's provider are aware of it. These computations are performed on the hash value and the resulting value is used as a key to encrypt the real time key on the service provide (Central health repository) side. This new encrypted key is then sent to the mobile device. Since the device is also aware of the same secret computation techniques, it will perform the same computations on the hash value which it already has. The resulting value is used to decrypt the encrypted real time key. The real time key is then used to decrypt the encrypted data.

The Figure 3.3 depicts the proposed architecture. The numbers shown in the figure refers to the different steps involved. They are briefed as follows:

1. Health care personnel(Doctor) requesting for the patient's information.

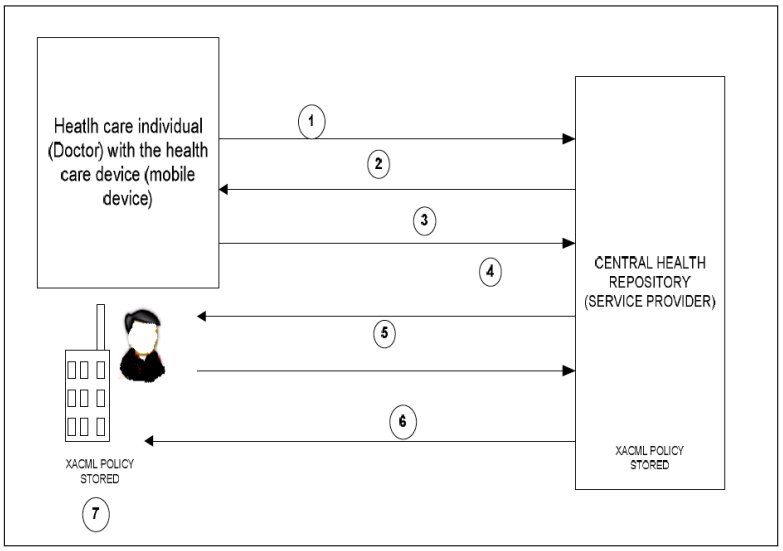

FIGURE 3.3: Proposed Architecture steps.

2. Encrypted patient information sent after encrypting it using real time key.

3. Health care mobile device requesting the real time key.

4. Challenge request and XACML policy request with policy decision checks sent to the device.

5. Challenge response and hash of the policy decision values returned to the service provider.

6. Real time key is encrypted using modified hash value ((hash undergoes some computations decided during registration) and is sent to the device.

7. Real time key is decrypted using the modified hash value. Real time key is used to decrypt the encrypted patient information.

3.6.2 New protocol

A proof of concept was developed to prove the proposed architecture. Mobile device is simulated using the Java 2 micro edition (J2ME) wireless kit in Netbeans by creating a Midlet application. The toolkit includes Mobile Information Device Profile (MIDP) libraries and JSR 172. MIDP is a key element of J2ME. When combined with the Connected Limited Device Configuration (CLDC) MIDP provides a standard Java runtime environment for mobile phones and mainstream PDAs. All the communications between the mobile device and the WebServices are performed using the Simple Object Access Protocol (SOAP) protocol over HTTP. The main functions or methods are written in the Web Services and these methods are invoked accordingly.

The Web Services is deployed and appropriate stubs are created for it. XACML implementation can be done by using Sun's XACML Open Source implementation or other implementation methods as described in [109]. A WSDL is created for each of the Web Service methods and this is published in order to test and invoke the appropriate methods. WSDL is an XML based language for describing Web Services. Details are available in the specification document [108]. This WSDL document can also be used to perform quick tests on each Web Service method and to verify the results. SOAP monitor can be used to monitor all the traffic between the mobile client and the Web Services. A number of SOAP monitoring tools are available to do the same. The communications are secured using well known Java crypto and security libraries.

3.6.3 Results for XACML based access control

In this work we propose an architecture for controlling the access to the data in the health services sector. The main novelty is the approach used to control privacy of the data using XACML within the mobile environment. With the implementation of XACML policy in the mobile environment, data access control can be preserved and privacy can be maintained. This implementation also mitigates man-in-the middle attacks and maintains privacy of the information.

3.7 Geospatial access control

The main contribution of this work is to restrict access to data from mobile devices based on geospatial attributes of the requester. XACML policy is applied to check the access rights before making a decision on an access request. However, it does not have operators and construct to handle geospatial attributes. There are several well-established open source XACML libraries, such as SUN's implementation[1]. In an XACML policy, several attributes of a requester are checked against their values to grant access. In this work, I propose to extend the existing XACML implementations with the ability of handling geospacial attributes efficiently. For this purpose, I introduced new functions and attributes that support geospatial functionality.

[1] http://sunxacml.sourceforge.net/

I have two main *attribute* extensions: GeoPoint and GeoPolygon. A GeoPoint attribute represents a particular point which stores a latitude and a longitude. For instance, position of an access requester is represented as a GeoPoint instance. A GeoPolygon instance represent a region or geospatial boundaries, e.g., a building, or a room. This can be achieved by storing multiple GeoPoints, which when combined create a polygon. The polygon generation can be done with polygonization provided by existing topology libraries such as JTS[2]. In addition to these attribute extensions, I implemented a function called *geo-contains*,which takes a GeoPoint and a GeoPolygon as arguments and checks to see if the point is contained by the polygon. The geopolygon can be defined in any shape and form, it can be a rectangle, square, circle or any other shape and form.

With these extensions, the requester may send its current location together with its access request using XACML request to a Policy Decision Point (PDP). The PDP checks geospatial policies to decide if it may allow the requested access or not. Once the access is allowed, Policy Enforcement Point (PEP) provides the requested resource. A simple example policy and access request are listed in Listing 3.1 and Listing 3.2 respectively. In the example policy, it is stated that doctors can access medical records within the geospatial boundaries defined through a geoPolygon object. In the request example, a doctor with name John request medical records for Alice. The PDP may examine the request and provide access based on the policy.

```
<Policy PolicyId="ExamplePolicy"...>
<Target><AttributeValue>/reports/*</AttributeValue></Target>
<Rule RuleId="ReadRule" Effect="Permit">
<Target>
<Subjects><AnySubject/></Subject><Resources>  <AnyRespurce/></Resources>
<Actions><Action>...
<AttributeValue DataType="...XMLSchema#string">read</AttributeValue>
 ...</Action></Actions>
</Target>
<Condition FunctionId="urn:oasis:names:tc:xacml:1.0:function:and">
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
<Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-one-and-only">
<SubjectAttributeDesignator AttributeId="doctor" DataType="...#string" />
</Apply>
```

[2]http://www.vividsolutions.com/JTS

```
<AttributeValue DataType="...XMLSchema#string">John</AttributeValue>
</Apply>
<Apply FunctionId="geo-contains">
<Apply FunctionId="geoPoint-one-and-only">
<SubjectAttributeDesignator AttributeId="location" DataType="geoPoint" />
</Apply>
<AttributeValue DataType="geoPolygon">
        41.027514,29.189435;
        41.029514,29.189435;
        41.029514,29.191435;
   ....
</AttributeValue>
</Apply>
</Condition>
</Rule>
</Policyl>
```

LISTING 3.1: A GeoSpatial policy example.

```
<Request>
<Subject>
<Attribute AttributeId="doctor"
  DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>Ihsan</AttributeValue>
</Attribute>
<Attribute AttributeId="patient"
  DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>Alice</AttributeValue>
</Attribute>
<Attribute AttributeId="location"
  DataType="geoPoint">
<AttributeValue>41.028514,29.190435</AttributeValue>
</Attribute>
</Subject>
<Resource>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
  DataType="http://www.w3.org/2001/XMLSchema#anyURI">
<AttributeValue>http://server.example.com/reports/alice</AttributeValue>
</Attribute>
```

```
</Resource>
<Action>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:action:action-id"
  DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>read</AttributeValue>
</Attribute>
</Action>
</Request>
```

LISTING 3.2: A GeoSpatial request example.

In the policy example, the geospatial boundaries are unnamed. In other words, I explicitly define the geospatial boundaries instead of addressing this region. This makes it harder to author policies, since the policy author may not correctly enter the geopoints composing the polygon. Authoring policies become an important issue as the number of policies. To handle this, I propose to store geoPolygon objects in a separate database and use alias to refer to them. For instance, the geoPolygon object used in the example policy is stored in a database table with alias *CentralHospital*. Then, the alias can be used while defining the policy, instead of explicitly using the geoPolygon object. Therefore, the policies could be more human friendly and less error prone.

The *geo-contains* function is responsible for checking if the location of the requester is within a region defined by geoPolygon object. This may be a costly operation if I represent the polygon as a set of geopoints. However, if I use *GeoHashing* [110] to represent geoPolygon objects instead of multiple geopoints, I can very efficiently perform the containment test. Geohash is a geocode system invented by Gustavo Niemeyer. For each geopoint, geohashing produces a code. By gradually removing characters from the end of the code will reduce the precision and the code would be representing a region instead of a single point. Therefore, geohash codes belong to nearby points may have similar prefixes. That is, the longer a shared prefix is, the closer the two points are. In other words, we can efficiently check if a point falls into a region by checking the prefixes. As a result, I propose to store geohashing prefixes for geoPolygon objects and exploit these while checking if requesters' location falls into predefined regions.

Figure 3.4 shows the components of our framework. Policies and requests are composed based on GeoXACML standards using the introduced techniques. In this framework,

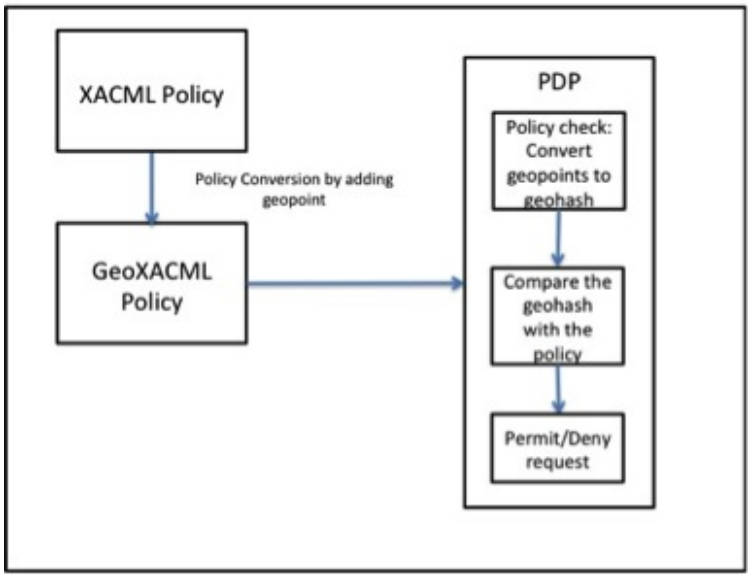

FIGURE 3.4: GeoSpatial Access control framework.

geospatial attributes are handled in preprocessing step in which a GeoXACML policy is converted into a regular XACML policy (with equality/prefix/range constraints on geohash) and during policy verification step the input geospatial attributes of requesters are converted to their geohashes and the regular XACML policy can be applied. The main advantage of this solution is that we can use an existing XACML engine to implement a geoXACML engine very easily and perform geospatial policy reasoning efficiently.

3.8 Geospatial access control for healthcare application

In this section, I will introduce a scenario from healthcare domain to evaluate and demonstrate our framework for geospatial access control in mobile devices.

3.8.1 Architecture

Medical records contains important information for diagnosing and curing various health problems. Authorised doctors with expertise may access medical information about their patients while fulfilling their responsibilities such as diagnosing, monitoring, and curing patients. However, these doctors may not access these records for a different purpose, since they may have sensitive data. There may be a relationship between purpose of access and the location of access requester. For instance, a doctor may be in the hospital

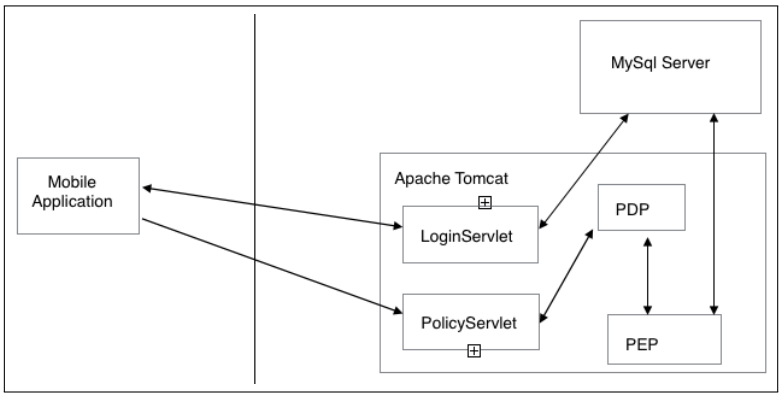

FIGURE 3.5: Architecture of the demo application.

while diagnosing or examining an outpatient. Therefore, she may request medical reports of the patient from the hospital, e.g., her office. It is less likely that the doctor will be using the requested information for the right purpose (i.e., diagnosis) if he is trying to access it outside of the hospital. Doctors may access their patients' records only when they are in the right location. In order to restrict the access based on attributes and location of the requester, I may use policies written in GeoXACML. An example policy and related request are listed in Listing 3.1 and Listing 3.2 respectively.

3.8.2 Protocol implementation

In order to demonstrate how our solution is deployed for this and similar problems, we implemented a mobile iOS application. Architecture of this demo application is shown in Figure 3.5. In the PDP, we store policies such as "Doctor's must be in the premises of their hospitals if they are required to access particular patients records". Then, we have extended SUN's XACML implementation as described in the previous section to implement GeoXACML policy decision and enforcement points. In order to store patient records, we have used a relational database at the back-end.

Figure 3.6 shows the polygon defined and a point within the polygon which refers to the location of the doctor making the request. Since the doctor's location is within the polygon, the doctor will be allowed to access the records. One of the snapshot from our demo application is shown in Figure 3.7. The figure shows the patient details that is visible to the doctor, once he has passed the policy check.

In our framework, the location information of the requester is sent along with other attributes in the request. In other words, the requester's device provides the location

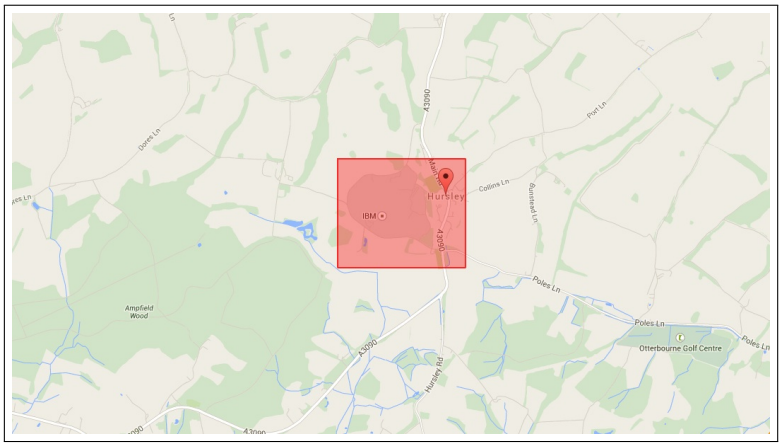

FIGURE 3.6: Geopolygon showing a point inside the polygon.

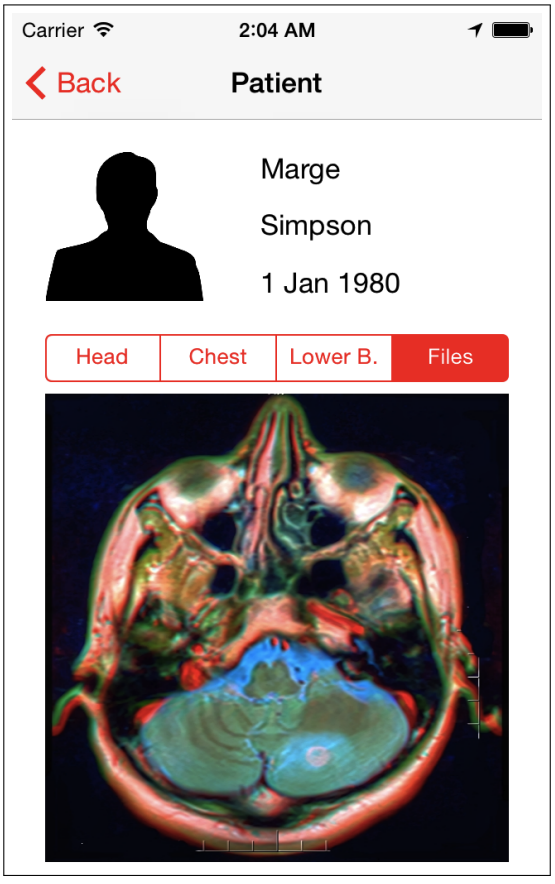

FIGURE 3.7: A screen shot from the mobile application.

information. A malicious requester can forge the location information to get access to a resource from an unauthorised location. Therefore, it is not wise to relay only on the location information provided by the requester while reasoning with geospatial policies. In other words, we may use one or more reliable location services to correctly determine true location or at least need to check the claimed location of the requester. For this purpose, we can use information from various sensors. If a device is bluetooth enabled, we can use bluetooth surveillance techniques to determine if the device of the requester is around. In bluetooth surveillance, the mac addresses of a bluetooth devices are registered and tracked using various bluetooth sensors [111]. Similarly, we can use information from wi-fi hotspots and GSM operators about the true location of requester's devices.

3.8.3 Results for GeoXACML based access control

With increasing volumes of data tagged with space and time (e.g., from smartphones) it is becoming increasingly important to support contextual (e.g., location-based) access control to data and resources. Through this research, I have explored solutions to realise the GeoXACML access control model that allows a security administrator to specify location-based access control policies. This work presents the first implementation and architecture for GeoXACML. The key novelty in the approach is the ability to use geohashes to translate a GeoXACML policy into a conventional XACML policy - this allows us to fully reuse existing implementations of the XACML engine. The research also describes a case study in the context of healthcare services wherein access control to handheld devices is moderated based on the location of the device. The code for the GeoXACML based access control is made publicly available [112].

3.9 Summary

This chapter has focused mainly on looking at the policy extensions for preserving privacy in mobile environments. It has detailed the methods of extension using P3P and XACML. The chapter also highlighted the new method I have introduced to achieve geospatial access control using XACML. This is a novel contribution to the research as there is a new way of using XACML engine with the geographical extensions and

geohashing to ensure that the information is received by the same person requesting for the information and that privacy is preserved. The chapter also introduces a new model for attestation of location which is a global attestation model and is the next step to the existing research in local attestation. Introducing global attestation through global consistency check is a novel contribution to the research community which has been possible through my work. This chapter has provided details on the implementation of the P3P and XACML extension for access control, GeoSpatial access control through GeoXACML showing the details on how the solution is novel. The chapter also detailed how global attestation of the location for mobile devices can be achieved using the trust matrix and calculations of EigenTrust. "How to enforce location privacy" has been addressed through this thesis and this chapter shows an implementation of this solution. This chapter has detailed the evaluation setup for policy based access control and highlighted the results of using 2 policy languages including P3P and XACML. It also shows the importance and benefits of using GeoXACML based access control model.

Chapter 4

Global attestation of location for mobile devices

4.1 Overview

As far as privacy is concerned, there is always a privacy Vs utility trade-off. This is very well known and in order to access sensitive information, one has to ensure a specific level of trust on the location. In order to achieve this, location attestation is the way forward and in this chapter, I will provide a new solution using global attestation of location. This chapter shows a detailed implementation of the solution that I designed through the various stages in my thesis. This chapter also shows a detailed evaluation of the various mechanisms used to achieve privacy in mobile environments. It details the evaluations for policy based access control and global location attestation. It explains that scenario used to evaluate the setup and shows the results.

4.2 Attestation of location

Location proofs are very important to ensure that the user providing the location co-ordinates are not misrepresenting the location. This could potentially be very important. Users provide their location information through location proofs to be able to access location based services. There are a number of solutions available in research that address the local attestation of the users. By making use of the Access Points (APs)

like WiFi or Bluetooth, the users can get the APs to give location proofs in order to access the services they require. Local attestation is a means where the user gets a location proof from it's local Access Point. Research shows that there are number of challenges and drawbacks in the local attestation scheme as the users can by various means cheat the system and get location proofs for the location where they are not located. Consider an example where a user from London is requesting for a service that is available only to users located in New York. In order to receive the service, the user requests his friend in New York to communicate with the corresponding AP in New York and then somehow asks the friend to send the communication channel to the user in London. By using local attestation, I can very well verify that the MAC address related to the request that came form New York to the corresponding AP was actually in New York. One of the mechanisms by which the user in London can use this service is to physically receive the WiFi card that was used in New York for accessing the service from London. Using local attestation, there is no way one can check if the card used in New York was reused in London in a short span of time. This is an open problem and is addressed extensively in this work. I introduce the concept of global attestation through consistency check where the user is checked for being located at different locations specified through various APs. Individual local attestation may be successful but when multiple local attestation are put together and scrutinised, the global attestation will show any fake activity at any stage in the pattern. I detail the challenges involved with the local attestation schemes that exist today and also explain how our solution using global attestation through consistency check is a better solution for providing location proofs in mobile environments. The future sections explains the global consistency check that is done based on the block chain methodology in Bitcoin. The block chain is a distributed global log of all the transactions. In this work, I use the block chaining model for tracking the details of the entities reporting the location information. Global trust values can be calculated using a number of methods. In this work, I have used EigenTrust and PeerTrust models to estimate the trust values.

4.2.1 System model

There are increasing number of mobile applications being used by end users to access all kinds of services. Location based services are mainly used for accessing location related capabilities. To ensure that the right user is receiving the location based services, the

requirement of location proofs have been commonly seen and understood. So by making use of the Access Points, the location proofs are provided by the APs in the form of a proof to ensure that the mobile device requesting a service is at a particular location at a particular point in time. This has been very well implemented and researched by various researchers including Saroiu et al. [74], Zhu et al [76] and through VeriPlace in [75]. When the user and the AP are colluding, it clearly indicates that the user is in the vicinity of the AP and hence it is convenient for the AP to attest for the user's location. Another example for location attestation could be where a user is in the vicinity of 3 APs nearby. Individually each of the APs can attest the user's location and the attestation could be verified. However, what are the consequences when a global check is done for capturing inconsistency. If all the 3 APs are put together and if a consistency check is done, it could either result in a positive feedback where all the points add up or it could also result in a negative feedback where all the points don't add up. This proves that one of the locations provided by the user is a false location. By doing a global check through a mechanism named global attestation, it is possible to check whether the user has been lieing all through the path. This brings us to a strong point that even if local attestation passes the check, does it mean that its correct. If it is correct, when multiple local attestations are done, will all the points add up? Our system model using global attestation scheme proves that a global consistency check is very crucial to prove the location proofs add up and that the user is not faking the location information through the entire journey of the request. Bitcoin has been using Block chain as the transaction database shared by all the nodes in a system [113] . This has been used as part of our system model for the global consistency check. A global log of the contacts is maintained in the database similar to block chain which is used to ensure that the locations reported by the entities themselves and other contacts in the proximity add up to result in a positive/negative feedback.

4.2.2 Global attestation scheme

In our model, each device (e.g., users and access points) provide reports about their locations. These devices register to our system with their Bluetooth or WiFi MAC addresses and each is given a unique ID. A location report from an entity x does not only contains its location, but also the MAC addresses sensed in the proximity. Therefore, if two devices are close in location, they may sense and report each others' MAC addresses.

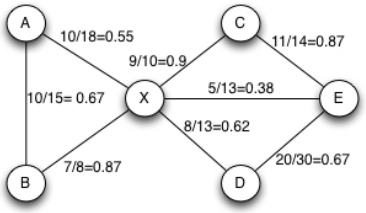

FIGURE 4.1: Trust Feedback Graph.

If all of the devices honestly report their locations and others they sensed, we may easily confirm the locations of these devices by cross checking the reports from different devices. However, these devices may not be honest, and even they may collude to mislead the system. Therefore, we do not assume the reliability of devices and compute their trustworthiness while reporting their and others' locations.

There are various statistical trust models. On the other hand, most of them require some sort of ground truth to come up with positive and negative evidence (or feedback) for the behavior of entities. In these models, each report from an entity is evaluated with respect to ground truth. The report serves as a positive feedback for the trustworthiness of the entity if it comply with the ground truth. Similarly, it serves as a negative feedback if it does not complies with the ground truth. While the computation of positive and negative feedback is trivial when ground truth is available, in our setting, I do not have ground truth for the location of devices.

I propose to use report consistency instead of ground truth to derive positive and negative feedback for the computation of trust. Our system uses a global log of location reports in the system. This global log can be implemented similar to blockchain [114], which is a transaction database shared by all nodes participating in Bitcoin protocol [113].

Let us consider two devices i and j that provide a number of location reports over time. At time t, let us assume that they share their reports R_i^t and R_j^t, which include the locations of these devices as l_i^t and l_j^t, respectively. If l_i^t and l_j^t are in proximity, report of each device may confirm the existence of other by including its MAC address. If l_i^t and l_j^t are not in proximity, report of each device may not include the MAC address of other. In these cases, the reports are considered consistent; otherwise, they are not.

I use consistent reports from pairs of devices as positive feedback for their trustworthiness. Similarly, inconsistent reports serve as negative feedback. If devices i and j

have n positive and m negative feedback, their overall positive feedback ratio is computed as $c_{i,j} = n/(n+m)$. Using positive and negative feedback, I compose a feedback graph where edges are weighted based on the computed overall positive feedback ratios. Figure 4.1 demonstrates a sample feedback graph.

Once a feedback graph is computed, I can use existing graph-based trust models to calculate trustworthiness of the nodes in the graph. While various trust models can be used, in our system, I use two specific graph-based trust models: EigenTrust [15, 115] and PeerTrust [16, 116]. These are models that compute global trust values for the nodes a graph based on their local trust values, e.g., edge weights.

EigenTrust [15] provides an efficient and robust method for computing global trust values. The calculation of the trust values are similar to the ranking calculations of the well-known page rank algorithm. It generates a matrix C whose each entry $C(i,j)$ corresponds to

$$\frac{c_{i,j}}{\sum_k c_{i,k}} \tag{4.1}$$

Then, the principal eigenvector of the feedback matrix C gives the global trust values for the nodes.

PeerTrust [16] computes a node's trust value based on the number of feedback and the credibility of feedback. The credibility of feedback for pairs of nodes may be measured by the personalised similarity of these nodes. In order to compute the similarity between two nodes i and j, I first create their feature vectors $\mathbf{f}_i$ and $\mathbf{f}_j$, respectively. The k^{th} element of $\mathbf{f}_i$ is set as $c_{i,x}$ – the weight of the edge between i and k in the feedback graph. Then, $cos(\mathbf{f}_i, \mathbf{f}_j)$ – the cosine distance between $\mathbf{f}_i$ and $\mathbf{f}_j$ – is computed as

$$cos(\mathbf{f}_i, \mathbf{f}_j) = \frac{\mathbf{f}_i \cdot \mathbf{f}_j}{||\mathbf{f}_i|| \times ||\mathbf{f}_j||} \tag{4.2}$$

and taken as the personalised similarity of the nodes. Details of the PeerTrust algorithm can be found in [16].

After calculating global trust scores for the entities, our location attestation algorithm considers these trust scores as follows. When an entity reports its location, I find a set

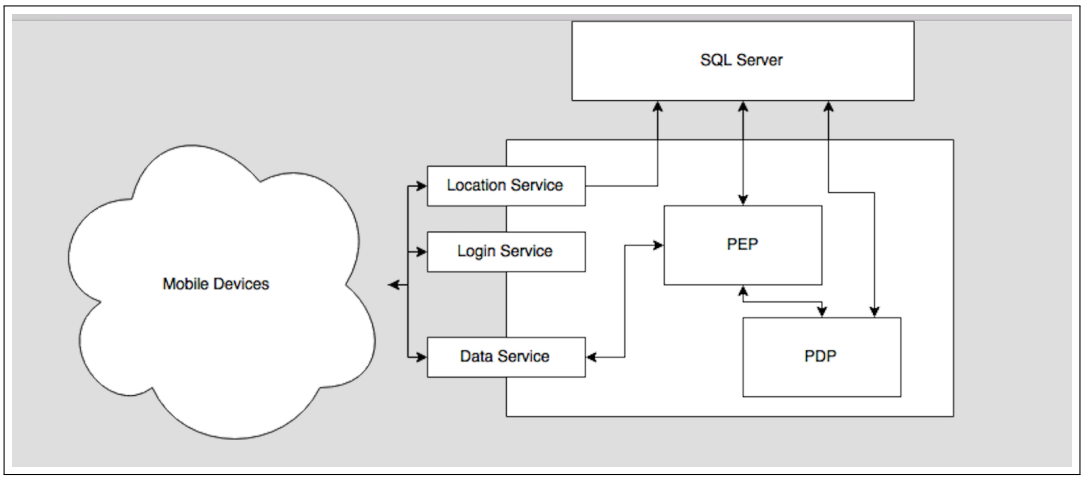

FIGURE 4.2: Demonstration application framework.

of positive reports (reports that concur with the location claimed by the entity) and negative reports (reports that claim that the entity was located elsewhere, or a lack of report from a trusted entity close to the claimed location). Then, I select the report from the most trusted entity (in this set) as the consensus report. If the consensus report agrees with the entity's claimed location then it is accepted; otherwise it is rejected.

4.3 Global location attestation application

I have implemented a mechanism to show the importance of global attestation of location and how it works in a mobile environment with a user requesting for information.

4.3.1 Architecture

My demonstration capability shows how the mobile device user requesting for certain information is being attested before getting access. Doctor requests the patient records. In order to gain access to this information, the doctor has to send the location information once logged in along with the names of the other devices in the vicinity seen through the Near field communications protocol. The server will then verify the location received and ensures that the other devices are also using the same access point as the requester. On verification, the access permission decision is made.

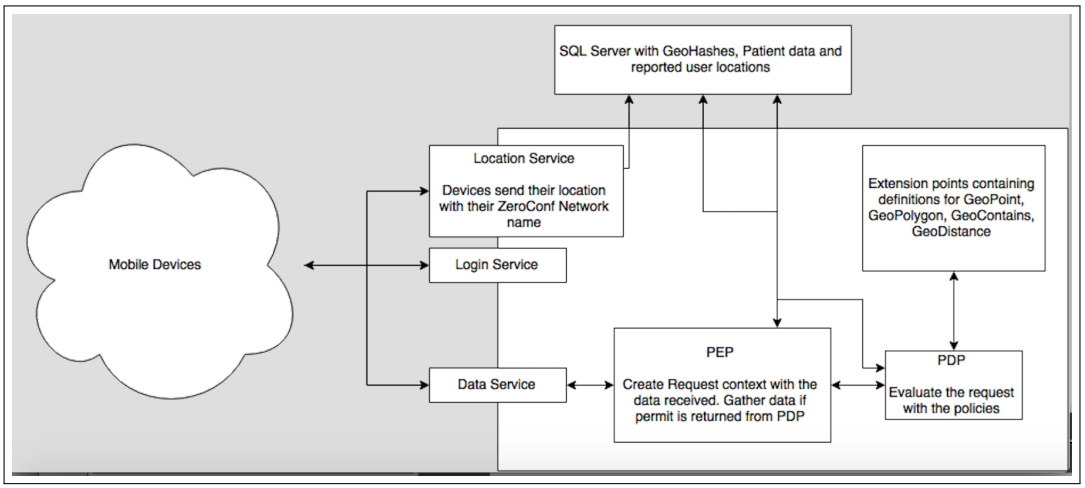

FIGURE 4.3: Workflow for the demo application.

4.3.2 Detailed design of the solution

Devices registered to the system once logged in will generate a name for the zero configuration network. The device will send the location and the name to the server and keep these information up to date. When the device requests for data it will send the following information for the creation of the Request Context that will be sent to the Policy Decision Point.

- Requesters ID
- Requesters location
- Names of the near by devices that are discovered using both wireless network and bluetooth.

Once the server receives the request it will run the request against the policies defined and generate a verdict. While generating a verdict the server will check for dishonest nodes.

4.4 Evaluation setup for global location attestation

This section details a scenario for evaluating the need for using Global location attestation and the benefits that our evaluation concluded through the results. The evaluation shows the results in a collusive and non-collusive setting with honest and dishonest users.

	San Francisco	**MIT Reality**	**Infocom06**
Num Entities	512	75	78
Location Source	GPS	WiFi	Bluetooth
Sampling Interval (secs)	30	300	120
Spatial Extent (km^2)	600	10	1
Temporal Extent (days)	30	30	4
Num contacts	28,241	18,665	182,951

TABLE 4.1: Datasets.

4.4.1 Scenario

I present an experimental evaluation of the proposed approach using three publicly available datasets [117], [118] and [119] (see Table 4.1). The `San Francisco taxicab` dataset includes GPS location traces from about 500 taxicabs over a period of 30 days. These cabs covered a spatial extent of about 600 km^2 around the city of San Francisco. Contacts between cabs were synthetically induced when two cabs happened to be within 600 meters of each other. When the spatial resolution of contacts was reduced below 200 meters, contacts were very infrequent and that over 1200 meters induced a large number of contacts; hence the choice of 600 meters for experimentation. The `MIT Reality Mining` dataset includes WiFi location traces from about 75 entities (typically personal laptop and handheld devices) over a period of 30 days. These entities covered a spatial extent of about 10 km^2. Contacts between two entities were naturally induced when two entities connected to the same WiFi access point. The `Infocom06` dataset includes Bluetooth contacts from about 78 entities (a subset of attendees of IEEE Infocom 2006 conference) over a period of 4 days. These entities covered a small indoor spatial extent of under 1 km^2. Contacts between two entities were naturally induced by pair-wise Bluetooth contacts. We remark that these datasets were intentionally chosen with varying degrees of contact density (number of contacts per entity per unit time). The taxicab trace from San Francisco has the least contact density, while the indoor Bluetooth contact trace from Infocom06 has the largest contact density.

In our experiments we emulate both honest and dishonest entity behavior. A honest entity faithfully reports contacts with other entities; e.g., in the taxicab data a honest taxicab would report contacts with all the other taxicabs that are within the chosen threshold distance of 600 meters. A dishonest entity can report both false positives and false negatives: a false positive report is one wherein a dishonest entity reports contact,

that which has not really occurred (e.g., a dishonest taxicab a claims that it is in contact with taxicab b when b is currently more than 600 meters from a); a false negative report is one where a dishonest entity fails to report a true contact (e.g., a dishonest taxicab a fails to report contact with taxicab b when b is within the threshold 600 meters from a).

In our experiments we also emulate both collusive and non-collusive settings for the dishonest entities. In a non-collusive setting a dishonest entity randomly chooses to induce a false positive or a false negative report. In a collusive setting a dishonest entity is more strategic: a dishonest entity would also concur with its colluder, i.e., if a dishonest taxicab a reports a contact with its colluder b, then b would also report a contact with a. Further, the choice of the location in the contact report between two colluding entities could be arbitrarily chosen (including a location that which neither a and b are currently located). In this setting dishonest entity would continue to randomly induce false positive and false negative reports against non-colluders (i.e., honest entities).

When two entities a and b concur (e.g., a reports a contact with b at location l at time t and b reports a contact with a at location l at time t) then the trust management system would treat this as a positive feedback between a and b. When two entities a and b fail to concur, the trust management system would treat this as a negative feedback. Indeed based on one instance of non-concurrence it is impossible to say whether a is dishonest or b is dishonest or both are dishonest (both being dishonest is feasible only under the non-collusive setting). We combine multiple positive and negative feedbacks to determine a trust score (between 0 and 1) in an entity using the EigenTrust and the PeerTrust algorithms.

Figures 4.4 and 4.5 show the error in trust score under non-collusive and collusive settings (respectively). Given the trust score estimate for an entity a ($\hat{ts_a}$) and the ground truth trust score ($ts_a = 0$ for dishonest entity and 1 for honest entity), the error in trust estimate is computed as the root mean square error in the estimate. I observe that in a non-collusive setting, then both the EigenTrust and the PeerTrust approach is very effective in estimating trust even when the fraction of dishonest entities is large (and hence do not offer corroborating evidences). Also, the localised approach to trust estimation helps PeerTrust outperform the EigenTrust solution that attempts to compute a global trust score for each entity. However, in a collusive setting both the EigenTrust is generally effective when the fraction of dishonest entities is small than 0.5. The

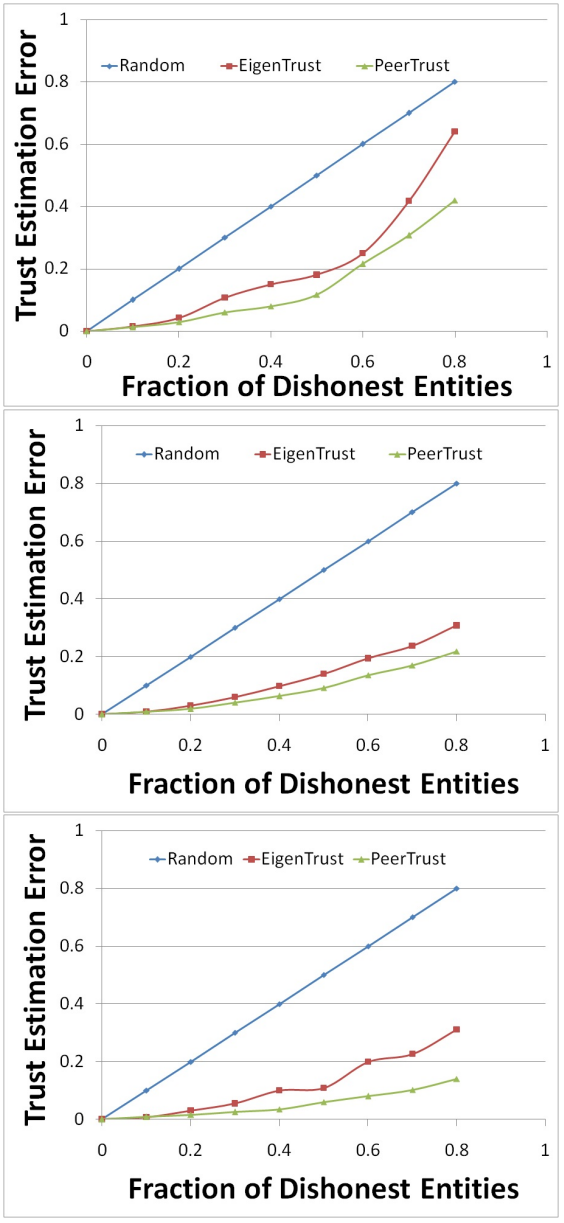

FIGURE 4.4: Trust Estimation Error in a Non-Collusive Setting: San Francisco, MIT Reality and Infocom06.

PeerTrust approach (again due to its localised trust estimation strategy) is relatively more robust. Nonetheless both the approaches are ineffective (compared to the baseline random strategy) when an overwhelmingly large fraction of entities are dishonest.

Figures 4.6 and 4.7 show the error in location attestation under non-collusive and collusive settings (respectively). Location attestation considers the trust scores of entities that report the target entity's location as described in the previous section. That is, I select the report from the most trusted entity (in this set) as the consensus report. If the consensus report agrees with the entity's claimed location then it is accepted; else

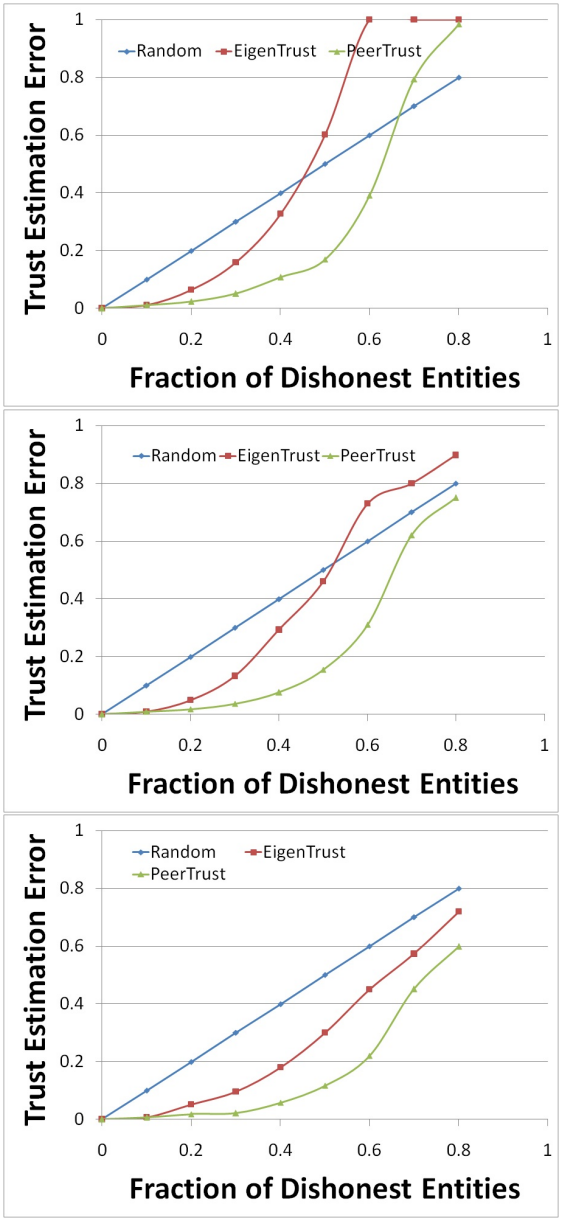

FIGURE 4.5: Trust Estimation Error in a Collusive Setting: (a) Using San Francisco dataset (b) MIT Reality (c) Infocom06.

rejected. I also check for absence of reports from highly trusted nodes that are in the vicinity (e.g., within 600 meters of the location claimed by a taxicab in San Francisco dataset). Location attestation error is captured as the root mean square error in accepting / rejecting a location claim, with "accept claim" assigned a numerical value of one and "reject claim" assigned a numerical value of zero.

I observe that under a non-collusive setting both the EigenTrust and the PeerTrust solution are effective even when a large fraction of entities are dishonest. Since location estimation relies on the entity with the highest trust score, under a collusive setting I

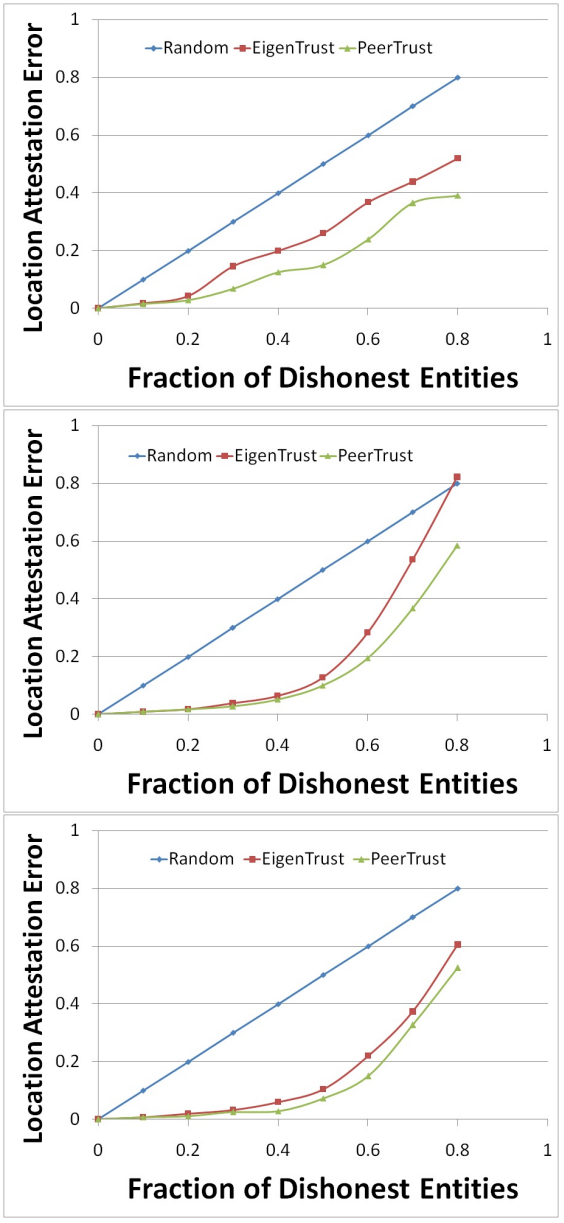

FIGURE 4.6: Location Attestation Error in a Non-Collusive Setting: (a) Using San Francisco dataset (b) MIT Reality (c) Infocom06.

observe that when an overwhelming fraction of entities are malicious, neither of the solutions are more effective than the baseline random strategy. However, for most practical settings wherein the fraction of dishonest entities is under 0.5, the EigenTrust and the PeerTrust solution offer a viable solution for robust location attestation.

4.4.2 Results for global attestation

A summary of key observations from our experiments are as follows:

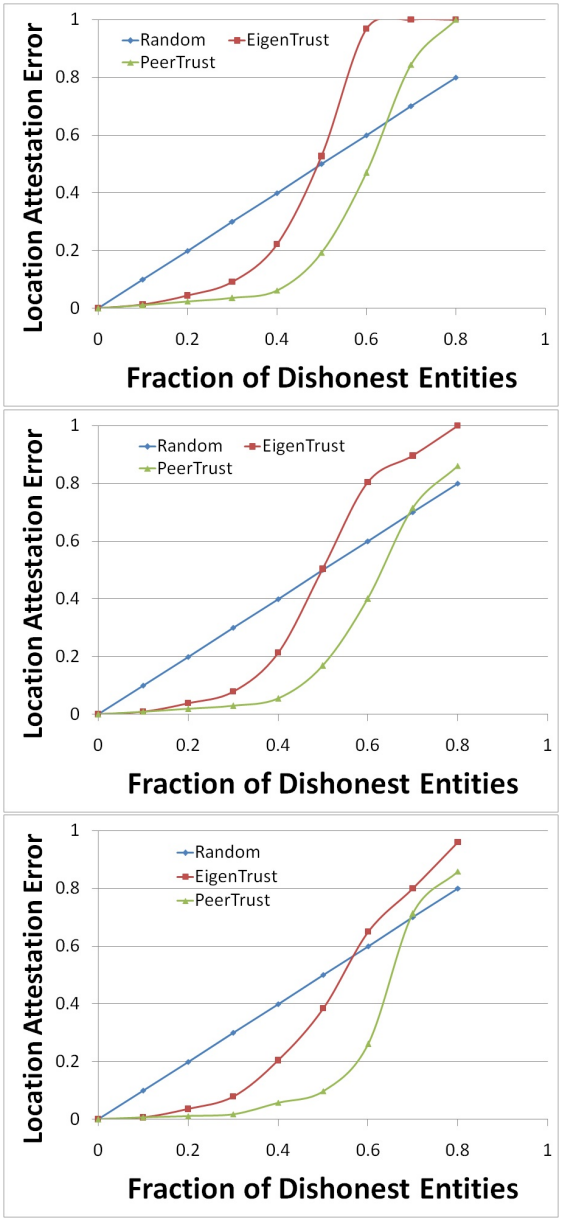

FIGURE 4.7: Location Attestation Error in a Collusive Setting: (a) Using San Francisco dataset (b) MIT Reality (c) Infocom06.

- In general, the trust estimation error and location attestation error are acceptably small ($< 15\%$) when I have fewer than one-third dishonest entities.

- In general as the contact density increases the error in trust estimation and location attestation decreases. Recall that the San Francisco dataset has the lowest contact density, while the Infocom06 dataset has the highest contact density.

- The PeerTrust approach of computing the trust score is more robust when I have a large fraction of dishonest nodes. The PeerTrust approach can handle more

dishonest nodes because it uses personalised trust scores for each entity, rather than a global trust score (as in EigenTrust).

4.5 Summary

This chapter details the steps used for evaluating global attestation. In order to show this, 3 publicly available datasets including SFO taxi cab dataset, MIT reality dataset and Infocom06 datasets have been used. The results show the errors in location attestation under collusive and non-collusive settings and includes outputs from both honest and dishonest users. In general, the trust estimation error and location attestation error are acceptably small ($< 15\%$) when I have fewer than one-third dishonest entities.

Chapter 5

Trust assessment in mobile environments

5.1 Overview

Fusing of location information is very crucial as information needs to be sent to the right resource for the right reasons and in order to achieve this, it is very important to evaluate and estimate some level of trust on the fused information. The more you trust an entity the more relaxed the privacy is when it comes to sharing information with that specific entity. Here is an example of why trust is very important in the overall context of privacy. On facebook, we have friends and we have friends of friends and rest of the world. We share more information with our friends as we trust our friends much more than we trust the others in the facebook. We share lesser information with friends of friends as we don't trust them so much. Further more, we would share even lesser information with the rest of the world as we don't trust them as much as we trust our friends and friends of friends. This is a classic example of the direct link between trust and privacy. In this chapter, trust has been shown with a specific use case and evaluation of trust value and its importance in the overall context of fusing information and resolving conflicts has been explained.

Humans or intelligent software agents are increasingly faced with the challenge of making decisions based on large volumes of streaming information from diverse sources. Decision makers must process the observed information by inferring additional information,

estimating its reliability and orienting it for decision-making. In this chapter, I propose a stream-reasoning framework that achieves all these goals. While information is streamed as unstructured reports (e.g., text in natural language) from unreliable sources, my framework first converts it into a structured form using Controlled English [82] and then it derives some facts that are useful for decision-making, and estimates the trust in these facts. Lastly, various facts are fused based on their trustworthiness. This process is totally undertaken on streaming information resulting in new facts being inferred from incoming information which immediately goes through trust assessment framework and trust is propagated to the inferred fact. In a crowdsourcing scenario there are multiple trust issues based on networking and connectivity. When the infrastructure is fixed, there are no trust issues as authorization is done through payment system. But in a mobile adhoc network, there are trust issues even for network connectivity. Further, if we consider the information layer, the real problem here is not whether it is fixed or mobile adhoc network. The problem here is information collected from multiple layers needs to be fused and trust is very important for this. Hence I came up with the idea of trust assessment in mobile environment.

In this chapter, I propose a comprehensive framework where unstructured reports are streamed from heterogeneous and potentially untrustworthy information sources. These reports are processed to extract valuable uncertain information, which is represented using Controlled Natural Language [82] and Subjective Logic [1]. Additional information is inferred using deduction and abduction operations over subjective opinions derived from the reports. Before fusing extracted and inferred opinions, the framework estimates trustworthiness of these opinions, detects conflicts between them, and resolve these conflicts by analysing evidence about the reliability of their sources. This chapter shows the detailed implementation of trust assessment framework. Further, evaluation setup for trust assessment is explained and the results of conflict resolution is described. The performance and accuracy achieved using the trust assessment framework is highlighted in this chapter.

5.2 Trust

A number of effective decision-making processes devised to support strategic operations and/or tactical missions and tasks in both military and civilian settings follow a recurring pattern from information gathering to action taking. This pattern is summarily represented by USAF Col. John Boyd's OODA loop [120]. The phrase OODA loop refers to the decision cycle of observe, orient, decide, and act, developed by military strategist and USAF Colonel John Boyd. It contains the following four phases (see Figure 5.1):

- The `Observe` phase comprises the processes of accumulating information pertinent to a situation of interest.
- The `Orient` phase comprises the processes of analyzing the information within the context of the particular task (operation or mission) at hand.
- The `Decide` phase comprises the processes of determining a proper course of action based on the knowledge gained during the `Orient` phase.
- The `Act` phase comprises the processes involving the actual (physical) actions taken.

The OODA phases continuously interact with each other, with, for example, the `Orient` phase *requesting* additional information from the `Observe` phase to drive better decisions. The interpretation of facts and information analysis in the `Orient` phase is heavily influenced by the existing (background) knowledge and understanding of the current situations at various levels including social, political, economic, cultural, administrative, logistic capabilities, resource availability, past experiences, and so on. In proposing the OODA loop, Col. Boyd had argued that in conflict situations, such as when acting to intercept an enemy airplane, in order to prevail one must execute his own OODA cycle at a pace faster than that of his opponent's OODA loop causing the opponent confusion by acting on stale information from situations that have already changed.

This chapter covers the first two phases that deal with the gathering, processing, and contextual placing of information to drive subsequent decisions and actions. It is motivated by the fact that decision makers (humans or software agents alike) are increasingly faced

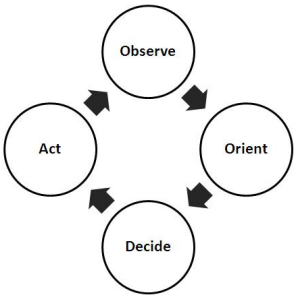

FIGURE 5.1: The OODA loop.

with the challenge of examining large volumes of information originating from heterogeneous sources requiring to ascertain opinion in various pieces of arriving information. Detection of conflicts between opinions and resolution of conflicts is a very important part of the solution. Resolving conflicts by analysing evidence about the reliability of the sources is shown very well in this chapter. The challenge is exacerbated when the decision makers employ uncertain rules to reason about the information collected from these heterogeneous sources. However, this is the case in most settings, since inference rules are uncertain and vague in many sensing domains and scenarios.

In typical applications of the OODA loop, the information sources feeding the `Observe` and supporting the `Decide` phases, such as the intelligence gathering scouts or the visual and electromagnetic sensors on a fighter airplane, are assumed to be administratively associated (if not physically owned and controlled) by the OODA beneficiary, e.g., the mission planners or the fighter pilot. In such cases, the capabilities of these sources are typically well known and documented and this background knowledge can be accounted for sufficiently well during the `Orient` phase. However, in coalition environments involving loosely coupled sensory systems [121] ownership, control and effectiveness of the sources cannot be guaranteed. The sensing resources may belong to different coalition partners or the informants may be locals. In such cases, the provenance and ultimately the *quality of the information* [122, 123] derived from these sources becomes questionable. Therefore, to properly support decision and action processes, the `Observe` and `Orient` phases must explicitly deal with uncertain information from sources whose capabilities are only partially (if at all) known and documented. In other words, the level of *trustworthiness* of the gathered information needs to be accounted for and properly weighed in the decisions and actions taken. There are various definitions of trust. McKnight and Chervany [30] defines trust as follows: Trust is the extent to which one party

is willing to depend on somebody, or something, in a given situation with a feeling of relative security, even though negative consequences are possible.

It is thus an objective of this work to design an opinion assessment framework that "scores" information arriving from uncertain sources and resolved conflicts. In supporting the `Observe` and `Orient` phases of the OODA loop, the framework will accommodate streaming information, acknowledging the need to deal with fast-paced information, such as that derived from sensors deployed across geographical spreads.

The main significance of this work shows the information being processed from streaming data, facts derived from it and new facts inferred from it. When new facts are inferred, trust is assessed for the new facts. So the process is executed on a continuous online mode. This is main difference between the previous works described in chapter 2 in this area and our work. In the offline mode, the facts are not processed immediately and the trust is not assessed instantaneously.

Another key observation from this chapter is that the work introduces key facts being derived from streaming data and the key facts lead to new facts being inferred. As the new facts are inferred it is passed through the trust assessment framework to propagate a trust value to the new inferred fact.

This solution is very applicable in social computing and social networks [124]. Social networking platforms like twitter can be used to collect information. In this case our information sources are users, information is received as unformatted text, which needs to be converted into Controlled English format. The formatted text would be used to infer new facts and a trust assessment framework is used to assess the trustworthiness of these facts. With more trustworthiness, the privacy can be relaxed as described in the facebook example scenario in the overview section of this chapter.

5.2.1 Use case and system highlights

As information accumulates and is read for consumption, its trustworthiness must be assessed for weighing properly in subsequent decision-making. Several authors have explored various opinion computation models on static data (e.g., eBay recommendation system [125], NetFlix movie ratings, EigenTrust [15], PeerTrust [16], etc.). A common data model subsumed by several trust computation models (as succinctly captured in

Kuter and Golbeck [126]) is the assignment of numeric trust scores between pairs of entities (e.g., in eBay recommendation buyers rate sellers, in Netflix movie ratings users rate movies, etc.). Such pair-wise numeric ratings contribute to a (dis)similarity score (e.g., based on the $\mathcal{L}_1$ or $\mathcal{L}_2$ norms, cosine distance, etc.) which is used to compute personalised trust scores (as in PeerTrust) or recursively propagated to compute global trust scores (as in EigenTrust).

There are two common assumptions in several such opinion models that are viable in commercial settings: (*i*) a statistically significant number of trust evidence (e.g., ratings) is available prior to opinion assessment, and (*ii*) assessed opinion values tend to vary slowly over time. In contrast, military settings warrant: (*i*) trust assessment over partial, uncertain and streaming (live and real-time) information from heterogeneous sources, and (*ii*) coping up with the dynamic and evolving nature of the ground truth.

Let us consider the following scenario comprising the events:

- *Event* e_0: An explosion occurs at location l_0 at time t_0. Event e_0 represents the ground truth and is assumed that there is no direct evidence available to support it in the opinion assessment system.
- *Event* e_1: An acoustic sensor reports a possible explosion at location l_1 at time t_1.
- *Event* e_2: A human agent reports unusual dust level at location l_2 at time t_2.

In this scenario, the goals are two-fold: (*i*) fact inference (e.g., infer an explosion from event e_2) and information fusion (e.g., fuse events e_1 and e_2 based on their spatio-temporal proximity), and (*ii*) assessing additional information (metadata) about the inferred/fused facts, such as an estimate of where and when ($\hat{l_0}$ and $\hat{t_0}$) the explosion took place, or an assessment of opinion in the fact that an explosion occurred based on previously seen events.

With reference to Figure 5.2, the first goal is addressed by the analytics component. In general, I expect to deal with a live (and potentially bursty) stream of events (e.g., textual or sensory reports) that are tagged with metadata (e.g., location, time, owner/authorship, beliefs). The stream of events may be subject to various analytics operations, e.g., fact extraction from event data using the Controlled English Store [82], event data fusion using the Information Fabric [83], etc. I do not make any specific assumptions on

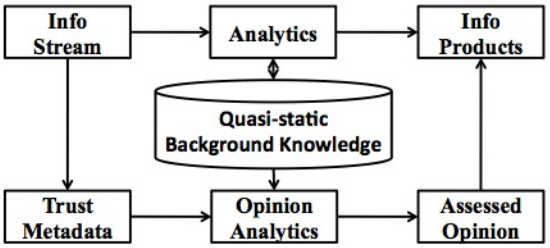

FIGURE 5.2: Opinion Assessment on Streaming Information.

the nature of such analytics operations, but for the fact that they are constrained to operate on streaming data. Hence, given latency constraints, storage, archival and post processing (or even batch processing) of event data may not be a viable option. I remark that this does not preclude the option of storing compact summaries (e.g., in a CE fact store [82]); only storing all (or a signification portion of) historical events is deemed infeasible. I also note that the goal of this work is not to support such analytics operations; instead the goal is to augment such operations with opinion metadata assessment.

While the internal workings of the analytics component are isolated from the opinion assessment system, I assume that the background knowledge is shared (e.g., rules for Bayesian inference, Controlled English fact inference rules, etc.). Intuitively, I model the analytics operation as accepting one or more inputs and a quasi-static background knowledge and generating one or more outputs. Here, the "quasi-static" qualifier means that the background knowledge is updated at a rate that is significantly slower than the event rate. This background knowledge is sometimes also referred to as the *domain knowledge*.

I model opinion assessment as a family of operators that accepts input opinion metadata and background knowledge and computes output opinion metadata. This model only requires that the background knowledge be shared between the analytics and the opinion assessment. Typically, such background knowledge is compact and has well defined semantics and isolates the opinion assessment system from understanding all the intricate details of the analytics operations (see Figure 5.2). Architecturally speaking, opinion is assessed as a part of a call-back function that is invoked each time the analytics operation generates a new output; the call-back function is supplied with the opinion metadata of the relevant inputs that were used to derive the said output.

Another interesting component of this work is the representation of rules with opinions. Past work assumed that the rules were always certain. Through this work, I have experimented the rules with opinions and certain uncertainty using the Subjective logic triples, i.e., *belief*, *disbelief* and *uncertainty*.

In this work I present a realization of this model using subjective logic triples (i.e., *belief, disbelief* and *uncertainty*) for representing opinion and Bayesian inference network for the background knowledge. In this setting, I have built a family of subjective logic operators that can be used to compute opinion metadata for an output event, given the opinion metadata on the input events and a Bayesian network that encodes probabilistic dependencies between one or more facts, e.g., Pr(dust|explosion) = 1, Pr(dust|¬explosion) = 0.25, and these probabilities may change over a time horizon that is significantly longer than the duration of an event burst.

5.2.2 Subjective Logic

Subjective logic (SL) is a type of probabilistic logic that explicitly takes uncertainty and belief ownership into account. In general, SL is suitable for modelling and analysing situations involving uncertainty and incomplete knowledge [1]. Arguments in SL are subjective opinions about propositions.

A binomial opinion of an agent A about the truth of a proposition x is represented by the quadruple $w_x^A = (b, d, u, a)$, where: b is the belief that x is true; d is the belief that x is false; u is the uncertainty (or, uncommitted belief) about x; and a is the base rate about x which represents the a priori probability about x in the absence of evidence; and $b + d + u = 1.0$ and b, d, u, $a \in [0, 1]$. Characteristics of a binary subjective opinion can be summarised as follows:

- $b = 1$ then it is equivalent to binary logic $TRUE$
- $d = 1$ then it is equivalent to binary logic $FALSE$
- $b + d = 1$ then it is equivalent to traditional probability
- $b + d < 1$ then it expresses some degrees of uncertainty
- $b + d = 0$ then it expresses total uncertainty

A binomial opinion can be represented as a Beta distribution. The probability expectation value of an opinion is defined as $E(w_x^A) = b + u \times a$, based on the corresponding Beta distribution [1]. Through the correspondence between binomial opinions and Beta distributions, SL provides an algebra for subjective opinions. Some of the operators defined over subjective opinions are listed in Table 5.1. Detailed descriptions and proofs for soundness of these operators are out of the scope, but they can be found elsewhere [1, 127–131].

TABLE 5.1: Some operators for binary opinions in Subjective Logic [1].

Operator Name	**Notation**
Complement [127]	$w_{\neg x} = \neg w_x$ $\neg(b, d, u, a) = (d, b, u, 1 - a)$
Discounting [127]	$w_x^{A:B} = w_B^A \otimes w_x^B$
Fusion [128, 131]	$w_x^{A \diamond B} = w_x^A \oplus w_x^B$ $w_x^{1 \diamond \ldots \diamond n} = \oplus\{w_x^1, \ldots, w_x^n\}$
Disjunction / OR [130]	$w_{x_1 \vee x_2} = w_{x_1} \sqcup w_{x_2}$ $w_{x_1 \vee \ldots \vee x_m} = \sqcup\{w_{x_1}, \ldots, w_{x_m}\}$
Conjunction / AND [130]	$w_{x \wedge y} = w_x \cdot w_y$ $w_{x_1 \wedge \ldots \wedge x_m} = \cdot\{w_{x_1}, \ldots, w_{x_m}\}$
Deduction [129]	$w_{y\|\|x} = w_x \circledcirc (w_{y\|x}, w_{y\|\neg x})$
Abduction [129]	$w_{y\bar{\|}x} = w_x \bar{\circledcirc} (w_{y\|x}, w_{y\|\neg x})$

Each opinion $w_x = (b, d, u, a)$ for a proposition x implies another opinion $w_{\neg x} = \neg w_x = (d, b, u, 1 - a)$. For instance, let x be the proposition "Dust storm in location X", the a priori probability for x be $\frac{1}{2}$, and the opinion of sensor s_1 about x be $w_x^{s_1} = (0.75, 0.1, 0.15, \frac{1}{2})$, then the sensor's opinion about "No dust storm in location X" would be $w_{\neg x}^{s_1} = (0.1, 0.75, 0.15, \frac{1}{2})$. If another sensor s_2 has another opinion $w_x^{s_2} = (0, 0.7, 0.3, \frac{1}{2})$, these two opinions are fused into a single opinion using the consensus (a.k.a., cumulative fusion) operator $\oplus$ [131]: $(0.75, 0.1, 0.15, \frac{1}{2}) \oplus (0, 0.7, 0.3, \frac{1}{2}) = (0.55, 0.34, 0.11, \frac{1}{2})$. The consensus operator is defined as follows in definition 4.1, [128].

Definition 5.1. Let $w_x^A = (b_x^A, d_x^A, u_x^A, a_x^A)$ and $w_x^B = (b_x^B, d_x^B, u_x^B, a_x^B)$ be opinions respectively held by agents A and B about the proposition x. When $u_x^A, u_x^B \to 0$, the relative dogmatism between w_x^A and w_x^B is defined by $\gamma = u_x^B / u_x^A$. The fused opinion $w_x^{A \diamond B} = (b_x^{A \diamond B}, d_x^{A \diamond B}, u_x^{A \diamond B}, a_x^{A \diamond B})$ can be computed as follows, where $\kappa = u_x^A + u_x^B - u_x^A u_x^B$:

for $\kappa \neq 0$	for $\kappa = 0$
1. $b_x^{A \diamond B} = (b_x^A u_x^B + b_x^B u_x^A)/\kappa$	1. $b_x^{A \diamond B} = (\gamma b_x^A + b_x^B)/(\gamma + 1)$
2. $d_x^{A \diamond B} = (d_x^A u_x^B + d_x^B u_x^A)/\kappa$	2. $d_x^{A \diamond B} = (\gamma d_x^A + d_x^B)/(\gamma + 1)$
3. $u_x^{A \diamond B} = (u_x^A u_x^B)/\kappa$	3. $u_x^{A \diamond B} = 0$
4. $a_x^{A \diamond B} = (a_x^A u_x^B + a_x^B u_x^A - (a_x^A + a_x^B) u_x^A u_x^B)/\kappa$	4. $a_x^{A \diamond B} = (\gamma a_x^A + a_x^B)/(\gamma + 1)$

Discounting operator $\otimes$ allows normalisation of opinions based on the trustworthiness of their owners [127]. In order to allow different trust models in our work easily, I adopt a slightly modified version of the discounting operator. Let us assume that I get the opinion (b_x^i, d_x^i, u_x^i, a) from source i about the fact x and have t_i as the trustworthiness of the source. The resulting opinion normalised based on opinion is computed as

$$(b_x^i \times t_i, d_x^i \times t_i, u_x^i + (1 - t_i) \times (b_x^i + d_x^i), a)$$

where the idea is to move a percentage of belief and disbelief to uncertainty based on the level of trust in the source. The trustworthiness of information sources can be calculated using different models. For example, beta reputation system [87] uses beta probability density functions [87] to model trust. A Beta distribution has two parameters (r_i+1, s_i+1), where r_i is the amount of positive evidence and s_i is the amount of negative evidence for the trustworthiness of the source i. The degree of trust t_i is then computed as the expectation value of the Beta distribution:

$$t_i^j = \frac{r_i^j + 1}{r_i^j + s_i^j + 2} \tag{5.1}$$

While this computation is simple and straight forward, there exist more complicated approaches for modelling trust. Our work is flexible enough to integrate any of such model as long as these models provide a trust value in range $[0, 1]$.

Now, I give a concrete example of discounting operation. Let the trustworthiness of s_1 and s_2 be 0.1 and 0.8, respectively. Then, normalised opinions of s_1 and s_2 would be $w_x^{s_1} \otimes 0.1 = (0.07, 0.01, 0.92, \frac{1}{2})$ and $w_x^{s_2} \otimes 0.8 = (0.0, 0.56, 0.44, \frac{1}{2})$, respectively. In this case, fusion of the normalised opinions would be $(0.03, 0.54, 0.43, \frac{1}{2})$. I assume that *all opinions are normalised based on trustworthiness of their owners* (i.e., sources).

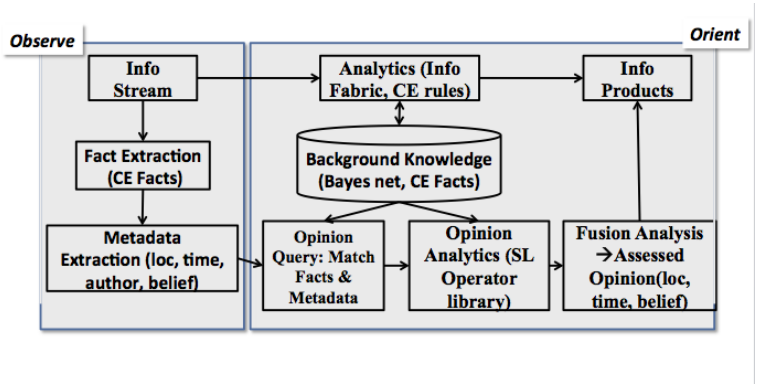

FIGURE 5.3: Opinion Assessment Framework.

Given opinions about different propositions (e.g., w_{x_1} and w_{x_2}), SL provides conjunction ($\cdot$) and disjunction ($\sqcup$) operators to compute opinions for the conjunction (e.g., $w_{x \wedge y}$) and disjunction (e.g., $w_{x_1 \vee x_2}$) of the propositions [130]. In addition to these operators, deduction operator ($\circledcirc$) is used to deduce new opinions from existing ones [129]. For instance, let x be the proposition "Country A is buying missiles" and y be "Country A is preparing to attack". Let us assume that I have the following conditional opinions about "y given x" and "y given $\neg x$".

- $w_{y|x} = (0.9, 0.0, 0.1, \frac{1}{2})$
- $w_{y|\neg x} = (0.0, 0.0, 1.0, \frac{1}{2})$

Given an opinion about x, I can deduce an opinion for y using $w_{y||x} = w_x \circledcirc (w_{y|x}, w_{y|\neg x})$ as described in [129]. If the opinion about x is $w_x = (0.75, 0.25, 0.0, \frac{1}{2})$, the deduced opinion about y is computed as $w_{y||x} = (0.68, 0, 0.32, \frac{1}{2})$ based on the algorithm in [129]. Details of the deduction operator is not in the scope, but it can be found in [129].

5.3 Opinion assessment framework

Figure 5.3 shows the architecture of our opinion assessment framework; wherever appropriate, I indicate ITA assets that could be leveraged in implementing the framework. As described earlier, I assume information analytics operates over a stream of input information and generates information products e.g., by fusing information, by extracting facts and using inferences to derive new facts, etc. Similarly to Figure 5.2, in Figure 5.3

I also show those parts of the architecture that relate to the `Observe` and `Orient` phases of the OODA loop.

Our opinion assessment framework operates in parallel with the information-processing pipeline. The opinion assessment framework extracts facts and opinion metadata from the stream of input information. The background knowledge store is queried for identifying relevant facts, i.e., facts whose opinion level may influence or may be influenced by the opinion level of the input information. For example, the input information may indicate unusual dust levels and background knowledge may encode the fact that unusual dust levels are often a result of an explosion; hence, the fact store may be queried for potential explosion information. Also, the background knowledge may encode the fact that explosion may be identified using its acoustic signature; hence, the fact store may be queried for acoustic information from a relevant space-time region (volume, to be more accurate) that surrounds the unusual dust level.

Once relevant facts are extracted, opinion in input information and information products is assessed using the family of opinion operators built using subjective logic. The choice of the opinion operator depends upon the nature of analytics operation (e.g., conjunction vs. deduction vs. consensus). In this section I present three examples, which illustrate the functionality of various components in our opinion assessment framework.

One of our key design features is that of decoupling the analytics component with the opinion assessment system. The analytics component is responsible for extracting and inferring facts (e.g., from unstructured text or other primitive facts). In doing so the analytics component may leverage (e.g., use a static road/terrain map to enrich facts) and update (e.g., Controlled English (CE) [82] rules to infer new facts) background knowledge. The opinion assessment system, on the other hand, does not infer new facts (as indicated by the unidirectional flow of information from background knowledge to the opinion assessment system in the figure). Instead it augments the inference task in the analytics component with opinion assessment capability, i.e., assesses opinion on the inferred fact.

In the rest of this section, I describe each of these steps in detail.

```
WISREP

Report number:      WIS/AF/008

Report date:        14th February 2010

Event date and time:    Disabling a potential suicide bomber wearing a vest

Event location:     Khoshal Village, Nad-e Ali District

Target

The target appears to have been a foot patrol operating in KHOSHAL
village. The suicide bomber, a young man  with an age
thought to be around 15-16 years, was wearing the vest and had been sent
from a terrorist stronghold to attack the foot patrol when a dicking network
had reported the patrol had entered the village market area. He was
riding a motorcycle which was also loaded with explosives disguised as
clothing.

    Device 1
```

FIGURE 5.4: A Snippet of an Unstructured Report.

5.3.1 Fact extraction

The input to our system is typically a stream of unstructured reports. Each report is subjected to fact extraction, resulting in a set of facts that are subsequently used by both the analytics and the opinion assessment components. The extracted facts and texts may be stored in Controlled English (CE) format in a CE Store. Here is a snippet of the intelligence report.

For example, given an intelligence report the following facts may be extracted[1]:

- `there is` a group named ent1_wisrep08 `that has` 'a foot patrol' `as` description
- `there is` an entity reference named er1_wisrep08 `that has` 'a foot patrol' `as` description `and`
  - `has` the document wisrep08 `as` source `and`
  - `has` the group ent1_wisrep08 `as` entity `and`
  - `has` 40 `as` start position `and`
  - `has` 53 `as` end position

5.3.2 Metadata extraction

Beyond fact extraction, I also extract various metadata attributes from a report including, Creator, Keywords, Opinion, Subject, Title, Security Level, Company, Category, Document parts, etc. For example,

[1] I use `teletype` lettering to emphasise CE specific word patterns.

- Source = “Dr Dave Sloggett”
- `has` ‘Opinion’ `as` (0.7, 0.2, 0.1, 0.5)
- Keywords = “IED, Pathfinder”
- Subject = “Daisy Chain Device”
- Title = “WISREP 001”
- Security Level = 0
- msole:codepage = 1252
- Company = “someCompanyName”
- Category = “WISREP”
- Document Parts = [(0, “WISREP 001”)]

Metadata attributes thus generated are also asserted as facts, thereby, permitting a uniform representation for both data and metadata. For example,

Document ‘20110228_WISREP_Report_001_ISAF’

- `has` the ms-word metadata ‘20110228_WISREP_-Report_001_ISAF.doc.meta.txt’ `as` metadata `and`
- `has` ‘ISAF’ `as` filename security level `and`
- `has` ‘001’ `as` filename subtitle `and`
- `has` ‘WISREP’ `as` filename report type `and`
- `has` ‘20110228’ `as` filename date `and`
- `has` ‘Report’ `as` filename title `and`
- `has` ‘20110228_WISREP_Report_001_ISAF.doc.plain.txt’ `as` plain filename `and`
- `has` ‘Wed Jun 8 13:24:33 2011’ `as` time processed `and`
- `has` ‘PFIII document processor’ `as` agent `and`

It is important to emphasise the following about the reports, facts, and opinions. Within the same report, there can be multiple facts written by multiple sources. When a fact is extracted from a report, the meta-data includes the source and opinion asserted by the source about this fact. Before asserting the opinion from the specific source into the knowledge base, I discount it with the trustworthiness of the source. If the opinion is not asserted by the source in the report, I take it as $(1, 0, 0)$ by default and discounted based on opinion. For instance, if the trustworthiness of the source is $(0.5, 0.1, 0.4)$, the opinion about the fact is taken as $(0.5, 0, 0.5)$ after discounting. Normally, the fact may appear in different reports. Therefore, there may be many opinions about the facts in our knowledge base. Some of these facts may be in conflict. For instance, I may derive opinions $(0.8, 0, 0.2)$ and $(0.1, 0.8, 0.1)$ for the same fact "there is a problem in the road". The first opinion indicates that the fact is probably true while other implies that the fact is not true. I will later describe how to handle these conflicting opinions about the same facts.

5.3.3 Opinion query

Each opinion is associated with a proposition, which may be a fact or meta-data. For example, the knowledge base stores the opinion $(0.9, 0, 0.1)$ from USGS about the fact "seismic activity has happened at location X on 20^{th} June 2014 at time 16:01" and the opinion $(0.8, 0, 0.2)$ for the proposition "USGS is a reliable source for seismic data". In this section, given a specific fact extracted from a report, I describe how to identify relevant facts (and meta-data), and retrieve opinions related to these. In particular, I extract three types of relevant facts and retrieve associated opinions from the knowledge base:

- The facts that exhibit causal relationships with the input fact (e.g., dust level, acoustic sensor measurements with an IED explosion)
- The facts that exhibit spatio-temporal similarity (e.g., two explosion reports whose space-time tags are relatively close to each other)
- The facts with matching meta-data attributes such as creator, keywords, etc.

After retrieving all the known opinions about the relevant facts and meta-data, I perform opinion analysis as described in the next section.

5.3.4 Opinion analytics

The opinion analytics component is designed to operate in parallel with the main analytics component; one could view our opinion analytics component as being triggered by the main analytics component whenever new facts are inferred or updated by the analytics component. The trigger would include as references the input and output facts from the analytics component and the relevant background knowledge queried in the previous step (Opinion Query step). The opinion analytics component uses a subjective logic (SL) operator library for deducing new opinions from the existing opinions about known facts. In this section I describe three sample opinion analytics operations.

5.3.5 Forward chaining (input opinions are known)

Consider two known facts that are annotated with metadata such as source (e.g., coalition member) C, location L and time T and an opinion tv. Given two annotated facts $(f_1, c_1, l_1, t_1, tv_1)$ and $(f_2, c_2, l_2, t_2, tv_2)$, the analytics operation may infer and generate a new opinion about another fact f_3 using deduction as: f_1 AND $f_2 \Rightarrow f_3$. The goal of the opinion assessment system in this case is to infer the opinion in the fact f_3. In order to do this, the opinion analytics system would identify facts f_1 and f_2 and the inference rule used to generate fact f_3, namely, f_1 AND $f_2 \Rightarrow f_3$. In this example, the conjunction of the opinions about f_1 and f_2 are computed using the conjunction operator of SL, which is described in Definition 5.2. The resulting opinion is the opinion for the left hand side of the rule, i.e., f_1ANDf_2. Based on the computed opinion and the deduction operator $\circledcirc$ mentioned in Section 4.2.2 (Subjective Logic), I compute the opinion tv_3 associated with the fact f_3.

Definition 5.2. Let $w_x = (b_x, d_x, u_x)$ and $w_y = (b_y, d_y, u_y)$ be opinions about two distinct propositions x and y, respectively. Then the opinion about $x \wedge y$ is computed using the conjunction operator as $w_{x \wedge y} = w_x \cdot w_y = (b_{x \wedge y}, d_{x \wedge y}, u_{x \wedge y})$ where,

1. $b_{x \wedge y} = b_x b_y$
2. $d_{x \wedge y} = d_x + d_y - d_x d_y$
3. $u_{x \wedge y} = b_x u_y + b_y u_x + u_x u_y$

Deduction operator of SL allows us to associate uncertainty with rules that are used to infer new facts. In our example, f_1 AND $f_2 \Rightarrow f_3$ may not be certain. That is, f_3 may not always follow from f_1 AND f_2. Let shortly refer to f_1 AND f_2 as θ. The deduction operator uses two conditional opinions about the rule $\theta \Rightarrow f_3$ while deducing opinions for f_3: i) $w_{f_3|\theta}$ and ii) $w_{f_3|\neg\theta}$.

The first conditional opinion $w_{f_3|\theta}$ refers to the belief, disbelief, and uncertainty for the proposition that f_3 follows from θ. It is computed using the evidence derived from observations. Given θ holds, if it is observed that f_3 holds as well, this observation is counted as a positive evidence while computing $w_{f_3|\theta}$; but if f_3 does not hold, the observation is taken as a negative evidence. Let r and s refer to the number of positive and negative evidence derived from observations. Then, following [129], the belief and disbelief in the opinion $w_{f_3|\theta}$ are computed as $r/(r+s+2)$ and $s/(r+s+2)$, respectively.

The second opinion $w_{f_3|\neg\theta}$ refers to the belief, disbelief, and uncertainty for the proposition that f_3 follows from $\neg\theta$. This opinion is also computed using the evidence derived from observations, as described above. If we set $w_{f_3|\theta}$ as $(0.8, 0, 0.2)$, this means that most of the time f_3 follows from θ, but sometimes this may not hold.

While describing rules, we can use arbitrary propositional logic statements (AND , OR , and NOT). Then, these rules can be used to infer a new fact, given a set of existing facts. For clarity, I have only shown conjunction operator, other SL operators can be found elsewhere.

5.3.6 Backward chaining (input opinions are unknown)

We can also infer new opinions about facts using backward chaining, i.e., through abductive reasoning. For instance, let us assume that we have a SL rule indicating that an explosion may lead to high level of dust. When we have a report asserting some opinions about high level of dust, we can use abductive reasoning or backward chaining to infer there might be an explosion resulting in the dust.

More formally, let us assume that f_1 and f_2 refer to facts representing *unusual dust level* and *explosion* at location l and time t, respectively. In our knowledge base, we may have a SL rule relating these facts, i.e., $f_2 \rightarrow f_1$ with conditional opinions $w_{f_1|f_2}$ and $w_{f_1|\neg f_2}$. The conditional opinions can be considered similar to conditional probabilities in

a Bayesian network, e.g., Pr(dust|explosion) and Pr(dust|¬explosion). Given an opinion w_1 retrieved from a specific source about f_1, we can infer another opinion w_2 using abduction operator $\bar{\otimes}$ of SL [132].

5.4 Spatio-temporal relevance and reasoning with location

Consider two facts f_1 and f_2 that assert the same statement (e.g., an explosion occurred) but with a small mismatch in its metadata. For instance, the location metadata l_1 associated with fact f_1 is slightly different to l_2 that is associated with fact f_2. Since the facts essentially agree with each other, the goal of opinion assessment is two-fold: (i) combining opinions in both the facts (because we now have collaborating evidence) and reduce uncertainty, and (ii) assuming that the location l_1 is not same as l_2, refine the location metadata associated with the facts. For sufficiently small deviations, we may simplify and assume that $l_1 = l_2$ in which case we can apply the SL consensus operator for fusing the opinions tv_1 and tv_2 to compute tv_3. However, in general, when $l_1 \neq l_2$, traditional SL cannot be used directly. The authors are exploring spatial-aware extensions of SL (coined `saSL`) to address cases like this, with the objective to compute a new location l_3 where the fact (e.g., explosion) could have occurred with high belief. Early results from this work and its application to a localization scenario for cognitive radio networks can be found in [133].

Location information in a report can be in the form of latitude and longitude values. In this case, the comparisons of locations in terms of proximity can be done using a set of geographical calculations. On the other hand, it may be more likely to have address information (e.g., street name, area, city etc.) in reports to indicate a specific location. Therefore, in order to identify proximity or spatio-temporal relevance between locations, we may need to do some reasoning. For this purpose, we can use an ontology that determines the relationships between different locations. The linked geographical data[2] can be used to determine if an area is within or close to another area. Once we determine the relevance between locations appearing in different facts, we can determine if the opinions about these facts are supporting one another or not.

[2]http://linkedgeodata.org

Let us describe this using the following example. We consider that a sniper has killed someone in an area. There are two reports about the sniper's location. The first report is from local police and indicates with high belief that "the sniper shot from a building x on street y". The second report is from a military force conducting operations in the region. The military force also knows that the sniper shot from the building x. However, it does not want to reveal the positions and density of its sensors on the region. Hence, it shares a report after obfuscating (e.g., generalising) the location information. For instance, if the street y is in the area z, it reports that "the sniper shot from a building in area z", being sure that this information does not reveal that it has sensors around the building x. When these two reports are received by the system, the locations indicated within the reports are analysed and the relevance between these facts and opinions are determined. That is, the system can figure out that the building x locates in the area z, therefore, the opinions appearing within these reports support each other. Using deduction (forward chaining), we can infer from the first report that the sniper was in the area z, which supports the second report. Similarly, through abduction (backward chaining), the second report supports the first report. That is, if the sniper was at a building on the area z, it is possible that this building is the building x.

5.5 Fusing of assessed opinions

There are different ways of getting various opinions about the same fact. We can get opinions direct from one or more information sources and discount them based on the trustworthiness of their sources or infer opinions through some sort of reasoning mechanism such as deduction (i.e., forward chaining) or abduction (i.e., backward chaining). Therefore, we may expect more than one opinion about the same facts. At this point, we may fuse these opinions using cumulative fusion operator of SL as described in Definition 5.1 or some other fusion method. On the other hand, some of these opinions may be in conflict. The conflicting opinions may significantly reduce the quality of the fused opinion, especially in the environments with malicious adversaries. In this work, I advocate that conflicts between opinions should be resolved before fusing. In this section, I detail our approach and justify it using a set of simulations.

5.5.1 Conflicts between opinions

When we have two or more opinions about the same fact, these opinions may conflict. Let us assume that we have two information sources that provide opinions some how related to the fact f_1 "The road R is bombed". These sources are Peter and Jane, whose levels of opinion are 0.8 and 0.75, respectively. Peter reports that the fact f_2 "the road is safe" with opinion $(1, 0, 0)$; this opinion is discounted as $(0.8, 0, 0.2)$ based on his trustworthiness. Given that we have a certain rule $hasExplosion(x) \rightarrow \neg safe(x)$, we infer, lets say, the opinion $(0, 0.8, 0.2)$ for f_1 using abduction based on Peter's report. On the other hand, Jane reports an explosion on the road with opinion $(0.9, 0, 0.1)$; this opinion is discounted as $(0.675, 0, 0.325)$ based on her trustworthiness. Therefore, we have two opinions for f_1: $(0, 0.8, 0.2)$ and $(0.675, 0, 0.325)$. It is easy to conclude that these opinions are in conflict. The first opinion indicates that most likely there is no explosion on the road while the other implies that there is a significant possibility of explosion.

The conflict indicates that *at least* one of these opinions is misleading. To resolve this conflict, these opinions may be discounted (possibly at different rates), hence the uncertainty within them is increased enough to resolve the conflict. For instance, discounting the first opinion $(0, 0.8, 0.2)$ with 0.0 makes it $(0, 0, 1)$, which absolutely does not conflict with $(0.675, 0, 0.325)$ or any other opinion, since it does not contain any belief or disbelief, but only uncertainty. Similarly, discounting both of these opinions with 0.5 leads to opinions $(0, 0.4, 0.6)$ and $(0.335, 0, 0.665)$, which are quite uncertain and do not conflict.

In this work, I argue that conflicts between opinions may serve as evidence for the necessity to increase uncertainty within them. Although there can be more than one way to define conflicts between subjective opinions, in this work, I introduce Definition 5.3 where conflicts are defined based on the satisfiability of *beliefs* and *disbeliefs* within opinions. In the rest of this section, I represent opinions using only belief and disbelief when a coincide notation is required; that is, I use (b, d) instead of (b, d, u) since $u = 1 - b - d$.

Definition 5.3. Let $O = \{w^0, w^1, \ldots, w^n\}$ be opinions about the same proposition, where each opinion $w^i = (b^i, d^i, u^i)$. These opinions are consistent if it is possible to have a valid opinion that can satisfy all of these opinions. An opinion $w^* = (b^*, d^*, u^*)$

can satisfy the opinion $w^i \in O$ *iff* $b^i \leq b^*$ and $d^i \leq d^*$. Although there can be infinitely many such w^*, the one with the highest uncertainty (i.e., one with the highest u^*) would be

$$(max(b^0, b^1, \ldots, b^n),\ max(d^0, d^1, \ldots, d^n)).$$

Therefore, there cannot be any opinion that can satisfy all opinions in O if and only if

$$max(b^0, b^1, \ldots, b^n) + max(d^0, d^1, \ldots, d^n) > 1.$$

That is, O is inconsistent *iff* $\exists w^i, w^j \in O$ s.t. $b^i + d^j > 1$.

The intuition behind the definition of conflicts between opinions is as follows. Let the ground truth about a proposition x be represented as an opinion $w_x^\tau = (b_x^\tau, d_x^\tau)$. Also, let $w_x^i = (b_x^i, d_x^i, u_x^i)$ be an arbitrary opinion about x. If w_x^i has a higher *belief* than w_x^τ does (i.e., $b_x^\tau < b_x^i$), then w_x^i is misleading. Similarly, if w_x^i has a higher *disbelief* than w_x^τ does (i.e., $d_x^\tau < d_x^i$), then w_x^i is misleading. How much w_x^i is misleading depends on how much extra belief and disbelief it imposes. On the other hand, if $b_x^i < b_x^\tau$ and $d_x^i < d_x^\tau$, then w_x^i is not misleading, because it does not impose any extra belief or disbelief, but only extra uncertainty that conflicts with neither the *belief* nor the *disbelief* within the ground truth w_x^τ. In reality, I do not have the ground truth about x. Therefore, I cannot say whether w_x^i is misleading or not. However, if we have another opinion $w_x^j = (b_x^j, d_x^j, u_x^j)$, we can reason about if it is possible to have a ground truth for which neither w_x^i nor w_x^j is misleading. Such a ground truth exists if and only if $max(b_x^i, b_x^j) + max(d_x^i, d_x^j) \leq 1$. If such a ground truth cannot exist, we say w_x^i and w_x^i are in conflict, i.e., at least one of them must be misleading at some degree.

5.6 Resolving conflicts

Before fusing opinions regarding a specific fact, for example using cumulative fusion operator in Definition 5.1, the conflicts between opinions should be resolved. Once we determine conflicting opinions, several approaches can be used for this purpose. In this section, I introduce three conflict resolution approaches and demonstrate their performance in terms of the success of the fusion operation. In these approaches, discounting operator is used to resolve conflicts between opinions by increasing uncertainty in some of

these opinions. Let us have two conflicting opinions w_i and w_j from information sources i and j with opinion values t_i and t_j, respectively. These opinions are discounted by coefficients $0 \leq x_i, x_j \leq 1$ based on information like t_i and t_j.

1. **Trust-based deleting:** If two opinions w_i and w_j are in conflict, the opinion from the less trustworthy source is deleted, and if both sources are equally trustworthy both opinions are deleted. Thus, if the trust we have in the source of opinion w_i is greater than that of the source of w_j $(t_i > t_j)$ then $x_i = 1$ and $x_j = 0$, and when $t_1 = t_2$, we assign $x_1 = x_2 = 0$.

2. **Trust-based discounting:** If two opinions w_i and w_j are in conflict, they are discounted in proportion to the trustworthiness of their sources. That is, the more trust worthy opinion is discounted relatively less. Let us define a conflict over $w_i = (b_i, d_i, u_i)$ and $w_j = (b_j, d_j, u_j)$ as $b_i + d_j > 1$. Then, the additional discounting factors for w_i and w_j are computed as $t_i/(b_i t_i + d_j t_j)$ and $t_j/(b_i t_i + d_j t_j)$, respectively.

3. **Evidence-based discounting:** The previous two conflict resolution methods uses discounting based on trust to resolve conflicts. Trustworthiness of information sources are computed based on past experience, i.e., evidence. Evidence relates to how reliable or unreliable a source was in the past. The evidence-based discounting uses TRIBE [134] approach for conflict resolution and depends on evidence analysis instead of trust values. As described before, each opinion in the proposed system is already discounted by the trustworthiness of its source. Discounting an opinion further to resolve a conflict implies making its trustworthiness less than the trustworthiness of its source. For instance, let us assume that our trust for Jack and Jane as information sources are 0.8 and 0.9, respectively; and they give opinions $(1, 0, 0)$ and $(0, 1, 0)$ for the fact "the road is not safe", respectively. After discounting based on the trustworthiness of their sources we have opinions: $(0.8, 0, 0.2)$ and $(0, 0.9, 0.1)$. These opinions are still in conflict. To resolve the conflict, we need to discount them further. For instance, we can discount both of these opinions with 0.5 and have two non-conflicting opinions: $(0.4, 0, 0.6)$ and $(0, 0.45, 0.55)$. Hence, the trustworthiness of the first opinion becomes $0.5 \times 0.8 = 0.4$ and that of the second becomes 0.45. Let us assume that the trust in the source of the first opinion is computed as $(r + 1)/(r + s + 2) = 0.8$

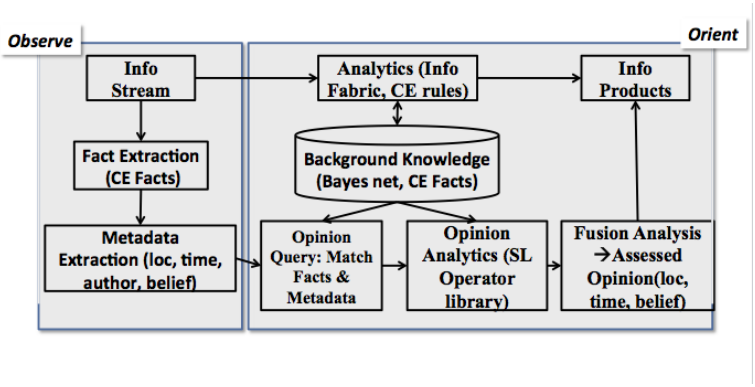

FIGURE 5.5: Opinion Assessment Framework.

based on Equation 5.1 where $r = 8$, $s = 0$. We can trivially calculate that 11 extra negative evidence, i.e., speculative evidence, should be added to s to decrease the trust from 0.8 to 0.4, i.e., $(8+1)/(8+11+2) = 0.4$. TRIBE aims at optimising the total amount of speculative evidence necessary to resolve all conflicts by finding the best discounting factors. Therefore, it considers not only the trust values, but also the evidence used to calculate them.

5.7 Trust assessment framework

The trust assessment framework has been implemented to show how streaming information can be assessed and annotated with a trust value and also to evaluate the level of trust when the information is fused.

5.7.1 Architecture

Figure 5.5 shows the trust assessment framework. Trust assessment model is implemented using Information Fabric along with the Subjective Logic triples. Information Fabric acts as the underlying framework that allows the trust assessment workflow to be a part of it. The fabric wrapper allows us to implement the required parts within it using services. The Fabric code has been implemented using Java code and built using Java Virtual Machine. The code is written using Eclipse SDK. The fabric wrapper calls out to these individual services as and when required. The Subjective Logic calculation code is imported into the Eclipse Development environment as a JAR file and is used with the code by the trust calculation and trust evaluation service.

5.7.2 Detailed design

The SL-based trust assessment process has been implemented on the *Information Fabric* (or *Fabric*) that serves as the underlying platform for executing the trust assessment workflows. The trust assessment operations appear as a service that is accessible over the Fabric. Specifically, packaged as a Java library, the SL-based operators are invoked as needed by the trust assessment service.

The trust assessment code is written using Java and Information Fabric has been made accessible as a wrapper to the trust service. The code connects to the Controlled English (CE) [82] store to check for the details on any existing information about the trust values of any of the sources and if not the Controleld English rule engine would do inference and the trust for the inference would be calculated using the appropriate Subjective Logic operator.

Assessed trust values are then used to annotate information products and appropriately update the background knowledge with revised trust values. We note that the trust assessment component only updates the trust values of information products and the facts in the background knowledge; it does not create new facts and thus does not trigger new rules in the analytics component. This decoupling ensures the stability (convergence) of the joint information processing and trust assessment pipeline.

5.8 Evaluation setup for trust assessment

All our experiments were performed on the SYNCOIN dataset [135]. I used the Controlled English Store to a priori extract facts from the dataset. These facts are then streamed over the Information Fabric to an in-memory database wherein facts are annotated with trust metadata.

5.8.1 Results of conflicts resolution

I evaluated the performance of the listed conflict resolution approaches in cumulative fusion through a set of simulations. In these simulations, there are two types of information sources that provide opinions about fact: honest and malicious sources. At each

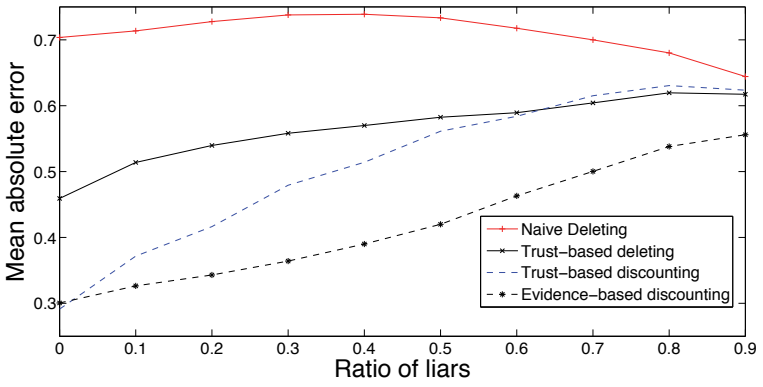

FIGURE 5.6: Error of cumulative fusion operator in the face of liars for varying R_{liar} .

time step, a fact is randomly selected and a ground truth is generated for the fact. Our fusion model collects opinions provided by various information sources and fuses these opinions by i) detecting conflicts, ii) resolving conflicts, and iii) applying the cumulative fusion operator to generate a single fused opinion. The success of the fusion operation is measured using absolute error. Let (b, d, u) be the ground truth and (b', d', u') be the fused opinion, then the *absolute error* is computed as $err_{B(a)} = abs(\delta_b) + abs(\delta_d)$, where $\delta_b = b' - b$ and $\delta_d = d' - d$.

The honest sources provide opinions close to ground truth; they add small amount of Gaussian noise with $N(0, 0.01)$ to belief and disbelief within the ground truth. The malicious sources behave like honest sources for a while until they build up some trust (above 0.7) and then they start to provide misleading opinions. The misleading opinions are generating by switching the belief and disbelieve within genuine opinions, which are close to the ground truth. In order to increase the uncertainty in the behaviour of sources, I also let honest sources provide misleading opinions with a probability of 0.1. During simulations there are 10 information sources and I have run the simulations for different ratio of liars.

For each ratio of liars, I repeat the experiments 10 times and I demonstrate the mean absolute error for our experiments in Figure 5.6. In the figure, *Naive Deleting* method is a baseline method in our evaluations. This method simply deletes all of the conflicting opinions before fusion. My results indicate that trust-based discounting and evidence-based discounting have similar performance for low ratio of liars. However, considering only trust is not enough for higher ratios of liars. Trust-based deleting and naive deleting approaches do not perform well. They lead to highly erroneous fused opinions. Evidence-based discounting is by far the best conflict resolution approach. It allows misleading

Metric	Mean	Std dev
Throughput (per sec)	33.6 / 33.5 / 2.98%	2.4 / 2.8 / 16.67%
Latency (ms)	32.2 / 36.4 / 13.04%	12.1 / 13.2 / 9.09%

TABLE 5.2: Performance Overhead. $x/y/z$

opinions to be discounted before the cumulative fusion takes place. Therefore, I suggest to use evidence-based discounting in the proposed system.

5.8.2 Performance

Table 5.2 summarises the performance overhead added by our trust assessment approach. I compared the latency and throughput of a `nop` (no operation) Fabric service with that of our Fabric trust service by streaming events through them at high speed. The `nop` Fabric service takes a stream of events as input and returns an identical stream of events as output. Thus, execution of `nop` reflects the cost of an event entering and exiting the Information Fabric platform without any processing within the Fabric platform. The trust service takes a stream of events as input and returns a trust annotated stream of events as output. I have repeated the experiment by calling each of the SL operators shown in Table 5.1. The variation in latency and throughput across these SL operators were hardly noticeable. All our experiments were repeated seven times until the numbers converged.

In Table 5.2 x denotes the performance of a `nop` Fabric service, y denotes the performance of the trust service and z shows the relative overhead (i.e., percentage decrease in throughput and percentage increase in latency)

My initial results indicate that including a trust service does not significantly alter the throughput of the Fabric platform. In general, I expect the information analytics operator is likely to be at least as complex (if not more) than the trust assessment operator; hence, the relative loss in throughput is expected to be reasonably small (roughly 3%). With regard to latency, the trust service adds a relatively larger overhead (13%). This is because in my current implementation, the trust assessment operator does not run concurrently with the analytics operator; I require the analytics operation to be complete and the output information product be available before the trust assessment operation can infer the trust metadata on the information product. For future implementations, I anticipate that once the choice of the analytics operation is determined, e.g., which

Controlled English rule will be applied, I could concurrently issue a request to the trust service without waiting for the analytics operation to complete. Hence, in this case, I expect that the latency overhead due to the trust service would decrease.

5.8.3 Accuracy

Next, I compare the accuracy of the assessed trust using our online trust assessment system versus an offline approach. The offline approach waits for all reports to be received before assessing trust in various facts. In contrast, my online approach operates on a stream of incoming reports with the goal of ascertaining trust in various facts as information arrives. Assuming that trust levels in reports do not change with time, the trust assessed by the offline approach can serve as a benchmark, representing a ceiling for the trust levels that can be ascertained as the entire body of information is available at decision time. Against this benchmark, I can assess how our online approach fares. The online approach operates on a compact set of facts where the incoming reports are discarded immediately after the extraction of facts and trust assessment. I will refer to the difference between the trust levels attained by the offline and online approaches as the *accuracy* in the trust assessment.

Figures 5.7, 5.8 and 5.9 compare the accuracy for different classes of trust analytics operators. Note that the forward chaining operators (such as the Subjective Logic `AND` and `OR` operators) and the Subjective Logic consensus operator are associative and commutative and, hence, the order in which these operators are applied does not change the outcome. As a result, the online trust assessment approach will eventually converge to the value attained by the offline approach when these operators are used; this behavior is seen in Figures 5.7 and 5.8. On the other hand, backward chaining operators (such as the Subjective Logic deduction and abduction operators) are neither commutative nor associative and, hence, the order of their application does matter. In the backward chaining case, in figure 5.9, the online trust assessment system exhibits inaccuracies by assessing trust at levels that remain "distant" to those attained by the offline approach.

The inaccuracy notwithstanding, the online approach allows for fast-paced evaluation of trust as information arrives. In the particular example, besides the computer processing time, there is practically zero delay from the moment the information arrives until a new trust level can be assessed. When trust level reaches satisfactory levels and remain

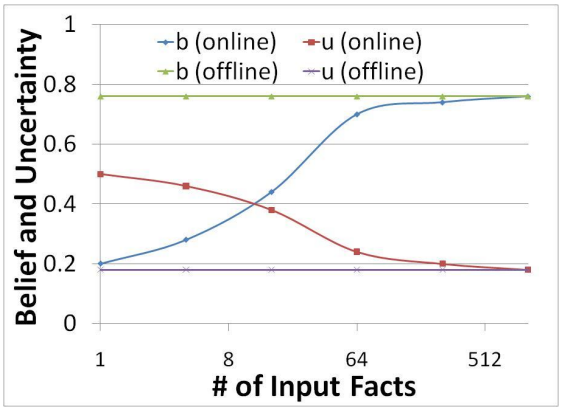

FIGURE 5.7: Forward Chaining.

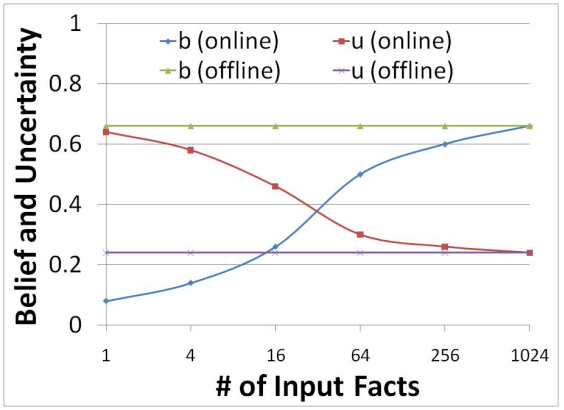

FIGURE 5.8: Consensus (Self Chaining).

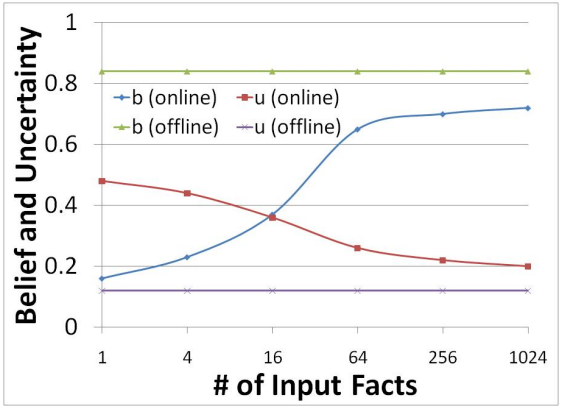

FIGURE 5.9: Backward Chaining.

at such levels for a sufficient period of time, decisions could be made and actions taken at earlier times, resulting in fast-paced execution of the OODA loop as well. Clearly, the level in trust accuracy attained and the responsiveness with which it is attained presents a design and system management trade-off parameter. To this end, in future work, I will explore operational implications of alternative designs where, for example, I trade-off delay to accuracy by operating in a near-streaming manner. In the latter case, the system could operate over time frames assessing trust based on reports arriving over a single frame in an offline fashion, but assessing trust over successive windows in an online fashion.

5.9 Summary

This chapter has proposed a new trust assessment framework that could be used to evaluate the trust on the fused information in a streaming setup. The chapter also discusses Subjective Logic operators and how these Subjective Logic operators can be used to achieve a level of trust on fused location information. The chapter also describes the methods used to resolve conflicting opinions based on a number of mechanisms including trust based deleting, trust based discounting and evidence based discounting. This chapter describes the implementation details for achieving trust on fused information using Information Fabric and CE store. Further, this chapter evaluates the trust assessment model and reviewed the performance with the conflicts resolution. It can be observed that the service does not significantly alter the throughput of the platform but with better analytics the latency could be decreased. I also compare the accuracy of different classes of trust analytics operators.

Chapter 6

Location and identity privacy enforcement

6.1 Overview

Mobile devices can request for location based services from service providers. This leads to various attacks made in order to tamper the end user security. Hence end user security is another major issue along with the data security flowing from one end to the other. On requesting a service from a mobile service provider, the location and identity of the individual making the request is unknowingly accessed by the Service Provider leading to taking advantages and misusing it. Hence preserving the security of the individual including location and identity is a very difficult open research problem today. One of the approaches is to ensure the location security and the other is through identity security. Both these techniques are equally important as either of them can compromise the security in the mobile space. There is a vast body of work in the area of location security [44] -[46]. However, there is no thorough evaluation on where to enforce location security and what are the trade-offs involved. There are two approaches for information flow control in mobile environments.

Privacy has been an issue from the time technology has been into existence. In the mobile world, there is a lot of data being transmitted from and to and hence there is a very important place for privacy. The information being transmitted could be simple information or could be much more sensitive information that could be misused if fallen

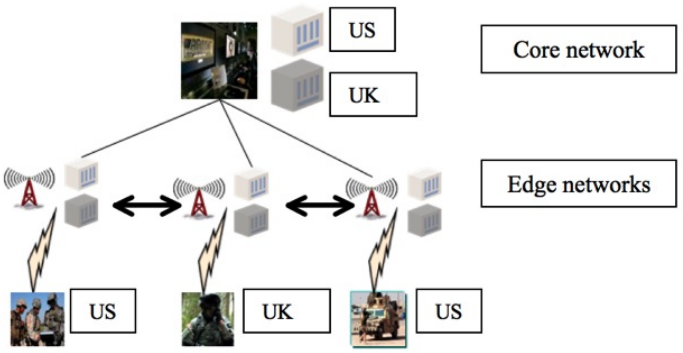

FIGURE 6.1: A tactical network scenario – enabling efficient computations over dynamic networks.

into the hands of a malicious user. Mobile devices are being used not just for data transmission but also for requesting services from various service providers available. When a user makes requests to service providers asking for a specific location based service, the service providers would be using the information of the user and also the location of the user. This information can be misused by the service provider or can also be handed over to other third parties without the consent of the user. Hence it is very important and critical to preserve the self information and also the location information of the user making the request to the service provider. The identity of the individual making requests needs to be preserved as the requestor would not like his personal information to be public. Similarly the mobile device holder would not be agreeable in giving away the location information as it would be a breach of his security. The results of knowing once location and identity can open doors to a lot of activities that a malicious user can perform. Hence preserving the privacy of any individual is very important.

It is very important to understand the placement of the location security solution and where to enforce security is key to this work. Introducing mobile micro-cloud in this section will help in understanding the placement of the solution. Mobile micro-cloud [136] envisions that applications (or computing tasks) will be deployed in a mobile micro-cloud, a logical network composed of two components, the core (e.g., the command and control center) with access to large quantities of static (and possibly stale) information and the edge (e.g., the forward operating base) with access to smaller quantities of more real-time and dynamic data. The edge and core are separated by dynamic and performance constrained networks with a many-to-one relationship between the core and

the edge. It is also possible for edge nodes to communicate with each other. Further, the (edge and core) nodes can belong to different coalition partners, raising the question of security and operational policies for handling of data and computation. Figure 1 illustrates a typical architecture of the mobile Microcloud in the army coalition context. The benefits of embedding storage and computation into such a micro-cloud tactical network are two fold: (i) Effective provisioning for diverse information requirements the micro- cloud supports users with different latency requirements and access rights and (ii) Effective information exchange in a constrained environment. Complete shuffling of information is impractical in a tactical network and the micro-cloud reduces congestion by providing computation at the edge. This chapter shows a detailed implementation of the location privacy application. The chapter also details the evaluation setup for enforcing location privacy and the results.

6.2 Device Vs edge based solution

This section focuses on Device Vs Edge based implementation and the trade-offs in them. It quantifies the trade-offs and proves that edge based solution is the better solution for enforcing security. My results show that while device-based solutions do not require trust in the edge location server, they either suffer from high false positive rate (about 25% probability of not meeting the desired security requirement) or low utility (about 600 meters higher error in obfuscated location data).

6.2.1 Solution at the core

The core is the centralised network and hence has a lot of bandwidth and can maintain huge repository of information. It also has a lot of computational power allowing it to process complex solutions. It is important to note that it takes longer time for transactions to work between the device and the core. This is a major drawback to the location based solution where spontaneous decisions needs to be made. Solution at the core will retain same false positive and false negative and will have a very high latency.

6.2.2 Solution on the device

The delays caused due the solution being placed at the core of the network gave rise to the new wave of solutions that were placed on the device. It is important to notice that the device doesn't have a lot of flexibility, bandwidth or computation power. Besides any of these, the device does not have visibility of the other devices in the network. Hence any kind of computations performed by the device will not be leading to accurate results. It could very well lead to misleading answers to the user's request. This leads to the new methodology that I introduce in this work called the solution at the edge of the network.

6.2.3 Solution at the edge

The edge of the network is closer to the device and is an intermediate channel between the device and the core of the network. The edge has visibility of all the other users in the network and the edge can perform computations faster and provide results spontaneously to the device. The advantage of having the solution at the edge is that edge will have information about other people and hence solution will have lower false positive and lower false negative. The only catch with this solution is that trust with the edge is needed. The edge will have the raw obfuscated data or slightly obfuscated location data. Latency with this solution is higher than device based solution and is lower than the solution at the core. This helps the device user make decisions on the location based service requests that one has. Hence this solution is the best solution compared to the other 2 solutions explained above. In this chapter I compare optimal choice on the device (based on historical data) with optimal choice on the edge and examines the trade-offs between enforcing location security at a device vs. enforcing location security at an edge location server. To the best of our knowledge this the first attempt to quantify the effectiveness gap between the optimal solution at the device versus that at the edge/core.

6.3 Location privacy enforcement

Location Security mainly deals with the location of the requester. In mobile environments, users requests for location based services very frequently, for example, when the

user requests nearby restaurant information from the location based server, the location based server needs to know the location of the user and hence the location information is normally requested. However, in most of the cases, the user doesn't want to disclose the location information to arbitrary location based service providers. This can be achieved by a number of different mechanisms. One of the well known methods is k-anonymity. In this method, users location information is updated with pseudo-ids and then the generalised location information is sent to the location based service provider. Due to some groups being created that fail to provide overall anonymity, another mechanism called s-proximity has been implemented [37] . This mechanism creates a larger number of anonymous user profiles to ensure that the location based service provider cannot identify the location of the requestor. Another location Security mechanism that is described is Casper [38]. Casper is a combination of location anonymiser and Security aware query processor. Few other mechanisms like the Encrypted data store [39], key agreement [40], privacy tools [41], In-device spatial cloaking assisted by cloud [42] is also part of the location security.

6.3.1 Anonymization methods

Preserving privacy using anonymization has been discussed in a number of research papers [37], [43], [44], and [45]. The authors in [46] have looked at the k-anonymity in order to generalise the location. The user of a mobile device usually requests information for 2 main types of resources namely static resources and mobile resources. In case of static resources, pseudo-identifiers are sent and the location is anonymized. In the case of mobile resources, IDs are updated with pseudo-ids and then the generalised location and profile are sent back to the requestor. Figure 6.2 shows the representation of this k-anonymity model. It shows how the mobile device makes a request to one of the location service providers asking for a location-based service. The anonymizer and the location information pick this up and the mobile user information is anonymized and is then transmitted to the location service provider. In this way the user and location information are hidden.

Although the profile anonymization model works well using the k-anonymity, there have been a number of attacks that can be performed on the k-anonymity model that has led to the identification of the query issuer in the location based services. To overcome

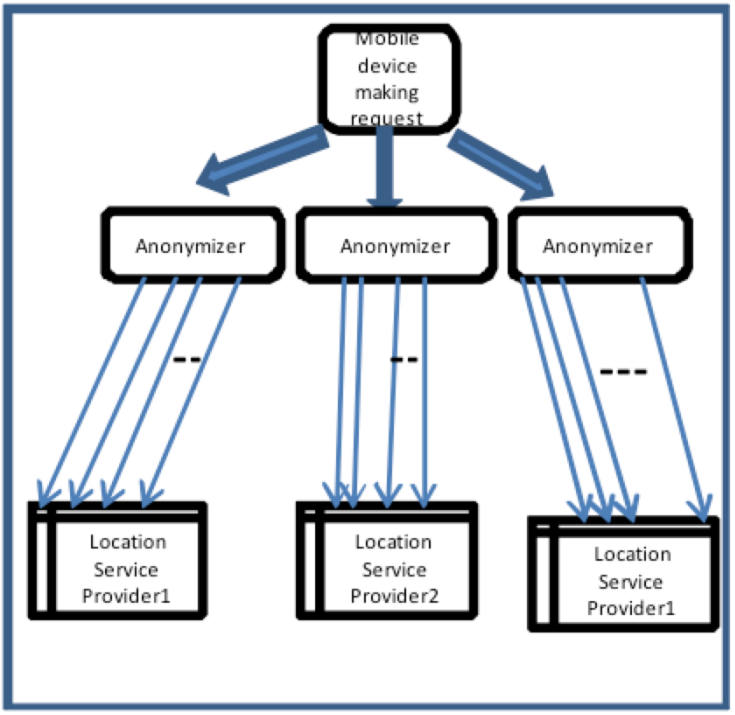

FIGURE 6.2: k-anonymity model.

some of the shortcomings in the current k-anonymity model the s-proximity model was proposed. The next paragraph discusses the advantages of the s-proximity compared to k-anonymity model.

The k-anonymity model as described above tries to hide the location of the query issuer who tried to request for location based information from the Location Service Provider (LSP). In [37], the authors shows that the k-anonymity is not enough as it can be easily prone to attacks thus resulting in the re-identification of the query requester. Two main attacks that have been explored are heterogeneity attack and conformity attacks. K-anonymity can create groups that fail to provide the overall anonymity due to lack of sufficient match among members with respect to some sensitive user attribute. The communication between the query requester and the Location Service Provider shown in the Figure 6.2 is as follows: Initially, the user sends a location-based query to the Anonymizer, which then replaces the exact location with a cloaked region. It is then passed on to the Location Service Provider. The attacks prove that in this process, by some combined work by the Location Service Provider's or a Location Service Provider can individually break down the anonymity set and prove the identification of the specific query requester in cases where the query is specific or not too generic. Hence [37]

comes up with a solution that generalises the query and hence makes it difficult for the Location Service Provider to identify the actual query requester. This is achieved in the s-proximity model. The paper suggests that both k-anonymity and s-proximity are needed to anonymize the query requester's identity in a location-based service. In the s-proximity model, the Anonymizer is replaced by context aware anonymizer with further modules such as query generalization, query analyser and partitioning agent. With the detailed implementation of these modules privacy of the user is preserved and hence the privacy preservation is achieved in allocation-based environments.

Hence I propose a model that uses both k-anonymity and s-proximity in order to perform anonymization at the edge so that the location security is enforced at the right place.

6.3.2 Mobile Integrated Server

I propose a new architecture that would ensure that the privacy information of the user is preserved. This is achieved by using one of the main components called Mobile Integrated Server (MIS). MIS is a dedicated server that is mainly used to perform two tasks. Its existence is mainly for the purpose of preserving the privacy. Privacy as discussed before is mainly consisting of any information that would be related to an individual and his existence. Hence MIS is mainly targeted in preserving the location information of the mobile user and the identity of the user. This is done by the two main parts of the MIS namely Location Concealer and Profile Cloner.

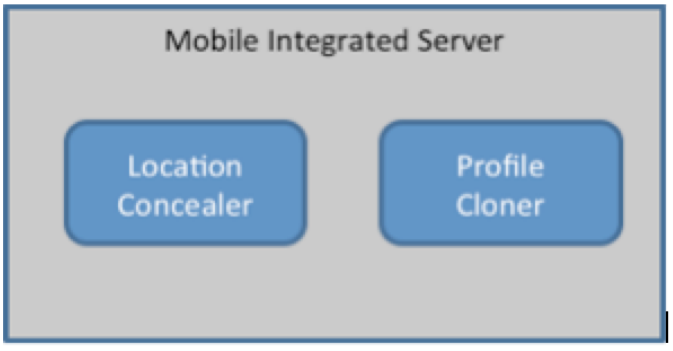

FIGURE 6.3: Mobile Integrated Server showing Location conealer and Profile cloner

The Location Concealer as the name suggests makes sure that the location of the user making the requests to the mobile service provider is hidden and is not retrieved for the service provider to misuse it. Location Concealer accepts the location information

from the mobile user making requests to the service provider and performs some manipulations on the location information and transforms the information sent to the service provider.Figure 6.3 shows the MIS components. Hence, the location information does not reach the service provider. The other part of the MIS is the Profile Cloner. When the mobile user sends requests for services to the service provider, MIS picks it and looks at the user's profile information that is sent as part of the request. It then makes a large number of similar profiles so that the profile of the actual requestor is cloned into a number of similar profiles making it difficult for the service provider to track the actual requestor's identity.

Location Concealer The Location Concealer is mainly used to hide the location information of the mobile user trying to request services from any mobile service provider. The concealer conceals the location information in the MIS and makes use of a general value which is passed on to the Location Service Provider.

Profile Cloner Profile Cloner clones the user's identity and then by using a large number of similar profiles send similar requests to the service provider. As a result, the service provider will not be able to identify the actual mobile requester. This larger set of profiles is created by using a cloning algorithm

6.3.3 Registration process of mobile user with MIS

The mobile user has to register with the MIS before performing any transactions. Mobile provider needs to have the unique identifiers of the mobile device in order to allow the MIS to authenticate with the device. Mobile provider receives the hash values of IMEI and IMPI numbers from the mobile device in encrypted form. A small application within the mobile device would encrypt the hash values of IMEI and IMPI numbers using the public key of mobile provider. The mobile provider would then decrypt and retrieve the hash values using its private key. Figure 6.4 shows the registration process steps.

Once the MIS is assigned by the Mobile provider it would also provide the hash of IMEI and IMPI numbers of the particular mobile device. MIS stores these information for establishing a secure connection between the MIS and mobile device. The initial registration process of the mobile device with the MIS will be done based on the unique identification of the mobile device. When the mobile provider assigns a MIS to the

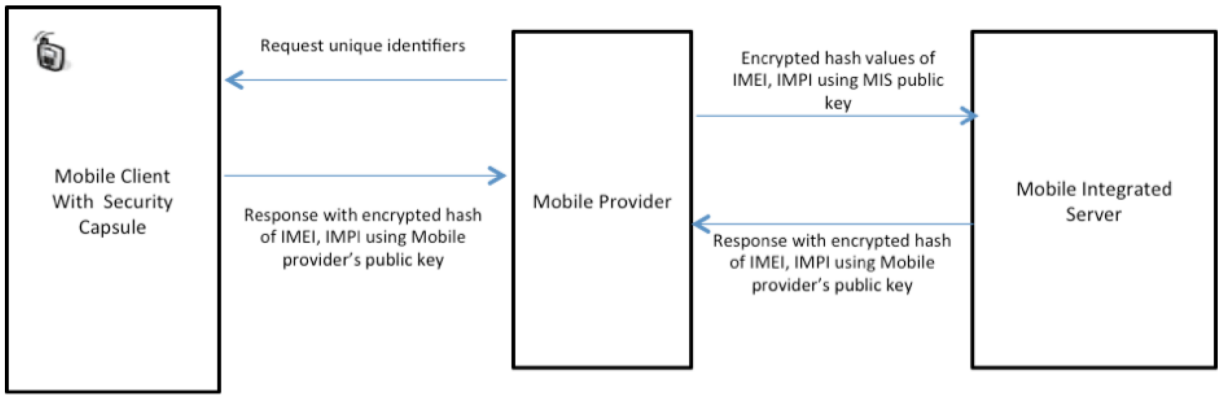

FIGURE 6.4: MIS registration process of a mobile device

mobile device, it would send the Mobile Intergrated Server's public key to the mobile device and parallely it would send the hash of IMEI and IMPI numbers of the mobile device to MIS. The hash values would be sent to MIS by encrypting them using mobile the public key of MIS. The MIS will then decrypt the hash value of the unique identifiers using it private key.

The mathematical representation of sending the mobile device's unique identifiers to the MIS is shown as below: On the Mobile client side, Hash (IMEI, IMPI) = Huni

Enc (Hash (IMEI, IMPI)), MPPub = CipherText1

Enc (Huni, MPPub) = CipherText1 = Enc1

On the Mobile Provider side, CipherText from mobile client that was generated for Hash (IMEI, IMPI) = Enc1

Dec [Enc (Hash (IMEI, IMPI)), MPPub , MPpriv] = DecMP

Enc (Hash (IMEI, IMPI)), MISPub = CipherText2 = Enc2

On the Mobile Integrated Server side, CipherText from mobile provider = Enc2

Dec [Enc (Hash (IMEI, IMPI)), MISpriv] = DecMIS

The Mobile Integrated Server stores a copy of the Hash(IMEI, IMPI) for the mobile device to be used during initial handshake of the mobile devicewith the MIS

6.3.4 Mobile user handshake with Mobile Integrated Server

The process involving the handshake of the mobile user with the Mobile Integrated Server involves the unique identifiers of the mobile users. The unique identifiers of the mobile device IMEI and IMPI are hashed and encrypted using the public key of the MIS. The mobile provider keeps a copy of the public key of the MIS in the security capsule within the mobile device and can be accessed by the security capsule only. Figure 6.5 shows the handshake between user and MIS.

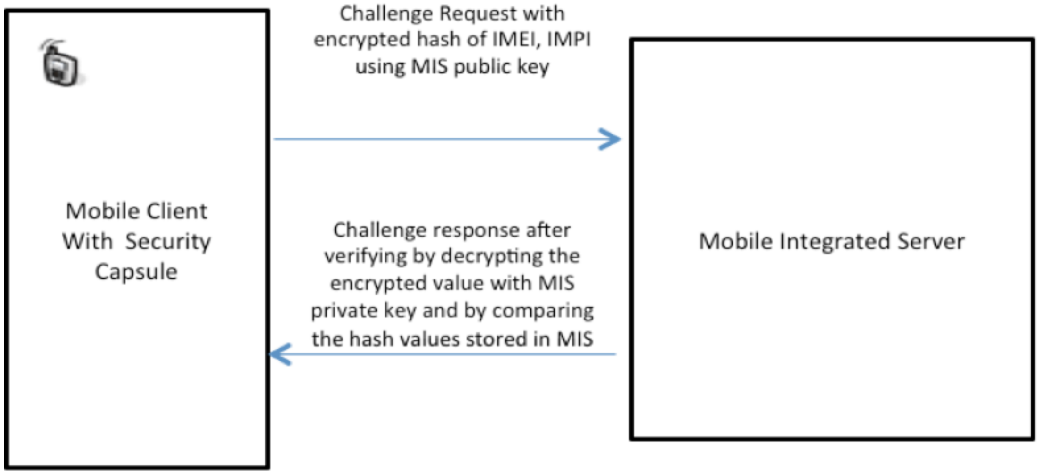

FIGURE 6.5: MIS handshake with mobile integrated server

On the mobile client side: Hash (IMEI, IMPI) = Huni

Enc (Hash (IMEI, IMPI)), MISPub = CipherText = E1

On the MIS side :

CipherText from mobile device = E1

Dec [Enc (Hash (IMEI, IMPI)), MISpriv] = D1

D1 = Hash (IMEI, IMPI) = Huni

The MIS already has a copy of the hash values stored in DecMIS A comparison of DecMIS and Huni are performed. If the comparison succeeds the initial handshake between the mobile device and the MIS is established for future transactions.

6.4 Identity privacy enforcement

Identity obfuscation is equally important when mobile device users are requesting for location based services. The location and identity are both being captured by the service providers in order to provide the user with the accurate results to the query. Hence it is very important to obfuscate both the identity and location so that the security is enforced but accurately providing the same set of results to the query.

In a mobile micro-cloud setting there are multiple entities with varying concerns for trust and vulnerability wherein a security solution may be deployed. I examine the interplay between security solutions and the mobile micro-cloud architecture. I use Identity based security as a use case. In particular I consider identity cloning to throttle information release to coalition partners. An identity cloning module may operate in one of the following three layers in a mobile micro-cloud: (i) device, (ii) edge and (iii) core. Each of these solutions has diverse implications on the overall performance of an application (e.g., latency), application quality and security objectives. For instance, a device based solution has incomplete information (on the location of other devices) and thus may either be over conservative (more obfuscation which leads to poor application quality) or too liberal (less obfuscation and thus fail to meet security requirements). In the implementation section, I showcase using simple performance, quality and security metrics the effectiveness of operating a security solution at different layers of a mobile micro-cloud. My observations highlight the need for edge computing for delay sensitive and security conscious coalition applications. Here I highlight the importance of preserving the identity information and still being able to get the same accurate results for the query. All the existing solutions out in the market today be it any smart phone takes the location, identity and query when one makes a location based service query. By default the Service providers use the identity information and hence knows where you are and who you are. This is a major breach of privacy if one does not want to share the identity information. In addition to this, the service providers also misuse the information and also pass it on to third party applications who further misuse the identity information. The key idea is to exploit the availability of low latency edge server to do identity cloning in order to preserve the identity of the mobile application user. Location anonymization has scope for improvement and hence identity cloning is our new concept to mobile security. The concept of mixing user identifiers have been around for several

decades (e.g., Chaumian mix). The key novelty here is to extract user profiles and mix only users whose profiles are within a certain distance from each other (i.e., mixing is only done amongst users with similar profiles).

In the identity cloning scenario, I mix the identifiers of multiple users with the identifier of the user making a query. By doing this the Service provider will not be able to distinguish who the actual user making the query is among the population of k users. k-anonymity matrix is a very well known matrix but this new concept of identity cloning is novel and is my contribution to research.

Take a hypothetical scenario where you have 1 coffee drinker among k- 1 other drinkers. Obfuscating the location would be sufficient and good enough as long as there is an envelope that includes other coffee drinkers. But in scenarios where there are no coffee drinkers at all except the one user making the query to the Service provider, obfuscating the location may not be sufficient. In such scenarios, introducing identity cloning is very important and a must. Identity cloning is achieved by taking the identity from the user and using the trust server on the edge to clone the users identifier by mixing it with one of the other identifiers that match the same criteria as the user in question. The key Steps are as follows:
(i) Hosting an edge server at the network edge (WiFi access points, cellular basestation, etc.).
(ii) Building user behavior models (e.g., based on click stream and interactions with applications).
(iii) Device trapping access to identification information (IMEI , IMSI) and location data from an app and redirecting the request to a (spatially) nearby edge server.
(iv) The edge server obfuscating the identifier and/or location data based on the location and identities of other devices in the vicinity, the proximity of user profiles and privacy preferences of the end-user and returning the obfuscated data to the device.
(v) device returning the obfuscated identifier and location data to the application.

6.5 Location privacy application

This section shows the implementation of “Where to enforce Location privacy” and highlights the archicture and design to show the usecases of the solution being at the

edge and on the device.

Here I describe the solution architecture for preserving the privacy in a mobile environment. As seen in the previous sections, the main component of the architecture is the Mobile Integrated Server (MIS). Additional players in the architecture are mainly mobile handset, mobile Providers and location based Service Providers. The below diagram shows the end-to-end flow of the solution architecture.

The mobile device is always tied to a mobile provider namely (Vodafone, O2 and so on). In our solution, it is assumed that the mobile provider randomly allocates a MIS to the mobile device. The allocation would be done in such a way that in future even if the mobile provider wants to find out which mobile device is tied to which MIS, it would be impossible. There is no way the mobile provider would be keeping track of the MIS assignment. Once a mobile handset is tied to a MIS, initial handshake between the mobile device and MIS will be done. This handshake is done based on the key that is generated on the mobile device. A random key is generated on the mobile device and this key with some additional shared secret is used to create a handshake with the MIS. Once the handshake is established, the mobile device is tied to the MIS. This initial handshake is done based on the unique characteristics of the mobile device. There is an initial authentication process that is performed using keys. Once the handshake is established, the mobile handset can then start making requests to the Service Provider asking for location based services.

Mobile Integrated Server is a Server which in a non-cloud setup will be a server installed in a specific secure location using webserver technology. In case of a cloud environment, the MIS can be located anywhere in the Cloud and hence it doesn't matter where the MIS is exists physically. MIS has 2 main components. One of them is the Location concealer which is used to hide the mobile user's location information when the mobile user makes a request to the Service Provider asking for a particular location based service (eg: nearest restaurant, nearest car park). The second component is the Profile cloner. Profile Cloner clones the user's identity and then by using a large number of similar profiles send similar requests to the service provider. As a result, the service provider will not be able to identify the actual mobile requester and his location. Figue 6.6 shows the application scenario and the interactions.

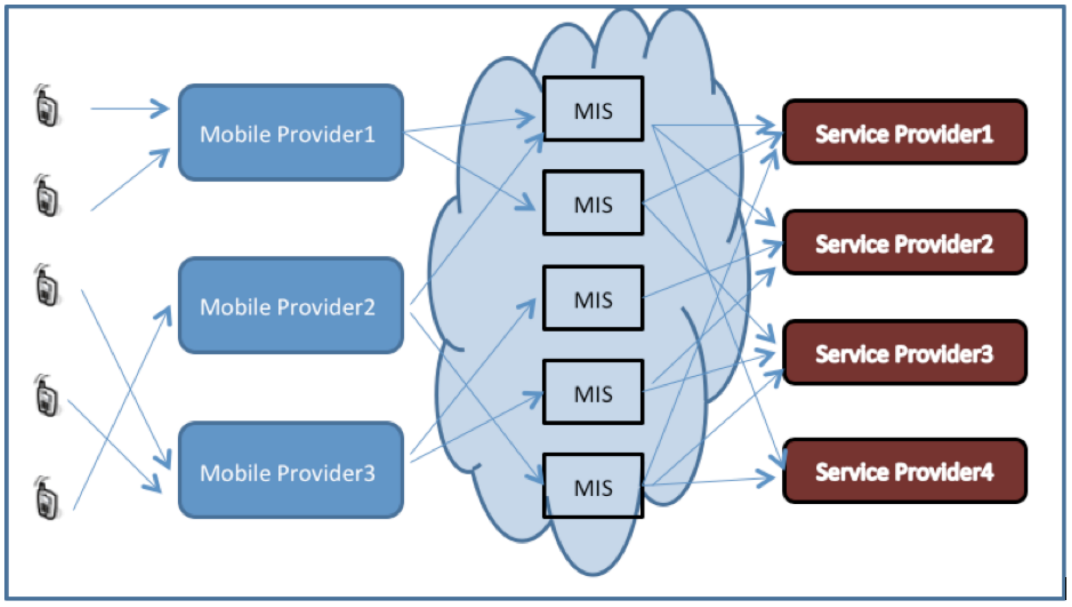

FIGURE 6.6: MIS architecture of the end to end solution

The location information is concealed in the MIS and is generalized by making use of some kind of a general value. The generalization of the location information is carried out by making use of a unique concealing algorithm. The concealing algorithm goes through the process of manipulating the location information and creating a generic value as the output of the algorithm. Once the generic value is created, the profile cloner then takes the profile of the service requesting mobile device and using a profile multiplier algorithm creates a huge list of similar profiles as the output. The generic value information from the location concealer along with a large number of similar profiles created by the profile cloner is sent to the service provider making the request for a particular service. The service provider will look at the incoming requests and assume that the requests are coming from a large group of profiles and it provides the location based service requested. This results in the service provider being totally unaware of the service requesting mobile user's location and identity. The Service provider sends back the information to MIS and MIS would be looking at this information and it can process the information in two ways. It could forward this information as it is to the specific mobile user or it can narrow it down further and send it to the user. There are a number of tasks involved within the MIS that are to be done by the Location concealer and the Profile cloner. This includes writing a Concealing algorithm, multiple profile creation and query formation. The expression can be described as follows: (Mu, Uq)

Mobile user information and User query This is passed through to MIS MIS transforms this into (Mu1, Mu2, Mu3. . . . . Mun), (Uq) Anonymization is applied to this and it transforms to Anony (Mu1, Mu2, Mu3. . . . . Mun), (Uq) The resulting output from the MIS is a generic value and a huge set of profiles requesting a particular location based service from the Location Service Provider. The proposed architecture would be simulated and implemented in the near future and the results for the performance and the accuracy would be published in the near future.

6.5.1 Architecture

This work has been implemented as an android based system. An application has been implemented in the android device in order to showcase the difference in the 2 methodologies. The solution at the device and the solution at the edge have been implemented using an example of the London Boris bikes. Boris bikes are the easiest way to hire a cycle, ride it where you like and return it to any docking station. In this implementation, we have shown the means of how the system solution works when the solution is at the edge and when its at the device. In order to perform the implementation, we have made use of an application in an android device and then have implemented an edge server on a windows server. This server behaves as an edge which has the visibility to all the devices in the network and perform computations accordingly. The device based solution shows an Android application with the map of London in it indicating the Boris bikes available for hire. Request from the mobile device is shown on the map by indicating the current location of the device. By performing obfuscation on the device, it can be noticed that the obfuscation is not accurate enough as the device does not have visibility to other devices in the network. When the user then makes a request for the bikes, the responses received are not accurate due to the drawback of inaccurate obfuscation. In the case of solution at the edge, the edge has visibility to all the devices. When the user makes a request asking for the nearest bike hire from the current location, the edge takes care of obfuscating the current location of the device in comparison with the other devices in the network who would have made similar requests. The request is then sent from the obfuscated location and this results in accurate responses for the user requesting the locations of the bikes nearby from his location. Figures 6.7 and 6.8 show the different stages in the demonstration of the location based request with the anonymized location and the results of the query.

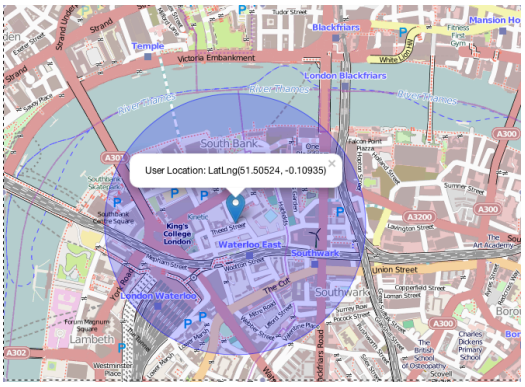

FIGURE 6.7: Deviced based solution view of the London Thames region (Blue dot is the user location).

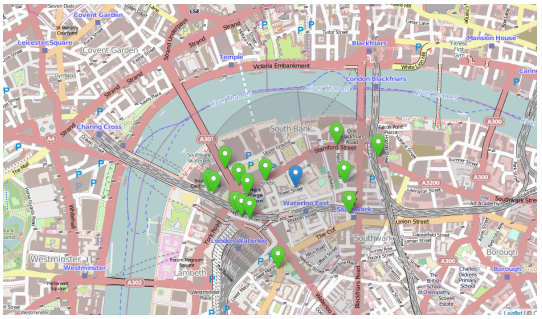

FIGURE 6.8: Green dots showing search results for the device based solution.

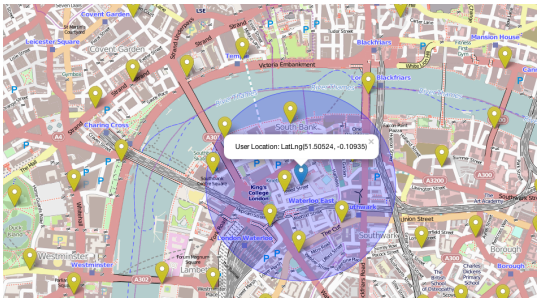

FIGURE 6.9: Yellow dots showing devices that are visible to the edge server with the user location shown in blue.

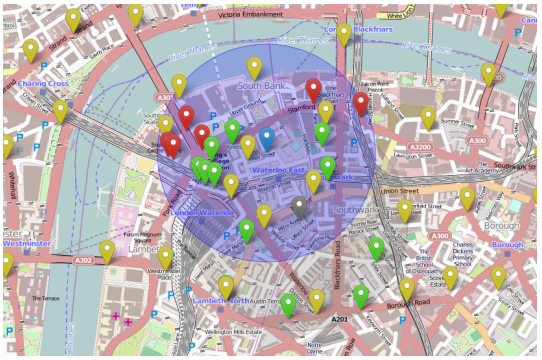

FIGURE 6.10: Green dots showing query results from true (blue dot) and obfuscated (grey dot) location. Yellow dots are other devices, red dots are the results missed when searched from obfuscated location

6.5.2 Detailed design of the solution

The solution has been fully implemented using the Eclipse development kit and has been tested with real use case scenarios. Figure 6.7 shows the device based solution where user clicks on a particular point and then checks are done to see if the chosen location has enough obfuscation. Device level obfuscation cannot be performed as the device has no visibility to the other devices. Hence checks are done at the edge server to ensure that the obfuscation is good enough to make a query. Figure 6.8 shows the search results for Boris bike using the device based solution. Figure 6.9 shows the view that the edge server would have of all the devices. Since the server can see all the devices, when a device makes a request for the bikes, the server can obfuscate the location based on the other devices in the area. The query resutls highlighting the bike locations from

the true and obfuscated location are shown in Figure 6.10. The comparison of results based on the search from the true location and the obfuscated location is shown using the 2 circles. This proves that the edge server functions close enough to the query made directly to the Boris bikes provider without any obfuscation.

6.6 Evaluation setup for enforcement of location privacy and results

In this section I present an empirical evaluation of the proposed location information flow control solution. Table 6.1 shows a summary of the datasets used for evaluation. Three of the datasets `Shanghai` [137], `San Francisco` [117] and `Stockholm` are taxicab traces obtained from the respective cities. The fourth (`Cellular`) is a user location trace and URL accesses obtained from a cellular network. The fifth (`Watson`) is an enterprise dataset obtained from WiFi location traces and URL accesses. The Stockholm dataset, cellular data and Watson data are proprietary datasets. These databases are the standard databases that have been used in a number of simulations on various privacy related experiments. I obtained it from a research website.

In the `Shanghai` and `San Francisco` datasets, there are explicit markers that indicate when the taxicab is occupied; in the `Stockholm` dataset collection of location trace is turned off when the taxicab is occupied (i.e., I only have trajectory information when the taxicab is not occupied). I use these datasets to quantify tradeoffs between the extent of obfuscation and anonymity.

In addition to these datasets, I use coarse grained mobility data from 16K mobile users obtained from CDRs (Call Detail Records) and from about 1.2K enterprise users obtained from WiFi and web data accesses. While a taxicab's trajectory may be viewed as a mixture of several user trajectories (i.e., multiple passenger trajectories), this dataset captures movement information at the granularity of each user. However, location information is captured is at the level of cellular Basestation association, which depending upon urban/rural areas can range from a few 100 meters to about 5,000 meters. From a population of about 11.6M users, I selected about 16K users that had more than 400 CDRs per day (i.e., >400 location samples and data accesses per day). While I use the taxicab dataset to analyse fine grained trajectories (each corresponding to one

Characteristic	Shanghai	San Francisco	Stockholm	Cellular	Watson
Sampling rate	2/min	12/min	1/min	>400/day	all web accesses
Number of entities	∼ 10,000	∼ 500	∼ 2000	∼ 16,000	∼1200
Source type	GPS	GPS	GPS	Cellular Basestation association	WiFi
Privacy	None	None	No sampling when taxi hired	Coarse grained samples	None
Timeline	1 month	1 month	1 month	1 month	1 month
Total number of trips	1,335,360	26,767	570,690	55,200	-
Total number of web accesses	-	-	-	12.4M	5.6M

TABLE 6.1: Summary of datasets.

trip), I use the cellular and enterprise dataset to analyse mobility across multiple trips undertaken by a single user.

Figures 6.11, 6.12, 6.13 and 6.14 show the average anonymity as the extent of obfuscation is varied for times 7am-10am, 10am-4pm, 4pm-7pm and 7pm-7am respectively. As the extend of obfuscation is increased so does the extent of anonymity; further anonymity is generally higher during busy hours in the morning and the evening because several mobile users are active within a small spatial extent. The key challenge in practice is that these measures of anonymities are averages over the respective dataset. Hence, given a user location at a point in date and time, the challenge is to identify the amount of obfuscation required to achieve a desired level of anonymity.

Figure 6.15 shows the number of users on the y-axis and similarity on x-axis. A point (x, y) in the figure indicates that there are at least y users whose profiles have a similarity of at least x with a randomly selected user. Similarity between user profiles is computed using a cosine distance on the set of URLs (web pages) accessed by a user with that of another user.

Figures 6.16, 6.17 and 6.18 show the complexity of a device-based model and false positive and false negative rates in enforcing the desired level of anonymity. A choice of obfuscation k is said to result in a false positive if it results in cloaking $< k$ users; and in a false negative if it results in cloaking $\geq k$ users. A false negative is an indicator of over obfuscation which would in turn affect the utility of the obfuscated data; while a false positive is in direct violation of the k-anonymity security requirement. In order to determine the level of obfuscation I analyse historical data using decision tree based machine learning algorithms − parameterised by location (typically encoded as latitude/longitude boxes) and timestamps (typically time of day and week). I tradeoff

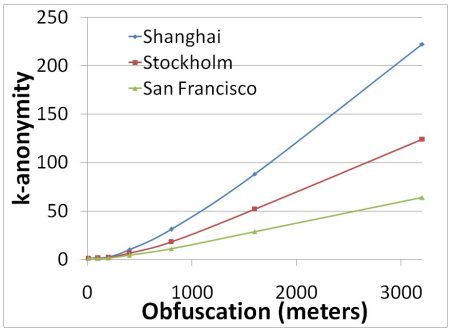

FIGURE 6.11: Average anonymity as the extent of obfuscation is varied for the time : 7-10am.

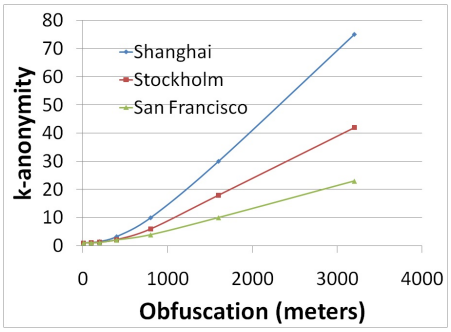

FIGURE 6.12: Average anonymity as the extent of obfuscation is varied for the time : 10am-4pm.

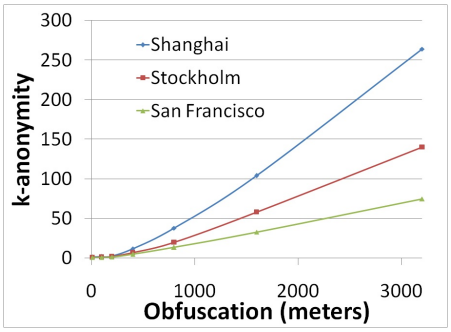

FIGURE 6.13: Average anonymity as the extent of obfuscation is varied for the time : 4-7pm.

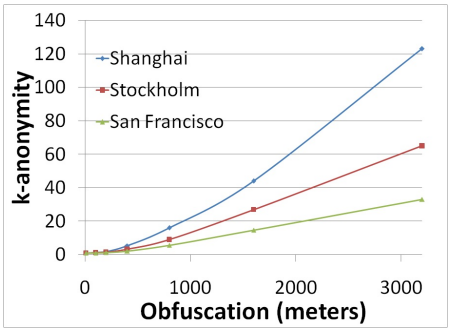

FIGURE 6.14: Average anonymity as the extent of obfuscation is varied for the time : 7pm-7am.

model complexity (i.e., number of nodes in the decision tree) with accuracy (i.e., being able to predict the desired level of obfuscation). It is clear that increasing model complexity beyond a desired level increases the error. primarily due to over fitting. I observe that in most cases the false positive and false negative rate of an optimal device-based algorithm (with large model complexity) varies between 0.12 and 0.25 for our datasets. This captures the extent of sub-optimality in a device-based solutions in comparison with an edge-based solution.

Figures 6.19, 6.20 and 6.21 show the false positive rate (i.e., the odds of not meeting the desired level of anonymity) and location error. Location error is only computed when the choice of obfuscation meets the desired level of anonymity. Otherwise, location error is computed as the difference between the extent of obfuscation chosen and the optimal obfuscation needed to achieve the desired level of anonymity.

Figures 6.22, 6.23 and 6.24 shows the false positive rate (i.e., the probability of not meeting the desired level of anonymity) and location error with and without consideration

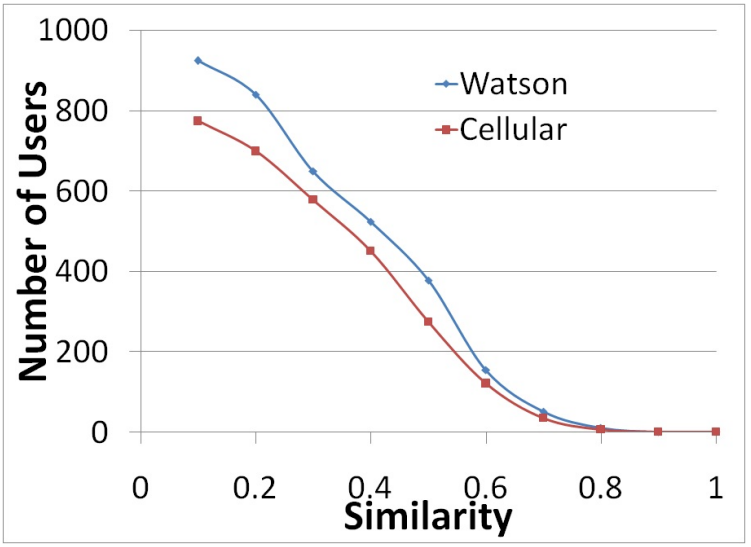

FIGURE 6.15: Similarity of user profiles (based on data accesses).

to user similarity respectively. For this experiment the desired level of anonymity $k = 16$ and the desired level of user similarity is 0.0 (first case that ignores user profiles), 0.7 (in the second case) and 0.9 (in the third case). For instance when user similarity threshold is 0.7, amongst the set of users that are within the extent of obfuscation only those users whose profiles are at least 70% similar to the given user are considered for quantification of anonymity. The above figures shows the additional cost (higher false positive rate and higher location error) that is incurred when enforcing location security based on profile cloning. I observe that when the similarity threshold is low the device-based solution pays a high penalty in terms of location error, while when the threshold is high the device-based solution pays a higher penalty in terms of false positive rate (i.e., the inability to meet the security requirement).

Figures 6.25, 6.26 and 6.27 show the false positive rate (i.e., the odds of not meeting the desired level of anonymity) and location error while requiring a user similarity threshold of 0.7. Profiles for entities are drawn at random from the `Watson` dataset with the goal of showcasing tradeoffs between location security and identity/profile based obfuscation. Similar to prior experiments, location error is only computed when the choice of obfuscation meets the desired level of anonymity. If the choice of obfuscation meets the desired level of anonymity and nothing more then location error is zero. Otherwise, location error is computed as the difference between the extent of obfuscation chosen and the optimal obfuscation needed to achieve the desired level of anonymity.

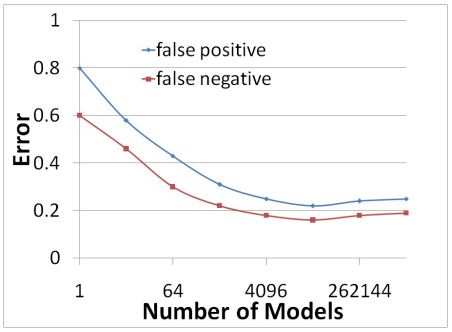

FIGURE 6.16: False positive/false negative rates of a device based model for Shanghai dataset.

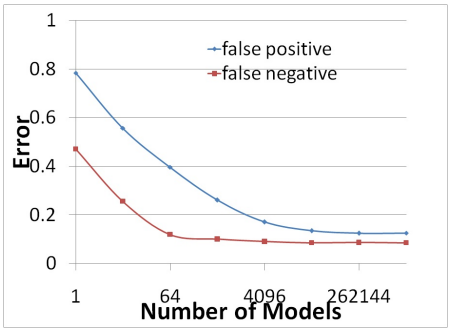

FIGURE 6.17: False positive/false negative rates of a device based model for Stockholm dataset.

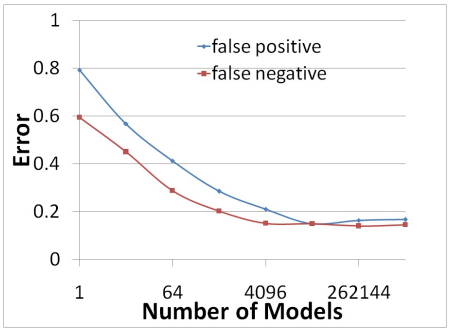

FIGURE 6.18: False positive/false negative rates of a device based model for San Francisco dataset.

This work builds upon the vast literature in location anonymity by investigating a large unexplored facet of this problem−where to enforce location security and what are the tradeoffs in doing so? I have explored both device and edge based enforcement of location security and quantified the gap between optimal device-based enforcement with that of the edge-based enforcement. In particular, I have identified machine learning algorithms that determine the extent of location obfuscation that is needed to achieve a desired level of anonymity. I have shown that even with good models a device based solution (that is unaware of the instantaneous locations of other entities or their profiles) is largely suboptimal in determining the extent of location obfuscation. The experiments on various mobility datasets show that device-based solutions either suffer from high false positive rate (about 25% chance of not meeting the desired security requirement) or low utility (about 600 meters higher error in obfuscated location data). The code for the enforcement of location privacy is made publicly available [138] .

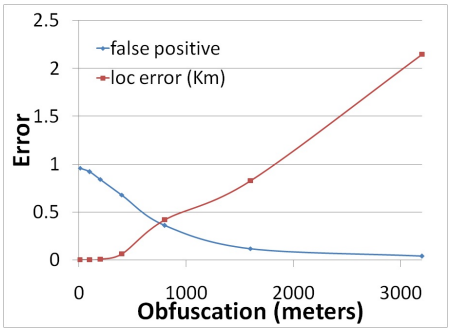

FIGURE 6.19: False positive rate and location error for Shanghai dataset.

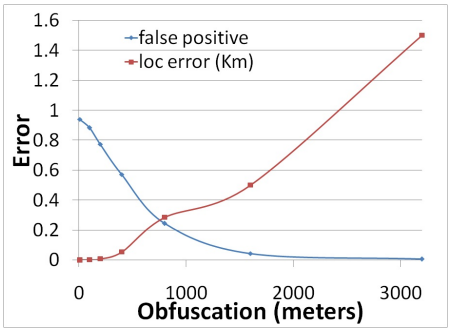

FIGURE 6.20: False positive rate and location error for Stockholm dataset.

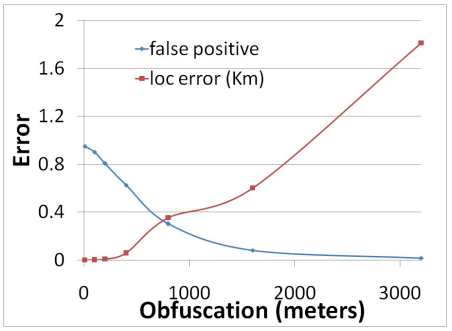

FIGURE 6.21: False positive rate and location error for San Francisco dataset.

6.7 Summary

This chapter describes the mechanisms used to enforce location privacy in the right place. It compares the solutions at the device level and at the core of the network and showcases that both the solutions have drawbacks. It proposes implementing location security and enforcing it at the edge of the network and explains the reasons for doing the same. It also showcases the anonymization methods including k-anonymity and s-proximity and explains how the location anonymizer replaces the exact location with the cloaked region to create a new anonymized location. Identity privacy enforcement is described in this chapter and the key steps involved in achieving the same has been shown through the chapter. This chapter also highlights the way the multiple users are mixed based on similar profiles and Chaumian mix. "Where to enforce location privacy" has been addressed through this thesis and this chapter shows an implementation of this solution. The chapter also shows the setup for enforcement of location privacy using a number of datasets. The experiments show that device-based solutions either suffer from high false positive rate or low utility.

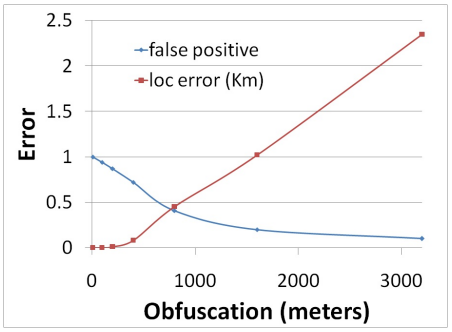

FIGURE 6.22: False positive rate and location error for Cellular dataset with Similarity threshold = 0.0.

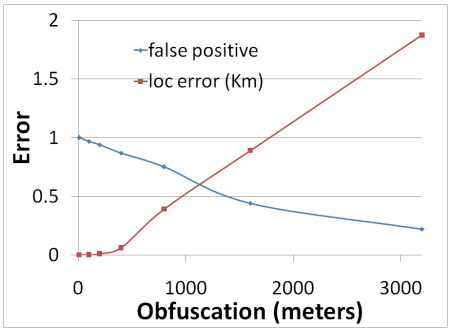

FIGURE 6.23: False positive rate and location error for Cellular dataset with Similarity threshold = 0.7.

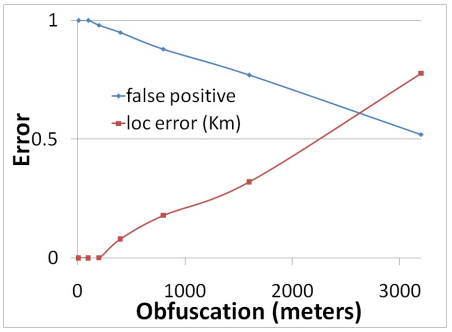

FIGURE 6.24: False positive rate and location error for Cellular dataset with Similarity threshold = 0.9.

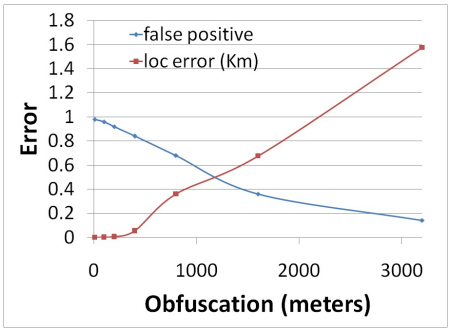

FIGURE 6.25: False positive rate and location error for Shanghai dataset with Similarity threshold = 0.7.

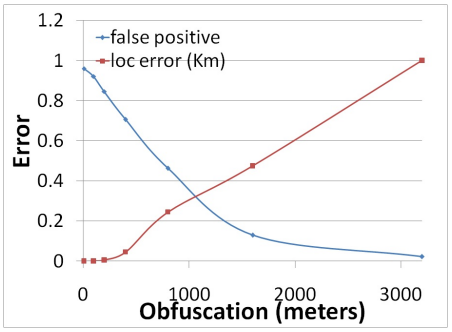

FIGURE 6.26: False positive rate and location error for Stockholm dataset with Similarity threshold = 0.7.

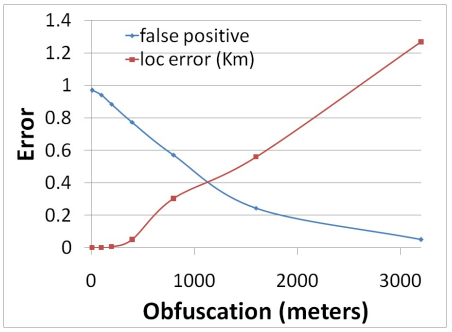

FIGURE 6.27: False positive rate and location error for San Francisco dataset with Similarity threshold = 0.7.

Chapter 7

Conclusion

7.1 Overview

In this thesis, I have addressed the open problems in the area of data access control and privacy in mobile environments. I have proposed new solutions for data access control using P3P and XACML extensions and GeoXACML. I have proposed novel solutions for the "where and when" to enforce location privacy and introduced a trust assessment framework for fused information. The overall thesis contributes towards better methodologies to ensure privacy in mobile environments.

This thesis details the methods and models of preserving privacy by introducing extensions to P3P and XACML. My research proposes a new model of using geospatial co-ordinates for geographical access control with the use of geohash and converting the XACML engine to GeoXACML. Through my research I have contributed novel solutions to resolve the problems of "where to enforce security" and "how to enforce security". Through my research I show a direct link between trust and privacy. Trust has been shown with a specific use case and evaluation of trust value and its importance in the overall context of fusing information and resolving conflicts has been explained through my work.

7.2 Contributions

I have focused mainly on looking at the policy extensions for preserving privacy in mobile environments and created new methods of data access control through extensions of P3P and XACML. I have shown geographical access control through GeoXACML with detailed description on the method, implementation and evaluation. With increasing volumes of data tagged with space and time (e.g., from smartphones) it is becoming increasingly important to support contextual (e.g., location-based) access control to data and resources. I have explored solutions to realise the GeoXACML access control model that allows a security administrator to specify location-based access control policies. The key novelty in my approach includes the ability to use geohashes to translate a GeoXACML policy into a conventional XACML policy - this allows us to fully reuse existing implementations of the XACML engine. I also describe a case study in the context of healthcare services wherein access control to handheld devices is moderated based on the location of the device. This is a novel contribution to the research as there is a new way of using XACML engine with the geographical extensions and geohashing to ensure that the information is received by the same person requesting for the information and that privacy is preserved. I have provided details on the implementation of the P3P and XACML extension for access control, GeoSpatial access control through GeoXACML showing the details on how the solution is novel. My research also details how global attestation of the location for mobile devices can be achieved using the trust matrix and calculations of EigenTrust. I have detailed the evaluation setup for policy based access control and highlighted the results of using 2 policy languages including P3P and XACML. It also shows the importance and benefits of using GeoXACML based access control model. This thesis also details the steps used for evaluating global attestation. In order to show this, 3 publicly available datasets including SFO taxi cab dataset, MIT reality dataset and Infocom06 datasets have been used. The results show the errors in location attestation under collusive and non-collusive settings and includes outputs from both honest and dishonest users. In general, the trust estimation error and location attestation error are acceptably small ($< 15\%$) when I have fewer than one-third dishonest entities.

I further discussed the main open challenges around 2 main questions, which are “Where to enforce Location privacy” and “How to Enforce Location Privacy” and provided

related work in the area and shown the limitations in the existing solutions in the area.

To address the "how", as per the privacy Vs utility tradeoff, when certain utility is required there needs to be a certain level of trust before providing the required information. To achieve this I have proposed a novel solution of Global Attestation of location which gives additional information compared to local attestation and improves the accuracy of the location information provided by the user requesting for location information services. My research has shown the implementation methods and the detailed evaluations done to prove the proposed model. Introducing global attestation through global consistency check is a novel contribution to the research community which has been possible through my work.

This research has also shown the problems that occur when there is a set of location information that comes in through streamed information that makes it difficult to derive at a consensus and introduced a new and novel model called the trust assessment model using Subjective logic and showed how the conflicts can be resolved when fused information is provided. I have proposed a new trust assessment framework that could be used to evaluate the trust on the fused information in a streaming setup. My research also discusses Subjective Logic operators and how these operators can be used to achieve a level of trust on fused location information. My work also describes the methods used to resolve conflicting opinions based on a number of mechanisms including trust based deleting, trust based discounting and evidence based discounting. Further it describes the implementation details for achieving trust on fused information using Information Fabric and CE store. This thesis evaluates the trust assessment model and reviewed the performance with the conflicts resolution. It can be observed that the service does not significantly alter the throughput of the platform but with better analytics the latency could be decreased. I also compare the accuracy of different classes of trust analytics operators.

My research work analysed the problem of "Where to enforce location privacy" and compared the solutions on the device and at the core and highlighted the problems with it. I have shown that by having the solution at the edge one can achieve acceptance latency with considerable security. My thesis explored both device and edge based enforcement of location security and quantified the gap between optimal device-based enforcement with that of the edge-based enforcement. In particular, this work identified

machine learning algorithms that determine the extent of location obfuscation that is needed to achieve a desired level of anonymity. I have shown that even with good models a device based solution (that is unaware of the instantaneous locations of other entities or their profiles) is largely suboptimal in determining extent of location obfuscation. The experiments on various mobility datasets show that device-based solutions either suffer from high false positive rate (about 25% chance of not meeting the desired security requirement) or low utility (about 600 meters higher error in obfuscated location data). I have shown that by having the solution at the edge one can achieve acceptance latency with considerable security. I have also detailed the anonymization methods including k-anonymity and s-proximity and explained how the location anonymizer replaces the exact location with the cloaked region to create a new anonymised location. My research includes Identity privacy enforcement and the key steps involved in achieving the same. I have highlighted how multiple users are mixed based on similar profiles and Chaumian mix. My research also details the setup for enforcement of location privacy using a number of datasets.

"Where and how to enforce location privacy" are the questions which has been addressed through this thesis and an implementation of this solution has been demonstrated in detail.

7.3 Future work

This thesis has presented new models and methods for achieving privacy in mobile environments. I have found some directions for research that originated through this work:

• Enhancing global attestation of location

It is interesting to evaluate the global attestation approach in real devices instead of experimentation over datasets. One would want to increase the feasibility of location change. For instance, a device cannot move from New York to London in a very short span of time. This information is useful when network is sparse. I use consistency of reports from different peers to assess positive and negative network feedback. On the other hand, in the case of network sparsity, I may not have sufficient number of feedback.

However, I can consider the previous location reports of a node to generate feedback. Future research could deal with such scenarios and enhance the solution.

• Variations in anonymisation methods

It would be interesting to design new anonymisation methods that would be capable of handling location privacy enforcement and identity privacy. Using fully homormorphic encryption methods, I am interested in creating new anonymisation solutions for location privacy.

• Design additional capabilities in trust assessment framework

A research area in trust assessment framework would be to consider handling information in other languages. This requires integration with other controlled languages. Research in extending my framework to integrate with other fusion operators that can handle malicious activities such as collision would be another area. In the case of collision, some sources come together and produce similar but deceptive opinions, which may mislead the system. This fusion operator could use trust models to handle these deceptive opinions. These trust models can also be used to estimate trust in sources when metadata is missing [131]. I use binomial opinions in our current framework, but there are cases when multinomial opinions are necessary to represent information [1]. Hence, it would be a good research topic to extend the framework in future towards these directions.

Bibliography

[1] A Jøsang. Subjective Logic : A critical tool in understanding and incorporating uncertainty in decision making, http://www.springer.com/us/book/9783319423357#aboutBook.

[2] U Varshney. Network access and security issues in ubiquitous computing. In *Workshop on Ubiquitous Computing Environment*, Cleveland, 2003.

[3] Ajay Brar and Judy Kay. Privacy and security in ubiquitous personalized applications. Technical report, http://www.it.usyd.edu.au/research/tr/tr561.pdf, 2004.

[4] M Karyda and S Gritzalis. Privacy and fair information practices in ubiquitous environments, research challenges and future directions. Technical report, Emerald, Internet Research Vol. 19 No. 2, Pages 194-208, 2009.

[5] M Hedenfalk A Escudero and P Heselius. Flying freedom, location privacy in mobile internetworking. Technical report, INET2001, 2001.

[6] Alastair R Beresford. *Location privacy in ubiquitous computing.* UCAM-CL-TR-612, ISSN 1476-2986, 2005.

[7] Marc Langheinrich. A privacy awareness system for ubiquitous computing environments. In *4th international conference on Ubiquitous Computing, Springer-Verlag*, pages 237–245, 2002.

[8] Danny Dolev and Andrew Yao. On the security of public key protocols. *IEEE Transactions on information theory*, 29(2):198–208, 1983.

[9] Jazilah Jamaluddin, Nikoletta Zotou, Reuben Edwards, and Paul Coulton. Mobile phone vulnerabilities: a new generation of malware. In *2004 IEEE International Symposium on In Consumer Electronics*, pages 199 – 202, 2004.

[10] Yu-Jia Chen and Li-Chun Wang. A security framework of group location-based mobile applications in cloud computing. In *Parallel Processing Workshops (ICPPW), 2011 40th International Conference*, pages 184–190, 2011.

[11] Collin Mulliner, Giovanni Vigna, David Dagon, and Wenke Lee. Using labeling to prevent cross-service attacks against smart phones. In *Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA)*, 2006.

[12] A Beresford and F Stajano. Mix zones: User privacy in location-aware services. In *IEEE Workshop on Pervasive Computing and Communication Security (PerSec)*, pages 127 – 131, 2004.

[13] Mobiloco: location based services for mobile communities. http://www.mobiloco.de. Accessed 2008.

[14] http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm, 2012.

[15] Sepandar D Kamvar, Mario T Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*, 2003.

[16] *Peertrust: Supporting reputation based trust for peer-to-peer electronic communities*, volume 16(7), 2004.

[17] Steven Bellman, Eric J Johnson, Stephen J Kobrin, and Gerald L Lohse. International differences in information privacy concerns: A global survey of consumers. In *The Information Society 20*, pages 313–324, 2004.

[18] A Westin. Commerce, communication, and privacy online for privacy & american business. privacy exchange. Technical report, H L Associates, http://www.privacyexchange.org/iss/surveys/computersurvey97.html. Accessed 2015.

[19] R K Chellappa. The role of perceived privacy and perceived security. Technical report, Consumers' Trust in Electronic Commerce Transactions, http://asura.usc.edu/ram/rcf-papers/secpriv.pdf. Accessed 2015.

[20] E A Kaluscha S Grabner. Empirical research in on-line trust: a review and critical assessment. Technical report, Academic Press, 2003.

[21] Federal Trade Commission. Privacy online: A report to congress. Technical report, http://www.ftc.gov/reports/privacy3/priv-23a.pdf. Accessed 2015.

[22] Annie Anton, Julie B Earp, Davide Bolchini, Qingeng He, Carlos Jensen, and William Stufflebeam. The lack of clarity in financial privacy policies and the need for standardization. Technical report, http://www.theprivacyplace.org/papers/glbsecPrivtr.pdf. Accessed 2015.

[23] N Hain W H Stufflebeam, A I Ant. Specifying privacy policies with p3p and epal: Lessons learned. Technical report, Proceedings of the 2004 ACM workshop on Privacy in the electronic society (WPES), 2004.

[24] R J Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. 2001.

[25] J Cannon. *Privacy : What Developers and IT Professionals Should Know*. J Cannon, 2004.

[26] L F Cranor. *Web Privacy with P3P*. Sebastopol, CA, USA, 2002.

[27] D Weerasinghe M Rajarajan and V Rakocevic. Device data protection in mobile healthcare applications. In *The First International Conference on Electronic Healthcare in the 21st century*, 2008.

[28] L Cranor M Langheinrich M Marchiori and J Reagle. The platform for privacy preferences 1.0 (p3p1.0) specification. In *W3C recommendation*, 2002.

[29] A Carlisle Q Xuebing. Xacml-based policy- driven access control for mobile environments. In *Canadian Canadian Conference on Electrical and Computer Engineering*, 2006.

[30] A Jøsang and S L Presti. Analysing the relationship between risk and trust. In *Proceedings of the Second International Conference on Trust Management (iTrust)*, 2004.

[31] Liang Cai, Sridhar Machiraju, and Hao Chen. Defending against sensor-sniffing attacks on mobile phones. In *1st ACM workshop on Networking, systems, and applications for mobile handhelds New York*, pages 31–36, 2009.

[32] Federico Cerutti, Supriyo Chakraborty, Geeth R de Mel, Lance M Kaplan, Timothy J Norman, Nir Oren, Stephen Pipes, Murat Sensoy, Mani B Srivastava, and Paul Sullivan. Managing information sharing in coalitions through credible obfuscation. Technical report, ITA, 2014. URL https://www.usukitacs.com/node/2770.

[33] Shiladri Chakraborty, Nicolas Bitouzé, Mani Srivastava, and Lara Dolecek. Protecting data against unwanted inferences. In *Information Theory Workshop (ITW), 2013 IEEE*, pages 1–5. IEEE, 2013.

[34] http://www.bbc.co.uk/news/business-17192234, 2012.

[35] http://www.bbc.co.uk/news/technology-17178954, 2012.

[36] M Kulicke R Fedler, J Schutte. On the effectiveness of malware protection on android. Technical report, Tech. rep., Technical report, Fraunhofer AISEC, Berlin, 2013.

[37] Chowdhury S Hasan, Sheikh I Ahamed, and Mohammad Tanviruzzaman. A privacy enhancing approach for identity inference protection in location-based services. In *33rd Annual IEEE International Computer Software and Applications Conference*, 2009.

[38] C Y Chow M F Mokbel and W G Aref. The new casper: Query processing for location services without compromising privacy. In *Proceedings of Very Large DataBases (VLDB)*, 2006.

[39] K Puttaswamy and B Zhao. Preserving privacy in location-based mobile social applications. In *HotMobile*, pages 1–6, 2010.

[40] S A Menesidou A Loukas, D Damopoulos, M E Skarkala, G Kambourakis, and S Gritzalis. Milc: A secure and privacy-preserving mobile instant locator with chatting. In *Springer Science plus Business Media, LLC*, 2010.

[41] J Freudiger R Shokri and J P Hubaux. A unified framework for location privacy. In *Hot Topics in Privacy Enhancing Technologies (HotPETS)*, 2010.

[42] W X Sean W Song. In-device spatial cloaking for mobile user privacy assisted by the cloud. In *Eleventh International Conference on Mobile Data Management (MDM)*, pages 381 – 386, 2010.

[43] R Wei Z Liang. Efficient k-anonymization for privacy preservation. In *Computer Supported Cooperative Work in Design, 12th International Conference*, pages 737 – 742, 2008.

[44] Daniele Riboni, Linda Pareschi, Claudio Bettini, and Sushil Jajodia. Preserving anonymity of recurrent location-based queries. In *16th International Symposium on Temporal Representation and Reasoning, IEEE Computer Society*, 2009.

[45] Dr V Kavitha P Deivanai, Mrs J Jesu Vedha Nayahi. A hybrid data anonymization integrated with suppression for preserving privacy in mining multi party data. In *IEEE-International Conference on Recent Trends in Information Technology, ICRTIT*, 2011.

[46] J Vaidya H Shin, V Atluri. A profile anonymization model for privacy in a personalized location based service environment. In *9th International Conference on Mobile Data Management. MDM'08*, pages 73–80, 2008.

[47] GFindster. Gfindster. http://www.androidapps.com/t/gfindster. Accessed 2008.

[48] IMEasy. Introduction to hi aim. http://im-easy.com/. Accessed 2008.

[49] BuddyMob. Buddymob. http://www.buddymob.com/. Accessed 2009.

[50] L Kagal A Patwardhan, V Korolev and A Joshi. Enforcing policies in pervasive environments. In *International Conference on Mobile and Ubiquitous Systems: Networking and Services*, 2004.

[51] Lalana Kagal, Tim Finin, and Anupam Joshi. A policy based approach to security for the semanticweb. In *2nd International Semantic Web Conference (ISWC2003)*, 2003.

[52] Lalana Kagal, Tim Finin, and Anupam Joshi. A policy language for a pervasive computing environment. In *IEEE 4th International Workshop on Policies for Distributed Systems and Networks*, 2003.

[53] Ian Horrocks, Peter F Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean, et al. Swrl: A semantic web rule language combining owl and ruleml. http://www.daml.org/rules/proposal/.

[54] S. Bechhofer. Hoolet swrl reasoner. http://owl.man.ac.uk/hoolet/.

[55] Vincent C Hu, D Richard Kuhn, and David F Ferraiolo. Attribute-based access control. *IEEE Computer*, 48(2):85–88, 2015.

[56] Sun-Moon Jo and Kyung-Yong Chung. Design of access control system for telemedicine secure xml documents. *Multimedia Tools and Applications*, 74(7): 2257–2271, 2015.

[57] M Zuidweg J G Pereira Filho and M van Sinderen. Using p3p in a web services-based context-aware application platform. In *Proceedings of EUNICE 2003 9th Open European Summer School and IFIP WG6.3 Workshop on Next Generation Networks*, 2003.

[58] L Cranor. M Langheinrich and M Marchiori. A p3p preference exchange language 1.0 (appel 1.0). Technical report, World Wide Web Consortium,, 2005.

[59] G Myles A Friday and N Davies. Preserving privacy in environments with location-based applications. Technical report, IEEE Pervasive Computing, 2003.

[60] M Fahrmair W Sitou and B Spanfelner. Security and privacy rights management for mobile and ubiquitous computing. In *Workshop on UbiComp Privacy*, 2005.

[61] P D Giang L X Hung R A Shaikh Y Zhung S Lee Y Lee and H Lee. A trust-based approach to control privacy exposure in ubiquitous computing environments. In *IEEE International Conference on Pervasive Services*, 2007.

[62] Y Wang J Zhang and V Varadharajan. Mobile agent and web service integration security architecture. In *IEEE International Conference on Service-Oriented Computing and Applications, SOCA*, 2007.

[63] Alessandro Giambruno, Muhammad Awais Shibli, Sead Muftic, and Antonio Lioy. Magicnet: Xacml authorization policies for mobile agents. In *Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for*, pages 1–7. IEEE, 2009.

[64] T Karygiannis S G W A Jansen and V Korolev. Assigning and enforcing security policies on handheld devices. In *Proceedings of the Canadian Information Technology Security Symposium*, 2002.

[65] T Karygiannis M I S G W Jansen and V Korolev. Security policy management for handheld devices. In *The 2003 International Conference on Security and Management(SAM'03)*, 2003.

[66] T Karygiannis V K W A Jansen and S G. Policy expression and enforcement for handheld devices. Technical report, National Institute of Standards and Technology -NIST, 2003.

[67] A Raghavendra D Weerasinghe, S Arunkumar and M Rajarajan. Policy extension for data access control. Technical report, In 6th IEEE workshop on Secure Network Protocols (NPSec),, 2010.

[68] S Arunkumar and M Rajarajan. Healthcare data access control using xacml for handheld devices. Technical report, In Developments in E-systems Engineering (DESE), IEEE, 2010.

[69] Geospatial extensible access control markup language (geoxacml). http://www.w3.org/Policy/pling/wiki/images/5/59/GeoXACML.pdf, 2008.

[70] Opengis geospatial extensible access control markup language (geoxacml) revision working group. http://www.opengeospatial.org/projects/groups/gxacmlrwg, 2007.

[71] John J Roese, Richard W Graham, David Frattura, and David Harrington. Location-based access control in a data network, March 3 2015. US Patent 8,972,589.

[72] Naeim Abedi, Ashish Bhaskar, and Edward Chung. Tracking spatio-temporal movement of human in terms of space utilization using media-access-control address data. *Applied Geography*, 51:72–81, 2014.

[73] Ai-juan Zhang, Jing-xiang Gao, JI Cheng, Jiu-yun Sun, and BAO Yu. Multi-granularity spatial-temporal access control model for web gis. *Transactions of Nonferrous Metals Society of China*, 24(9):2946–2953, 2014.

[74] S Saroiu and A Wolman. Enabling new mobile applications with location proofs. In *ACM Hotmobile*, 2009.

[75] W Luo and U Hengartner. Veriplace: a privacy-aware location proof architecture. In *In ACM GIS*, 2010.

[76] Z Zhu and G Cao. Towards privacy-preserving and colludingresistance in location proof updating system. In *In IEEE Transactions on Mobile Computing*, 2011.

[77] Xinlei Wang, Jindan Zhu, Amit Pande, Arun Raghuramu, Prasant Mohapatra, Tarek Abdelzaher, and Raman Ganti. Stamp: Ad hoc spatial-temporal provenance assurance for mobile users. In *Network Protocols (ICNP), 2013 21st IEEE International Conference on*, pages 1–10. IEEE, 2013.

[78] R Hasan and R Burns. Where have you been? secure location provenance for mobile devices. In *In CoRR*, 2011.

[79] H C B Davis and M Franklin. Privacy preserving alibi systems. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, 2012.

[80] A. Joshi A Patwardhan, F Perich and Y Yesha. Active collaborations for trustworthy data management in ad hoc networks. In *Proceedings of the 2nd IEEE International Conference on Mobile Ad-Hoc and Sensor Systems*, 2005.

[81] Hyo-Sang Lim, Yang-Sae Moon, and Elisa Bertino. Provenance-based trustworthiness assessment in sensor networks. In *Proceedings of the Seventh International Workshop on Data Management for Sensor Networks*, pages 2–7. ACM, 2010.

[82] D Mott. Summary of ita controlled english. Technical report, ITA Tech. Report, https://www.usukita.org/papers/5658/details.html, 2012.

[83] Joel Wright, Christopher Gibson, Flavio Bergamaschi, Kelvin Marcus, Ryan Pressley, Gunjan Verma, and Gene Whipps. A dynamic infrastructure for interconnecting disparate isr/istar assets (the ita sensor fabric). In *Information Fusion, 2009. FUSION'09. 12th International Conference on*, pages 1393–1400. IEEE, 2009.

[84] Shuiqing Yang. Role of transfer-based and performance-based cues on initial trust in mobile shopping services: a cross-environment perspective. *Information Systems and e-Business Management*, 14(1):47–70, 2016.

[85] Mehrbakhsh Nilashi, Othman Ibrahim, Vahid Reza Mirabi, Leili Ebrahimi, and Mojtaba Zare. The role of security, design and content factors on customer trust in mobile commerce. *Journal of Retailing and Consumer Services*, 26:57–69, 2015.

[86] R Schwitter. Handling defaults and their exceptions in controlled natural language. In *Language Production, Cognition, and the Lexicon, Springer*, 2015.

[87] A Jøsangand R Ismail. The beta reputation system. In *15th Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy*, 2002.

[88] A Jøsang A Whitby and A Indulska. Filtering out unfair ratings in bayesian reputation systems. *The Icfain Journal of Management Research*, 2005.

[89] N Jennings W Teacy, J Patel and M Luck. Travos: Trust and reputation in the context of inaccurate information sources. In *Autonomous Agents and Multi-Agent Systems,*, 2006.

[90] S Chakraborty L Kaplan, M Sensoy and G de Mel. Partial observable update for subjective logic and its application for trust estimation. In *Information Fusion, Elsevier*, 2015.

[91] B Yu and M Singh. Detecting deception in reputation management. In *Proceedings of Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2003.

[92] J Golbeck and C Halaschek-Wiener. Trust-based revision for expressive web syndication. *Journal of Logic and Computation,*, 19(5):771–790, 2009.

[93] L Berti-Equille X L Dong and D Srivastava. Integrating conflicting data: The role of source dependence. *35th International Conference on Very Large Databases, Lyon, France.*, 2009.

[94] J Han X Yin and P S Yu. Truth discovery with multiple conflicting information providers on the web. In *Proceedings of the Conference on Knowledge and Data Discovery.*, 2007.

[95] Roman Schlegel, Kehuan Zhang, Xiao-yong Zhou, Mehool Intwala, Apu Kapadia, and XiaoFeng Wang. Soundcomber: A stealthy and context-aware sound trojan for smartphones. In *18th Annual Network and Distributed System Security Symposium (NDSS)*, pages 17–33, 2011.

[96] William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth.

Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In *In OSDI'10: Proceedings of the 9th USENIX Symposium on Operating Systems Design and Implementation*, 2010.

[97] Lei Liu, Guanhua Yan, Xinwen Zhang, and Songqing Chen. Virusmeter: Preventing your cellphone from spies. In *In E. Kirda, S. Jha, and D. Balzarotti, editors, RAID,*, pages 244–264. Springer volume 5758 of Lecture Notes in Computer Science, 2009.

[98] Jian Liao, Ying-hao Qi, Pei-wei Huang, Meng-tian Rong, and Sheng-hong Li. Protection of mobile location privacy by using blind signature. *In Journal of Zhejiang University – Science A 7(6)*, pages 984–989, 2006.

[99] Qi He, Dapeng Wu, and Pradeep Khosla. Quest for personal control over mobile location privacy. *IEEE Communications Magazine, 42(5)*, pages 130–136, 2004.

[100] D Wu H Qi and P Khosla. A mechanism for personal control over mobile location privacy. In *Proceedings of IEEE/ACM First International Workshop on Broadband Wireless Services and Applications, BroadWISE*, 2004.

[101] Sheng Zhong, L Li, Yanbin Grace Liu, and Yang Richard Yang. Privacy-preserving location-based services for mobile users in wireless networks. Technical report, Technical report, State University of New York, 2005.

[102] L Barkhuus and A K Dey. Location-based services for mobile telephony: a study of user's privacy concerns. In *International Conference on Human-Computer Interaction, Switzerland*, 2003.

[103] A Escudero-Pascual T Holleboom and S Fischer-Hubner. Privacy for location data in mobile networks. In *Nordic Security Workshop, NORDSEC*, 2002.

[104] S Skiadopoulos G Ghinita1, P Kalnis. Mobihide: A mobile peer-to-peer system for anonymous location-based queries. In *Proceedings of the International Symposium on Advances in Spatial and Temporal Databases, SSTD*, 2007.

[105] L Ling B Gedik. Location privacy in mobile systems: A personalized anonymization mode. In *Proceedings of 25th IEEE International Conference on Distributed Computing Systems (ICDCS 2005)*, pages 620–629, 2005.

[106] R P Minch. Privacy issues in location-aware mobile devices. In *Hawaii International Conference on System Sciences*, pages 1–10, 2004.

[107] Carl A Gunter, Michael J May, and Stuart G Stubblebine. A formal privacy systems and its application to location-based services. In *Proceedings of Workshop on Privacy Enhancing Technologies, Canada,*, 2004.

[108] Web services description language (wsdl) version 2.0 part 1: Core language. http://www.w3.org/TR/wsdl20/, 2007.

[109] Oasis extensible access control markup language (xacml) version 2.0 oasis standard. http://docs.oasisopen.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf, 2005.

[110] Zoran Balkić, Damir Šoštarić, and Goran Horvat. Geohash and uuid identifier for multi-agent systems. In *Agent and Multi-Agent Systems. Technologies and Applications*, pages 290–298. Springer, 2012.

[111] John Gilliom and Torin Monahan. *Supervision: an introduction to the Surveillance Society*. University of Chicago Press, 2012.

[112] Code for geoxacml based access control. https://www.usukitacs.com/node/3017, 2015.

[113] S Nakamoto. Bitcoin: A peer-to-peer electronic cash system. In *Consulted, 1(2012):28*, 2008.

[114] J DuPont and A C Squicciarini. Toward deanonymizing bitcoin by mapping users location. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, 2015.

[115] F Z Filali and B Yagoubi. Global trust: A trust model for cloud service selection. In *Computing, 3(18):19*, 2015.

[116] S Moalla and M Rahmouni. Trust path: a distributed model of search paths of trust in a peer-to-peer system. In *Security and Communication Networks*, volume 8(3), pages 360–367, 2015.

[117] San francisco taxicab data set. http://crawdad.org/epfl/mobility/20090224/, 2009.

[118] Nathan Eagle and Alex (Sandy) Pentland. Reality mining: sensing complex social systems. Technical report, Journal Personal and Ubiquitous Computing Volume 10 Issue 4, March 2006 Pages 255 - 268, 2006.

[119] Infocom 2006 data set. http://crawdad.org/cambridge/haggle/20090529/, 2009.

[120] B Brehmer. The dynamic ooda loop: Amalgamating boyd's ooda loop and the cybernetic approach to command and control. In *Proceedings of the 10th international command and control research technology symposium*, 2005.

[121] Murat Sensoy, Chatschik Bisdikian, Nir Oren, Chris Burnett, Timothy J Norman, Mani B Srivastava, and Lance M Kaplan. Trust and obfuscation. In *SPIE Defense, Security, and Sensing Symposium, Baltimore, MD, USA*, 2012.

[122] Chatschik Bisdikian, Lance M Kaplan, Mani B Srivastava, David J Thornley, Dinesh Verma, and Robert I Young. Building principles for a quality of information specification for sensor information. In *In 12th Int'l Conf. on Information Fusion*, 2009.

[123] L Kaplan C Bisdikian and M B Srivastava. Quality of information in sensor networks. In *In IBM Res. Rep. RC25254*, 2011.

[124] Shiguang Wang, Lu Su, Shen Li, Shaohan Hu, Tanvir Amin, Hongwei Wang, Shuochao Yao, Lance Kaplan, and Tarek Abdelzaher. Scalable social sensing of interdependent phenomena. In *In 14th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN), Seattle, WA*, 2015.

[125] J Konstan J B Schafer and J Riedl. Recommender systems in e-commerce. In *In ACM Conference on Electronic Commerce*, 1999.

[126] U Kuter and J Golbeck. Sunny: A new algorithm for trust inference in social networks, using probabilistic confidence models. In *Association for the Advancement of Artificial Intelligence, AAAI-07*, 2007.

[127] A Jøsang. A logic for uncertain probabilities. In *Int. J. Uncertain. Fuzziness Knowl.-Based Syst*, volume 9(3), pages 279–311, 2001.

[128] A Jøsang. The consensus operator for combining beliefs. In *Artificial Intelligence*, volume 141(1/2), pages 157–170, 2002.

[129] A Jøsang and T Grandison. Conditional inference in subjective logic. In *Proceedings of the 6th International Conference on Information Fusion.*, 2003.

[130] A Jøsang and D McAnally. Multiplication and comultiplication of beliefs. In *Int. J. Approx. Reasoning, 38(1)*, volume 38(1), pages 19–51, 2005.

[131] A Jøsang. Probabilistic logic under uncertainty. In *Proceedings of the thirteenth Australasian symposium on Theory of computing*, pages 101–110, 2007.

[132] A Jøsang. Conditional reasoning with subjective logic. In *of Multiple-Valued Logic and Soft Computing*, volume 15(1), pages 5–38, 2008.

[133] C Bisdikian S Chakraborty, M B Srivastava and R Ganti. Localization in cognitive radio systems in the presence of spatially obfuscated data. Technical report, Technical Report TR-UCLANESL- 201205-01, NESL, 2012.

[134] Murat Sensoy, Geeth de Mel, Lance Kaplan, Thach Pham, and Timothy J Norman. Tribe: Trust revision for information based on evidence. In *Proceedings of 16th International Conference on Information Fusion*, pages 914–921, 2013.

[135] Jacob L Graham, David L Hall, and Jeffrey Rimland. A coin-inspired synthetic dataset for qualitative evaluation of hard and soft fusion systems. In *In 14th Int'l Conf. on Information Fusion (FUSION 2011), Chicago, IL, USA*, 2011.

[136] Shiqiang Wang, Guan-Hua Tu, Raghu Ganti, Ting He, Kin Leung, Howard Tripp, Katy Warr, and Murtaza Zafer. Mobile micro-cloud: Application classification, mapping, and deployment. In *Proceedings of Annual Fall Meeting of ITA (AMITA)*, 2013.

[137] Shanghai data set. http://crawdad.org/keyword-vehicular-network.htm/, 2012.

[138] Code for geoxacml based access control. https://www.usukitacs.com/node/3018, 2015.