Engineering Trusted Location Services and Context-aware Augmentations for Network Authorization Models

by

Christian John Wullems

BIT(Hons) *QUT*

Thesis submitted in accordance with the regulations for Degree of Doctor of Philosophy

Information Security Research Centre
Faculty of Information Technology
Queensland University of Technology

March 2004

Keywords

Pervasive, ubiquitous, authorization, access control, location, trusted location, context, context-aware, network security.

Abstract

Context-aware computing has been a rapidly growing research area, however its uses have been predominantly targeted at pervasive applications for smart spaces such as smart homes and workplaces. This research has investigated the use of location and other context data in access control policy, with the purpose of augmenting existing IP and application-layer security to provide fine-grained access control and effective enforcement of security policy. The use of location and other context data for security purposes requires that the technologies and methods used for acquiring the context data are trusted.

This thesis begins with the description of a framework for the analysis of location systems for use in security services and critical infrastructure. This analysis classifies cooperative locations systems by their modes of operation and the common primitives they are composed of. Common location systems are analyzed for inherent security flaws and limitations based on the vulnerability assessment of location system primitives and the taxonomy of known attacks.

An efficient scheme for supporting trusted differential GPS corrections is proposed, such that DGPS vulnerabilities that have been identified are mitigated. The proposal augments the existing broadcast messaging protocol with a number of new messages facilitating origin authentication and integrity of broadcast corrections for marine vessels.

A proposal for a trusted location system based on GSM is presented, in which a model for tamper resistant location determination using GSM signaling is designed. A protocol for association of a user to a cell phone is proposed and demonstrated in a framework for both Web and Wireless Application Protocol (WAP) applications. After introducing the security issues of existing location systems and a trusted location system proposal, the focus of the thesis changes to the use of location data in authorization and access control processes. This is considered at both the IP-layer and the application-layer.

For IP-layer security, a proposal for location proximity-based network packet filtering in IEEE 802.11 Wireless LANs is presented. This proposal details an architecture that extends the Linux netfilter system to support proximity-based packet filtering, using methods of transparent location determination through the application of a pathloss model to raw signal measurements.

Our investigation of application-layer security resulted in the establishment of a set of requirements for the use of contextual information in application level authorization. Existing network authentication protocols and access control mechanisms are analyzed for their ability to fulfill these requirements and their suitability in facilitating context-aware authorization. The result is the design and development of a new context-aware authorization architecture, using the proposed modifications to Role-based Access Control (RBAC). One of the distinguishing characteristics of the proposed architecture is its ability to handle authorization with context-transparency, and provide support for real-time granting and revocation of permissions.

During the investigation of the context-aware authorization architecture, other security contexts in addition to host location were found to be useful in application level authorization. These included network topology between the host and application server, the security of the host and the host execution environment. Details of the prototype implementation, performance results, and context acquisition services are presented.

Contents

List of Figures

List of Tables

Declaration

The work contained in this thesis has not been previously submitted for a degree or diploma at any higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signed: . . . . **Date:** . . . .

Previously Published Material

The following papers have been published or presented, and contain material based on the content of this thesis.

[1] Chris Wullems, Mark Looi, and Andrew Clark. Enhancing the Security of Internet Applications using Location: A New Model for Tamper-resistant GSM Location. In *Proceedings of the Eighth IEEE International Symposium on Computers and Communications (ISCC 2003)*, volume 1, pages 1251–1258, Antalya, Turkey, July 2003. IEEE Computer Society, Los Alamitos, CA, USA.

[2] Chris Wullems, Oscar Pozzobon, Mark Looi, and Kurt Kubik. Enhancing the Trust of Location Acquisition Systems for Critical Applications and Location-based Security Services. In *Proceedings of the Forth Australian Information Warfare and Security Conference (AIWSC 2003)*, pages 391–405, Adelaide, Australia, November 2003.

[3] Chris Wullems, Mark Looi, and Andrew Clark. Towards Context-aware Security: An Authorization Architecture for Intranet Environments. In *Proceedings of the First IEEE International Workshop on Pervasive Computing and Communication Security (PerSec 2004)*, pages 132–137, Orlando, FL, USA, March 2004. IEEE Computer Society, Los Alamitos, CA, USA.

[4] Chris Wullems, Kevin Tham, and Mark Looi. Proximity-Based Network Packet Filtering For IEEE 802.11 Wireless Devices. In *IADIS International Conference on Applied Computing*, Lisbon, Portugal, March 2004.

[5] Chris Wullems, Kevin Tham, Jason Smith, and Mark Looi. A Trivial Denial of Service Attack on IEEE 802.11 Direct Sequence Spread Spectrum Wireless LANs. In *Third IEEE Wireless Telecommunications Symposium (WTS 2004)*, pages 129–136, Pomona, CA, USA, May 2004. IEEE Computer Society, Los Alamitos, CA, USA.

[6] Chris Wullems, Harikrishna Vasanta, Mark Looi, and Andrew Clark. A Broadcast Authentication and Integrity Augmentation for Trusted Differential GPS in Marine Navigation. In *Workshop on Cryptographic Algorithms and their Uses*, pages 125–139, Gold Coast, QLD Australia, July 2004.

Acknowledgements

First I would like to express gratitude for the support and guidance my principal supervisor, Associate Professor Mark Looi has given me. Without his support and vision, this thesis would not have been possible. Second, I would like to gratefully acknowledge the support of my secondary supervisor, Dr. Andrew Clark who has also given much support and guidance during the course of my research.

Some of the research presented in this thesis was performed jointly with other postgraduate students. I would like to gratefully acknowledge Oscar Pozzobon, who in joint work with University of Queensland, contributed to the vulnerability analysis of GPS and the generalized location models in Chapter 2.

Special thanks goes to the members of the Network and Systems Security Group of the ISRC, for their support, friendship and insightful discussions. I would especially like to mention those whom I shared an office with at Margaret St., Harikrishna Vasanta, John Holford, Kevin Tham and Jason Smith.

I gratefully acknowledge the assistance provided by Kevin in developing the software and testing the proximity-based wireless LAN prototype in Chapter 5, and both Kevin and Jason for their assistance in investigating the properties of the Wireless LAN denial of service attack described in Chapter 6.

Without the support of the secure network laboratory staff, much of the development and testing would not have been possible. Many thanks to Sam Lor, Ricco Lee and Matt Bradford for their support and friendship.

I would like to thank those whom I have had the privilege of working with and knowing at the ISRC, Professor Ed. Dawson, Associate Professor Colin Boyd, Dr. Ernest Foo, Dr. Juanma Gonzalez Nieto, Jason Reid, Mina Yao, Terry Tin, Riza Aditya, Pal-Erik Martinsen, Lou Tang Seet, Yvonne Hitchcock, Dr. Greg Maitland, Dr. Kapali Viswanathan, Roslan Ismail, and Stig Andersson. Also special thanks to the administration staff, Christine Orme and Elizabeth Hansford who provided much support for the endless amounts of administrative paper work and university bureaucracy.

Finally, I would like to thank my parents Min and Rossana, my sisters Elizabeth and Lydia, and my extended family both in Australia and Italy, for their support and patience during the three years of my research.

This thesis is dedicated to my late grandmother, Innalaura Speltoni, who gave me much moral support during the first two years of my research

Chapter 1

Introduction

The overall goal of network security is to prevent misuse of resources, and where this is not feasible, the detection and recovery from misuse. Prevention of misuse is an increasingly difficult objective to achieve, given that anticipating and defending against all possible missuses is virtually impossible. The research presented in this thesis introduces a new vision for network security, "context-aware access control". Context-aware access control is a concept whereby hosts are granted or denied access to resources based on the perceived security of the host. In this thesis, we examine context-aware access control at two layers, the IP layer and the application layer.

Security contexts that could be considered include:

- *Host location.* This includes instances where access to a given resource or operation would constitute a breach of security if outside of a trusted location;
- *Network topology between host and application server.* This includes contexts such as connection security, bandwidth, and routing;
- *Host security.* This includes contexts such as operating system patch level, antivirus signature version, host firewall rules, routing tables and filesystem permissions / file-sharing settings; and
- *Host execution environment.* This includes contexts such as currently executing applications that have conflicts of interest and validating the state of executing applications such as resident virus scanners.

This thesis has a significant focus on the engineering of trusted location systems.

Traditionally, IP-layer access control is implemented in routers to manage the security of gateway and network perimeter traffic. Access control is typically managed through the use of Access Control Lists (ACLs), which provide a means to filter packets by allowing a user to permit or deny IP packets from traversing over specified interfaces.

Consider, for example, the situation where wireless clients connect remotely to a private network. Traditional security technologies such as Virtual Private Networks (VPNs) attempt to protect privacy and integrity of data transmitted over the air, and authenticate the client that is connecting. What traditional IP-layer security does not provide, is protection from the client's host. A client's host may be multi-homed and as such, an attacker may use the secure VPN tunnel established by a legitimate client to gain access to the private network from another interface. In addition, a client's host may contain viruses or Trojans that can pose a significant threat to the private network. Contextual IP-layer security would be able to mitigate such situations.

Context-aware access control implemented in a router, could provide higher levels of access to clients that are compliant with a given security policy. For clients that are not compliant, the router could allow them to use a restricted subset of services in a sandbox, such that vulnerabilities on the client's host would have a negligible impact on the private network.

Access control at the application layer is able to provide finer-grained access control than at the IP layer, such that access to specific resources can be controlled where required. In an application-layer access control system, objects are protected through the use of reference monitors that mediate all access to protected objects. An authenticated user with the appropriate security attributes would be able perform a set of authorized operations on a given object.

Consider the example where special access requirements exist for sensitive “In Strictest Confidence” information. Procedural and operational policy may specify particular locations, times and host security requirements for this information to be accessed. Context-awareness in application layer access control process would facilitate enforcement of such a policies that would not be enforceable using traditional access control mechanisms.

The research presented in this thesis additionally proposes an application-layer authorization architecture that uses context information to augment existing and time-tested authentication technologies such as Kerberos, facilitating effective enforcement of security policies through fine-grained access control to network resources and support for context information.

1.1 Goals

The research in this thesis was initiated to investigate the use of contextual information in access control processes at the IP and application layers. Special emphasis was placed on the context of location, identifying issues with its use for security applications and methods to facilitate trust. This goal was achieved through the following objectives:

1. *The development of a framework for the analysis of location systems for security services.* This purpose of this objective is to identify common components of location systems and the desirable security properties of these components. This provides us with a basis to perform vulnerability analysis of existing location systems. A set of requirements for the common components of location systems is to be established, such that a platform is provided from which to design new secure location systems.

2. *The development of various methods to increase the trust of existing location systems in order to support their use in security systems.* Having established an analysis framework and a set of requirements for secure trusted location, methods are to be developed to enhance existing location technologies to provide the security properties required for trusted location.

3. *The design of IP layer access control processes that support context-awareness.* Integration of location context data into IP layer access control is investigated. We peruse the development of a system that performs acquisition of location and uses this data to augment existing IP layer network security mechanisms.

4. *The investigation of existing access control models and their suitability for supporting context-awareness.* Integration of context-awareness into application-layer access control processes requires the investigation of access control models used in application-layer authorization. Existing access control models are investigated for their suitability in providing support for context awareness. Context awareness can be characterized by a model's ability to support specification of context constraints and context-influenced granting and revocation of privileges based on these constraints. Where necessary, existing models are modified or enhanced to provide such support.

5. *The design of an application-layer authorization architecture facilitating context-awareness in access control processes.* Once an appropriate access control model to support the context-aware access control paradigm has been chosen, we pursue the development of an application-layer authorization architecture. The architecture will be able to support existing network applications.

1.2 Outcomes

It is our belief that the above goals are accomplished in this thesis, as well as a number of additional contributions. The outcomes of this research are summarized below:

1. *A set of models for vulnerability analysis and building of trusted location systems.* We propose a set of location acquisition models that generalize common location technologies. The security properties of the model components were investigated. These properties were used to establish a set of requirements that trusted location systems should exhibit. A taxonomy of attacks against these common components is presented, based on known attacks against common location technologies. The taxonomy and requirements are used to classify the trust of a number of existing location technologies. The models and requirements are used throughout the first part of the thesis. Whilst access control is the primary focus for location in this research, the focus was broadened to include other security services and critical infrastructure applications in the analysis of location technologies.

2. *An authentication and integrity augmentation to differential GPS, to provide support for trusted correction and health data sourced from third-party observers.* As a result of the vulnerability analysis, an opportunity to provide a solution to differential GPS corrections for maritime applications arose. The vulnerabilities of differential GPS broadcasts are discussed in more detail, the first known discussion of such vulnerabilities. DGPS vulnerabilities are a significant concern, as they may have implications on safety-critical marine operations. A solution is proposed, augmenting the existing message broadcast protocol with an authentication and integrity scheme that attempts to mitigate the vulnerabilities identified.

3. *A new tamper-resistant GSM location system suitable for security applications.* We propose a scheme for providing tamper-resistant location acquisition in GSM

that can be used in security services such as auditing and access control. The proposal is an entire location scheme, including the geographical representations for trusted location and an association protocol that attempts to mitigate the possibility of a disassociation attack. A prototype application was implemented using the proposed location scheme, association protocol and a simulated Global Mobile Positioning Center.

4. *Proposal of a IP-layer context-aware access control.* As this research investigates the augmentation of context to access control at both the IP and application-layer, we first investigated IP-layer access control. A proximity-based packet filtering system, which can be integrated into firewall/router devices, was proposed. Proximity-based packet filtering can be used to restrict users from accessing a network from an unauthorized location, requiring that users are within a predefined location area. Experimentation was conducted through the development of a prototype system based on 802.11 wireless LAN, using signal pathloss for location determination.

5. *Discovery of a denial of service exploit for IEEE 802.11.* An unanticipated contribution of research was the discovery of a denial of service exploit for Wireless LANs. During the development of Wireless LAN location services, a vulnerability realizing a trivial denial of service attack was discovered. Experiments to determine the attack properties were conducted using the implemented attack tools. The attack is significant as it can be achieved using commercial off the shelf hardware and software; has low power requirements; and can be executed with minimal chance of detection and localization.

6. *Proposal of a new application-layer context-aware authorization architecture for Intranet environments.* A new authorization architecture is proposed that augments support for context-awareness to existing and time-tested technologies such as Kerberos, facilitating fine-grained access control to network resources and effective enforcement of security policies. A number of additions to traditional role-based access control are proposed in order to support context-awareness in access control policy. The proposed architecture was implemented.

1.3 Organization of Thesis

The thesis is organized as follows:

- Chapter 2 describes a framework for the analysis of location systems in security services. This analysis classifies cooperative locations systems by their modes of operation and the common primitives they are composed of. Common location systems are analyzed for inherent security flaws and limitations based on the vulnerability assessment of location system primitives and the taxonomy of known attacks.

- Chapter 3 proposes an efficient scheme for supporting trusted differential GPS corrections, such that DGPS vulnerabilities that have been identified are mitigated. The proposal augments the existing broadcast messaging protocol with a number of new messages facilitating origin authentication and integrity of broadcast corrections for marine vessels.

- Chapter 4 describes the proposal for a trusted location system based on GSM, in which a model for tamper resistant location determination using GSM signaling is presented. A protocol for association of a user to a cell phone is proposed and demonstrated in a framework for both Web and Wireless Application Protocol (WAP) applications.

- Chapter 5 considers the use of location context for access control at the IP-layer. A proposal for location proximity-based network packet filtering in IEEE 802.11 Wireless LANs is presented. This proposal details an architecture that extends the Linux netfilter system to support proximity-based packet filtering, using methods of transparent location determination through the application of a pathloss model to raw signal measurements.

- Chapter 6 presents a denial of service attack against IEEE 802.11 wireless LANs. During the development of Wireless LAN location services, a vulnerability realizing a trivial denial of service attack was discovered. In this chapter, we discuss the attack and its ramifications.

- Chapter 7 investigates context awareness in application-layer access control. Existing network authentication protocols and access control mechanisms are analyzed for their ability to support context-aware authorization. A new context-aware authorization architecture is developed using modifications to Role-based

Access Control (RBAC). One of the distinguishing characteristics of the proposed architecture is its ability to handle authorization with context-transparency, and provide support for real-time granting and revocation of permissions. Details of the prototype implementation, performance results, and context acquisition services are presented.

- Finally, Chapter 8 concludes by providing a summary of the research and its results, as well as proposing future directions for research, instigated by this thesis.

Chapter 2

A Framework for the Analysis of Location Systems for Security Services

2.1 Introduction

In this chapter we propose a set of models that generalize cooperative location systems for the purpose of identifying security properties that are required for location systems to be secure. These models provide a basis for analyzing existing location systems and developing secure location systems. This is the first known analysis of the security of general location systems to date.

This chapter is structured as follows. First § 2.2 introduces a number of critical civilian applications that have a requirement for trusted location. A set of models for assessing the trust of location systems are proposed in § 2.3. Primitives derived from these models are used to determine the required properties of trusted location systems. These properties are detailed in § 2.4. A taxonomy of attacks based on these primitives is detailed in § 2.5. The taxonomy and required properties are used to perform a vulnerability analysis of a number of common location technologies and classify them according to the trust they provide. The vulnerability analysis and classification are presented in § 2.6. § 2.7 details methods for increasing the trust of location systems, and § 2.8 discusses emerging technologies and issues that affect the development of trusted location systems. Lastly, a summary of the chapter is presented in § 2.9.

Portions of this chapter have been published in the paper: Enhancing the Trust of Location Acquisition Systems for Critical Applications and Location-based Security

Services. In Proceedings of the Forth Australian Information Warfare and Security Conference (AIWSC 2003), Adelaide, Australia, November 2003.

2.2 Location-based Security Services

There are numerous applications and services that have a requirement for trusted location. Whilst access control is the primary focus for location in this research, other security services and critical infrastructure applications were considered in the analysis of location technologies.

There are a number of military technologies that significantly improve the robustness and survivability of location systems against attack, however, it is not likely that they will be made available to the civil sector. As such, only civilian uses of location technologies are considered and military location technologies or augmentations are not discussed at length. A non-exhaustive list of location services used in security and critical infrastructure applications is detailed below:

- **Vehicle tracking:** Vehicle tracking is pertinent to critical infrastructure where there is the need to track vehicles carrying hazardous substances including chemicals, fuel and radioactive waste. In addition, location services can be used for the collection of taxes or tolls;

- **Personal tracking / emergency response:** this is particularly relevant in the US, where the Enhanced 911 (E911)[1] requirements have resulted in cell phones with embedded GPS functionality used for tracking by emergency services;

- **Electronic Commerce:** An emerging trend can be seen in the use of location for electronic commerce. An example is the TAD (Transaction Authentication Device) developed by WorldPay [105], using an embedded GPS chip to determine the location at which a transaction is initiated. This device is used for identifying the parties involved in large commercial Internet transactions. The integration of GPS receivers in cell phones as part of the E911 requirements will inevitably result in the widespread use of location for identification and security in M-commerce applications.

- **Control applications:** There are many uses of location for control applications including the following transport sector applications identified by Carroll et al. in [18]:

[1] See `http://www.fcc.gov/911/enhanced/`

   1. Railway traffic control and monitoring;
   2. Aviation systems including civil monitoring and landing system augmentations for precision and non-precision approaches; and
   3. Marine systems including harbor approach and constricted waterways control.

- **Access control / auditing:** Location can be used for the enhancement of access control and auditing. There are many applications where location can be used in the enforcement of security policy, assuming the location acquired can be trusted. This is the primary use for trusted location presented in this thesis; and

- **Time synchronization:** There are numerous critical applications that rely on location technologies such as GPS for time synchronization. Such applications include:

   1. Timing for phase synchronization of power during an intermesh or transferring load between substations in order to avoid complications such as tripping circuit breakers, power outages or damage to equipment within the power grid [92];
   2. Timing and synchronization of communication networks;
   3. Authentication and access control, e.g. RSA SecureID, Kerberos, etc. using time synchronization protocols such as Network Time Protocol (NTP); and
   4. Secure document timestamps (with cryptographic certification).

Because of the widespread use of location for critical services, it is imperative that the location systems be trusted.

2.3 Models for Assessing Trust of Location Systems

The trust of a location system can be determined by assessing the security vulnerabilities of a location system's components and communication protocols, and by determining how they affect the trust of the location data they acquire.

This section introduces a number of location models based on components and high-level protocols that are common to cooperative location systems. These models will facilitate the establishment of security requirements for a wide range of cooperative location technologies.

Cooperative location systems are defined by their use of location infrastructure, such that the device being located takes part in the location acquisition. These components are as follows:

1. *Location infrastructure:* The supporting architecture except the mobile device;

2. *Location device:* The device whose location is estimated;

3. *Signaling:* The signals that the infrastructure or location devices observe for the purpose of calculating location;

4. *Observation:* The method utilized to make a measurement based on the signaling. Three common types of measurements can be observed:

    (a) The propagation time of a signal (e.g. Observed Time Difference (OTD) and Time Of Arrival (TOA));

    (b) The arrival angle of a signal; or

    (c) The attenuation of a signal.

5. *Calculation:* The computation of the position using the observations:

    (a) Lateration (hyperbolic or circular) for signal propagation-time and signal attenuation observations, where signal attenuation observations require a pathloss model to estimate the distance based on the logarithmic decay of the signal; and

    (b) Angulation for angle-of-arrival based observations.

6. *Communication:* The transfer of observations and calculations between the infrastructure and device, and the transfer of location data results to the application; and

7. *Application:* The system that will request / receive the location data. An application component is present for situations where the location result is transmitted to a 3rd party. An example of a 3rd party application is a tracking system where location devices transmit their location to the tracking system. This component is omitted for cases where the location device is the endpoint for the location result.

In the following sections, we show that common location technologies can be generalized to a number of location acquisition models. In addition, we show that a general augmentation model can be applied to these models for augmentation systems of these location technologies.

2.3.1 Location Device Observed and Calculated Model

This model, as illustrated in Figure 2.1, generalizes location systems where the location is both observed and calculated by the location device. In this model, the location device communicates the calculated location result, or other data such as time, to the destination application. Figure 2.2 illustrates a variation of this model in which the application performs the location calculation instead of the device. This model is typically used in situations where the location device does not have sufficient computational power to perform calculations.

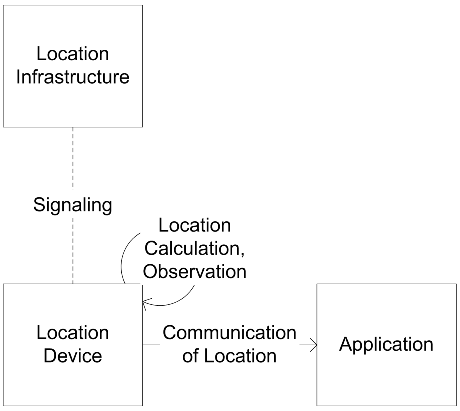


Figure 2.1: Location Device Observed and Calculated

The following technologies correspond with the *location device observed and calculated model*.

2.3.1.1 Global Positioning System

The Global Positioning System (GPS) Infrastructure consists of 24 satellites (location infrastructure) that each transmit two bands, the L1 (1575.42 MHz) frequency and the L2 (1227.60 MHz).

The signaling is broadcast on the L1 band in two channels, a narrowband channel occupied by the C/A code and a wideband channel that is intended for precision mea-

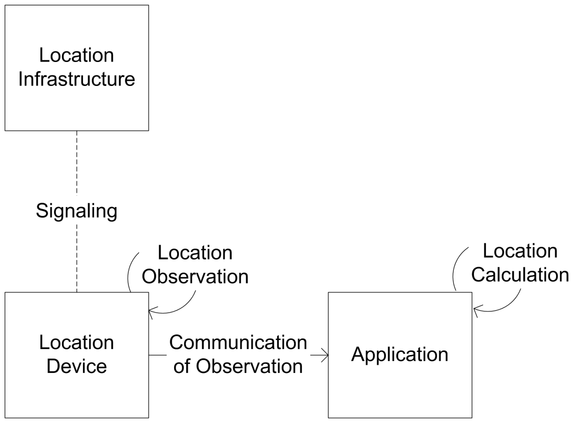


Figure 2.2: Location Device Observed and Calculated Variation

surements with the classified P(Y) code. The L2 band contains only the P(Y) code and serves mainly to act as an ionospheric effects calibrator for the L1 band [60].

The satellites transmit a Direct Sequence Spread Spectrum (DSSS) signal and use Code Division Multiple Access (CDMA) techniques to share frequencies. Data including satellite ephemeris data is modulated using CDMA.

The distances between the receiver and satellites are measured by determining pseudoranges. There are numerous factors such as receiver clock bias error, satellite clock error and physical effects such as ionospheric and tropospheric delays that affect the accuracy of a measured pseudorange. Tsui in [97] defines a measured pseudorange as:

$$\rho_i = \rho_{iT} + \Delta D_i - c(\Delta b_i - b_{ut}) + c(\Delta T_i + \Delta I_i + v_i + \Delta v_i)$$

where ΔD_i is the satellite position error effect on range, c is the speed of light, Δb_i is the satellite clock error, b_{ut} is the receiver clock bias error, ΔT_i is tropospheric delay error, ΔI_i is the ionospheric delay error, v_i is the receiver measurement noise error, and Δv_i is the relativistic time correction. ρ_{iT} is the true value of the pseudorange from user u to satellite i is given by:

$$\rho_{iT} = c(t_u - t_{si})$$

where t_{si} is the true transmission time of a satellite and t_u is the true time of reception at the receiver.

The ionospheric and tropospheric errors can be corrected by providing the receiver with information such as tropospheric models and ionospheric corrections, however, user clock bias error remains unknown and must be found in order to calculate time of arrivals.

A minimum of four equations as given below are required to solve for unknowns x_u, y_u, z_u and b_u:

$$\begin{aligned}
\rho_1 &= \sqrt{(x_1 - x_u)^2 + (y_1 - y_u)^2 + (x_1 - z_u)^2} + b_u \\
\rho_2 &= \sqrt{(x_2 - x_u)^2 + (y_2 - y_u)^2 + (x_2 - z_u)^2} + b_u \\
\rho_3 &= \sqrt{(x_3 - x_u)^2 + (y_3 - y_u)^2 + (x_3 - z_u)^2} + b_u \\
\rho_4 &= \sqrt{(x_4 - x_u)^2 + (y_4 - y_u)^2 + (x_4 - z_u)^2} + b_u
\end{aligned}$$

where $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)$ and (x_4, y_4, z_4) are the satellite positions calculated by evaluating the ephemeris data, and b_u is the user clock bias expressed in distance ($b_u = cb_{ut}$).

Solving the above equation is represented as the calculation in the location device observed and calculated model. Refer to [97] for more details on the solution of the user position from pseudoranges.

Where the GPS receiver is not the endpoint of the location data, the receiver communicates the location data to an application, typically though a serial communications interface. (Figure 2.3)

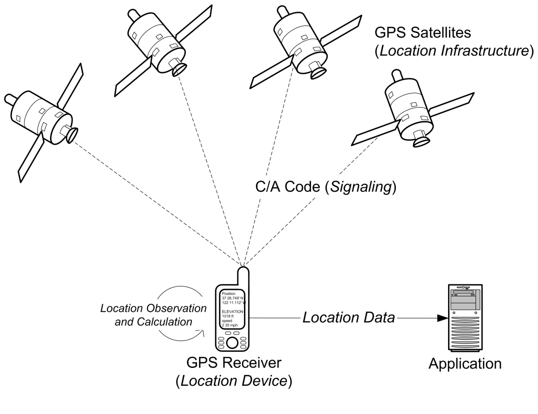


Figure 2.3: Global Positioning System (GPS)

2.3.1.2 GSM - Mobile Station Based Enhanced Observed Time Difference

Global System for Mobile Communication (GSM) is a cellular network technology providing second generation voice and data services. As location is inherent in the operation of GSM signaling, a cell phone's location can be calculated using a number of location methods based on signal timing. There are two types of Enhanced-Observed Time Difference (E-OTD) based on different measurements and calculation methods [43]:

1. *Hyperbolic Type:* This type of calculation requires the MS (Mobile Station) (location device) observe the OTD (Observed Time Difference) of signal bursts from at least three geographically disparate Base Transceiver Stations (BTS) (location infrastructure). This information is used in combination with the relative synchronization difference of the BTSs to perform hyperbolic trilateration. The E-OTD calculations given below are detailed in [43].

   The OTD for a pair of BTSs is given as: $OTD = t_2 - t_1$, where t_1 is the time a signal burst is received by the MS from BTS 1 and t_2 is the time a signal burst is received by the MS from BTS 2. If the BTSs in the network are not synchronized, the relative synchronization difference in the network between 2 BTS must be known.

   This synchronization difference is the Real Time Difference (RTD): $RTD = t_4 - t_3$, where t_4 is the time signal bursts are transmitted from BTS 1 and t_3 is the time signal bursts are transmitted from BTS 2. An RTD of 0 implies the BTSs are synchronized. The OTD is therefore related to the Geometric Time Difference (GTD) by the relationship: $OTD = RTD + GTD$. Based on the OTD and RTD value, a hyperbola can be obtained. The Hyperbola is defined by:

   $$GTD = \frac{(d_2 - d_1)}{c}$$

   where the GTD is constant and d_1, d_2 represent the distance of the propagation path between the MS and BTS 1 and 2. At least 3 BTSs are required to obtain two hyperbolas as illustrated in Figure 2.4. The intersection of the hyperbolas is the location of the MS.

2. *Circular Type:* This type of location calculation requires the MS (location device) observe the time at which signal bursts from BTSs (location infrastructure) arrive at the MS using its internal clock. Based on the observed time at which the same bursts arrive at a Location Measurement Unit (LMU), the MS clock error can be corrected and hence provide Time of Arrival (TOA) measurements.

   As the MS position is based on individual TOA measurements, the position of the MS can be found through circular trilateration (intersection of circles) as shown

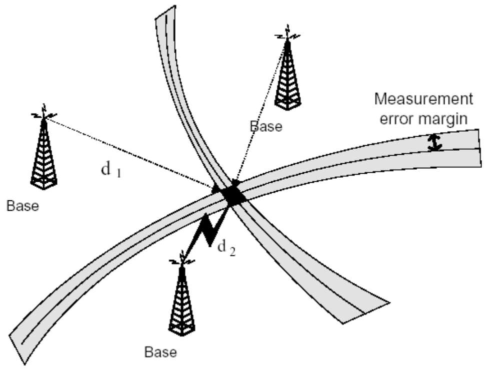


Figure 2.4: Hyperbolic Type E-OTD Positioning

in Figure 2.5. Each TOA measurement is related to the geographic distance through the relationship: $DMB - DLB = c(MOT - LOT + \epsilon)$, where c is the speed of light, MOT is the mobile observed time, LOT is the LMU observed time, and ϵ is the unknown MS clock offset.

Three equations are needed to solve for the three unknowns: x_m, y_m and clock offset ϵ are shown below:

$$\sqrt{(x_m - x_{b_1})^2 + (y_m - y_{b_1})^2} - \sqrt{(x_m - x_l)^2 + (y_m - y_l)^2} = c(MOT - LOT + \epsilon)$$
$$\sqrt{(x_m - x_{b_2})^2 + (y_m - y_{b_2})^2} - \sqrt{(x_m - x_l)^2 + (y_m - y_l)^2} = c(MOT - LOT + \epsilon)$$
$$\sqrt{(x_m - x_{b_3})^2 + (y_m - y_{b_3})^2} - \sqrt{(x_m - x_l)^2 + (y_m - y_l)^2} = c(MOT - LOT + \epsilon)$$

where (x_m,y_m) is the MS position, (x_{b_i},y_{b_i}) is the position of the BTS i, and (x_l,y_l) is the position of the LMU. More details of the E-OTD calculations given above are detailed in [43].

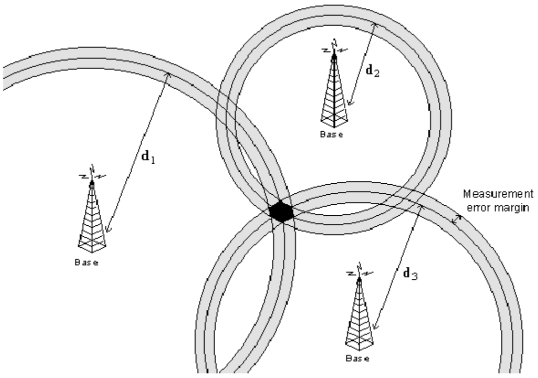


Figure 2.5: Circular Type E-OTD Positioning

E-OTD can operate in two modes: MS-Based and MS-Assisted. The MS-Based mode of operation corresponds to the *location device observed and calculated model*. The MS is able to calculate the location based on assistance data provided by the location infrastructure. This assistance data contains information including the neighbor channel RTD values, RTD drift factor values, and the serving BTS and neighbor BTS coordinates (UTM easting and northing). Refer to [42] for details of the assistance data broadcast message. §2.3.5 details the augmentation model that corresponds to the communication of augmentation data. The result of the calculation can be provided to an application via a proprietary interface on the MS and a communications link between the MS and an application.

2.3.2 Location Infrastructure Observed and Calculated Model

This model, as illustrated in Figure 2.6, generalizes location systems where the location is both observed and calculated by the location infrastructure. In this model, the location infrastructure communicates the calculated location result to the application.

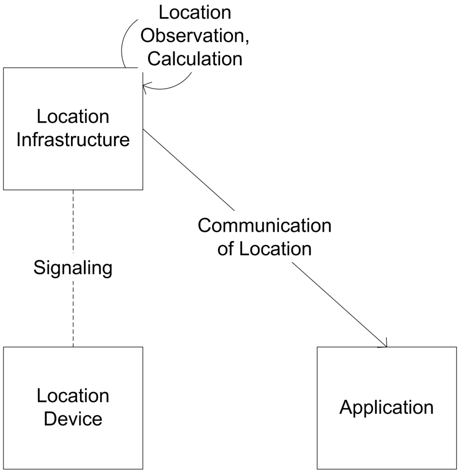


Figure 2.6: Location Infrastructure Observed and Calculated

The following technologies correspond to the *location infrastructure observed and calculated model*.

2.3.2.1 GSM - Timing Advance

Timing Advance (TA) based location acquisition is specific to GSM, where the round trip propagation delay is measured by the Base Transceiver Station (BTS) as part of the Time Division Multiple Access (TDMA) adaptive frame alignment process for ensuring a Mobile Station (MS) transmits in the correct time slot. The TA is observed by the serving BTS in the GSM location infrastructure and may be used to calculate the location in the infrastructure or MS. In the method of obtaining the TA that complies with this model, the application (an authorized LCS client) makes a request to the Gateway Mobile Location Center (GMLC) in the form of a Location Immediate Request (LIR). The cell-ID and TA are obtained by the Serving Mobile Location Center (SMLC) where the MLC-PCF (Positioning Calculation Function) calculates the position based on knowledge of the serving BTS coordinates [43]. The calculated location is returned via the GMLC to the application in the form of a Location Immediate Answer (LIA). (Figure 2.7)

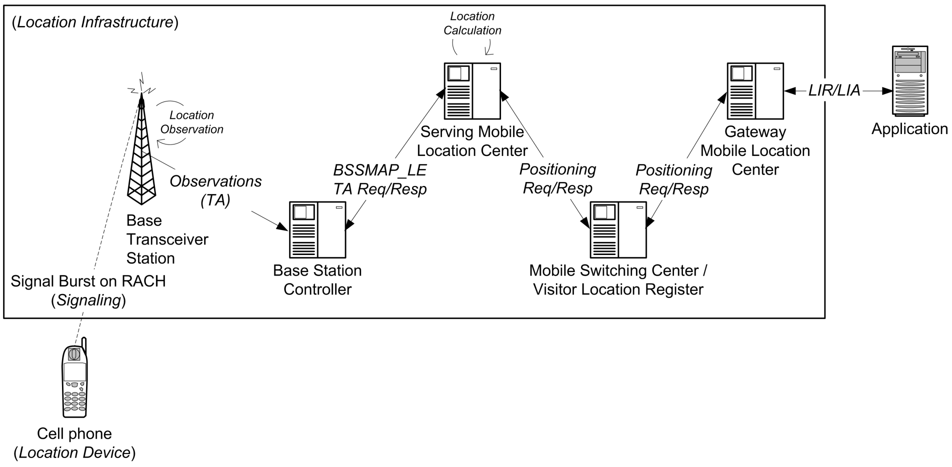

(Location Infrastructure)
Location Calculation
Location Observation
Serving Mobile Location Center
LIR/LIA
Gateway Mobile Location Center
Application
Observations (TA)
BSSMAP_LE TA Req/Resp
Positioning Req/Resp
Positioning Req/Resp
Base Transceiver Station
Base Station Controller
Mobile Switching Center / Visitor Location Register
Signal Burst on RACH (Signaling)
Cell phone (Location Device)

Figure 2.7: GSM Timing Advance

2.3.2.2 GSM - Time of Arrival

The Time of Arrival (TOA) is observed by the GSM location infrastructure and may be used to calculate the MS position in the location infrastructure. Location Measurement Units (LMU) are used to observe the TOA of a MS access burst. An LMU can be integrated into a BTS or be a stand-alone unit. The timing offset between LMUs must be known in order to calculate the location. This can be achieved through the use of absolute GPS time or a reference measurement unit placed at a known location, such that the Real Time Difference (RTD) can be determined. Figure 2.8 illustrates BTSs with integrated LMUs using GPS time, such that the RTD between pairs of BTSs is equal to 0.

The application (an authorized LCS client) makes a location request to the Gateway Mobile Location Center (GMLC)[2]. The Serving Mobile Location Center (SMLC) makes a TOA request to the BSC, which requests a handover by sending a Handover Command request to the MS. The MS then transmits Handover Access Commands (access bursts) until expiry of the T3214 timer, resulting in handover failure (Handover Failure Command). The LMUs measure the TOA of the access bursts received, and forward them to the SMLC, where the Positioning Calculation Function (MLC-PCF) calculates the position of the MS based on knowledge of the RTDs and the LMU coordinates [43] using hyperbolic trilateration. The calculated location is returned via the GMLC to the application.

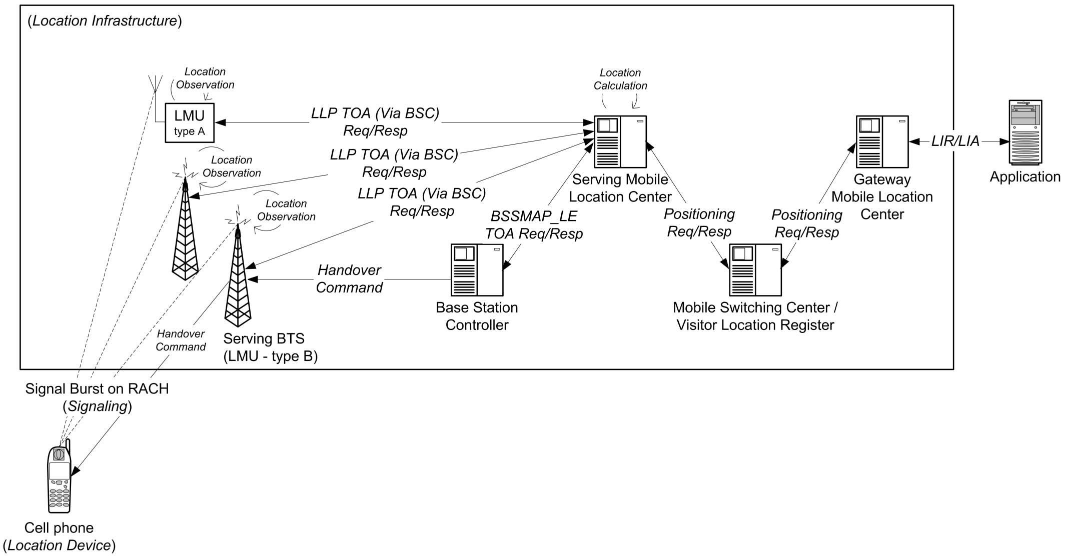

(Location Infrastructure)
Location Observation
LMU type A
LLP TOA (Via BSC) Req/Resp
Location Calculation
LLP TOA (Via BSC) Req/Resp
Location Observation
LLP TOA (Via BSC) Req/Resp
Location Observation
Serving Mobile Location Center
BSSMAP_LE TOA Req/Resp
Positioning Req/Resp
Positioning Req/Resp
Gateway Mobile Location Center
LIR/LIA
Application
Handover Command
Base Station Controller
Mobile Switching Center / Visitor Location Register
Handover Command
Serving BTS (LMU - type B)
Signal Burst on RACH (Signaling)
Cell phone (Location Device)

Figure 2.8: GSM Time Of Arrival

2.3.3 Location Device Observed, Location Infrastructure Calculated Model

This model as illustrated in Figure 2.9, generalizes location systems where the location is observed by the location device, communicated to the location infrastructure, and calculated by the location infrastructure. In this model, the location infrastructure communicates the calculated location result to the location application. The following technologies correspond to this model.

[2] Refer to Appendix F for an overview of GSM LCS services.

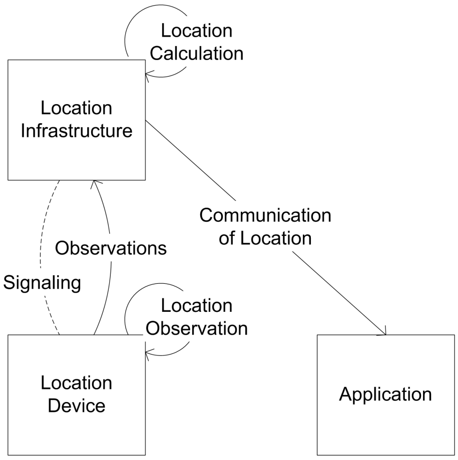


Figure 2.9: Location Device Observed, Location Infrastructure Calculated

2.3.3.1 GSM - Enhanced Observed Time Difference

In this method of E-OTD location acquisition, the application (an authorized LCS client) makes a location request to the Gateway Mobile Location Center (GMLC). For hyperbolic calculation, the Serving Mobile Location Center (SMLC) obtains E-OTD measurements from the MS, where the MLC Positioning Calculation Function (MLC-PCF) calculates the position of the MS using its knowledge of Real Time Differences (RTD) , the BTS coordinates and other supplementary data. For the circular calculation, the SMLC obtains the Mobile Observed Time (MOT) from the MS, the Location Measurement Unit (LMU) TOA measurements and other supplementary data and calculates the position using the MLC-PCF [43]. The calculated location is returned via the GMLC to the application. (Figure 2.10)

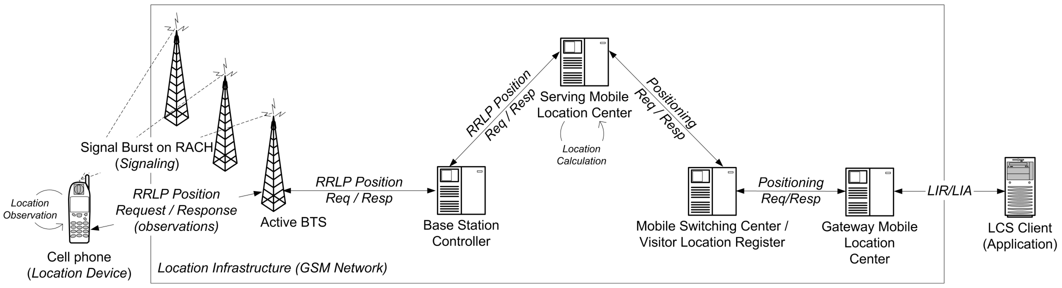


Figure 2.10: GSM Enhanced-Observed Time Difference

2.3.3.2 MS-Assisted AGPS

This method of GPS corresponds to the location device observed, location infrastructure calculated model. Refer to § 2.3.5.2 for details on MS-Assisted AGPS.

2.3.4 Location Infrastructure Observed, Location Device Calculated Model

This model, as illustrated in Figure 2.11, generalizes location systems where the location is observed by the location infrastructure, communicated to the location device, and calculated by the location device. In this model, the location device communicates the calculated location result to the location application. Similarly to the *location device observed and calculated model* in § 2.3.1, a variation of this model exists. The location observations are communicated to the application, where the application performs the location calculation.

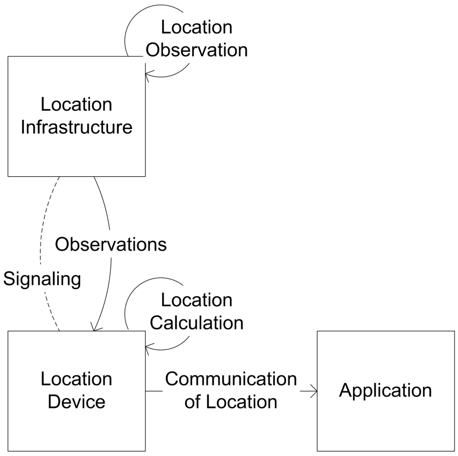


Figure 2.11: Location Infrastructure Observed, Location Device Calculated

This model is shown for completeness. While there are no standardized location methods for this model, an example of this model can be shown in GSM. The infrastructure observed measurement of the Timing Advance (TA) is communicated to the MS in layer 3 messages for adaptive frame alignment. The MS could calculate its location based on the TA measurement and the coordinates of the active BTS. The coordinates could conceivably be sent via cell broadcast messages to the MS. (Figure 2.12)

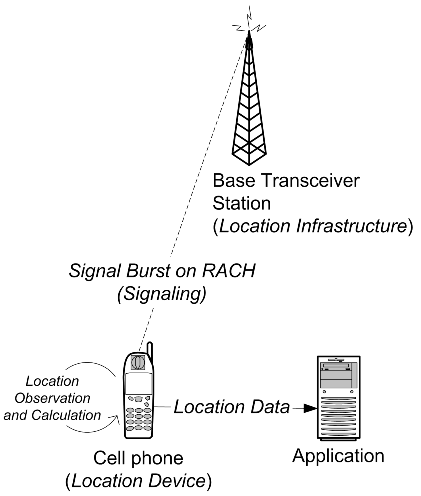


Figure 2.12: GSM Timing Advance

2.3.5 Assistance Data Augmentation Model

Basic location devices can be augmented with assistance data as shown in Figure 2.13, to provide better location accuracy. This model generalizes three modes of assistance data augmentation:

1. *Mode 1: Assistance data is communicated to the location infrastructure.* In this mode, the location infrastructure applies the corrections based on the assistance data received from a 3rd party observer and location data received from the location device. The infrastructure communicates the final location result to the application;

2. *Mode 2: Assistance data is communicated directly to the location device.* In this mode, the location device applies the corrections (assistance data) received from a 3rd party observer. The final location may be communicated to an application if the location device is not the endpoint of the location data; and

3. *Mode 3: Assistance data is communicated to the location device via the location infrastructure.* This mode is identical to Mode 2 except that assistance data is communicated via the infrastructure to the location device. Similarly to Mode 2, the final location may be communicated to an application if the location device is not the endpoint of the location data.

The following augmentation technologies correspond to this model.

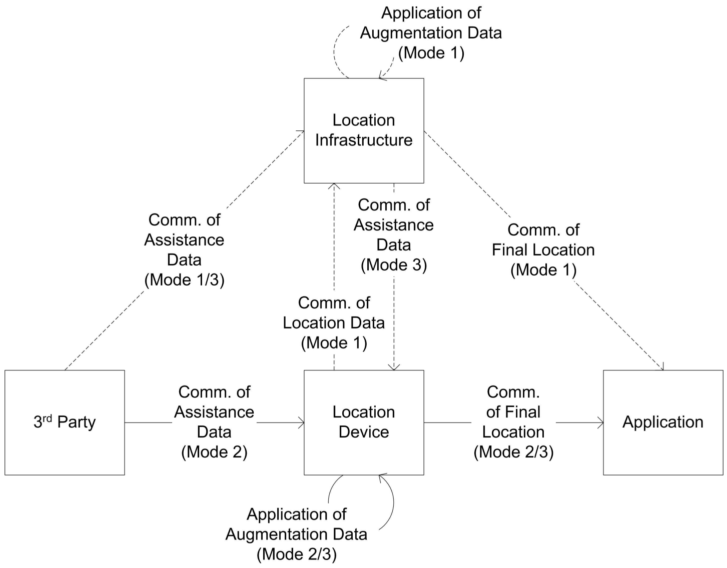


Figure 2.13: Assistance Data Model

2.3.5.1 Differential GPS

Differential GPS (D-GPS) is a type of GPS augmentation system. D-GPS versions can be shown in terms of this model, where pseudorange corrections and integrity information are derived from observations at one or more monitoring stations (3rd party) at known locations. As the 3rd party in this case is another GPS receiver, the 3rd party can be modeled by the *location device observed and calculated model*. The pseduorange corrections from the 3rd party are then broadcast to the user location device or an application for improvement of its location calculation. This broadcast corresponds to the (mode-2) communication of assistance data depicted in this model.

Ground and space-based augmentation systems can also be shown in terms of this model. The information from the monitoring stations (3rd party) is signaled to a master control station (infrastructure), as depicted by the first part of communication of (mode-3) data. The control station pre-processes the information in order to compress data to reduce data rate and to improve data consistency, then signals this processed information to the user location device. The signaling for the various functions may be either ground based (GBAS) or space based (SBAS). The latter usually allows the service to cover wide areas of service and use a standard protocol (RTCA). Examples of these satellite services are the American Wide Area Augmentation Service (WAAS), the European Geostationary Navigation Overlay System (EGNOS), the Japanese Multifunctional Transport Satellite Augmentation Systems (MSAS), the planned Indian

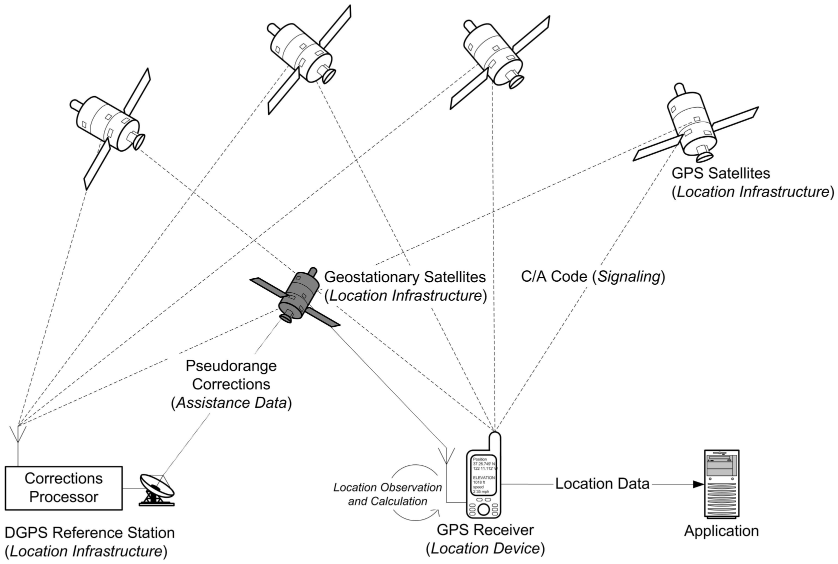


Figure 2.14: Differential GPS

GPS and Geostationary Augmented Navigation (GAGAN) and a number of commercial services such as OmniSTAR. Examples of GBASs are the Australian GRAS system as proposed by Air Services Australia [23], the future VICNET of the State of Victoria, Australia, and Trimble's virtual base station network. Figure 2.14 illustrates a differential space-based GPS augmentation system.

2.3.5.2 GSM - Assisted GPS

AGPS is a hybrid GPS technology that uses assistance data from the GSM network to facilitate demodulation of weaker signals than in conventional receivers, allowing accuracies of 15 meters when outdoors, and 50 meters when indoors [26]. Conventional GPS receivers typically require line-of-sight to the satellites, and hence do not function well, if at all in indoors.

As defined in the GSM LCS Functional Description [43], the MS (location device) makes GPS measurements aided by assistance data transmitted by the network. This assistance data consists of a list of visible satellites, satellite signal Doppler, and the code phase search window, facilitating significantly faster GPS acquisition times.

Djuknic and Richton in [26] note that in a typical cellular sector the uncertainty of a satellites signal is about $\pm 5\,\mu s$ which corresponds to ± 5 chips of the C/A code. The AGPS server can predict the phase of the C/A code and the Doppler shift due to satellite motion, which greatly reduces the search space allowing an AGPS receiver to achieve significantly faster time-to-first-fix (TTFF). These predictions are made based on the location of the cell phone obtained from the serving BTS cell ID, or using cellular location techniques such as Timing Advance.

The GPS pseudorange data processed by the MS is sent to the network infrastructure, where the Serving Mobile Location Center (SMLC) calculates the position of the MS.

There are two types of GSM Assisted GPS (A-GPS): MS-Based and MS-Assisted AGPS. In the MS-based AGPS solution, the MS consists of a fully functional GPS receiver able to perform the position calculation. In MS-Assisted AGPS, only minimal GPS receiver functionality is provided at the MS and the majority of the GPS functionality is supported by the network infrastructure in order to save power and reduce computational complexity (Figure 2.16).

In both the MS-Based and MS-Assisted AGPS, the MS location observations are made using assistance data provided by the network (Figure 2.15). This type of assistance data is contained in an RRLP position request and is modeled by mode-3 of the augmentation model, where the third party corresponds to the location device observed and calculated model. The third party additionally corresponds to the location infrastructure observed and calculated model, as a coarse cellular location acquisition is required to predict the C/A code phase and Doppler shift.

For MS-based AGPS implementations, LCS broadcast data containing differential corrections, ephemeris data, clock correction data, almanac data, ionospheric delay elements and the UTC offset is provided for increased location accuracy. This type of broadcast data is modeled by mode-3 of the augmentation model, where the third party corresponds to the location device observed and calculated model. Refer to § 2.3.5.3 for details on the LCS broadcast data transmission techniques.

In MS-Assisted GPS, the calculation function of the GPS receiver is shifted to the network processor, where the location calculation is made in the infrastructure. This type of GPS can be augmented with DGPS correction data, obtained by the network from a DGPS monitoring station. This type of augmentation is depicted by mode-1 of the augmentation model.

Pseudorange data observed by the MS is communicated to the Serving Mobile Location Center (SMLC) and GPS differential corrections are communicated from the DGPS monitoring station (3rd party) to the SMLC, where the corrections are applied to the pseudorange data, and the position of the MS is calculated.

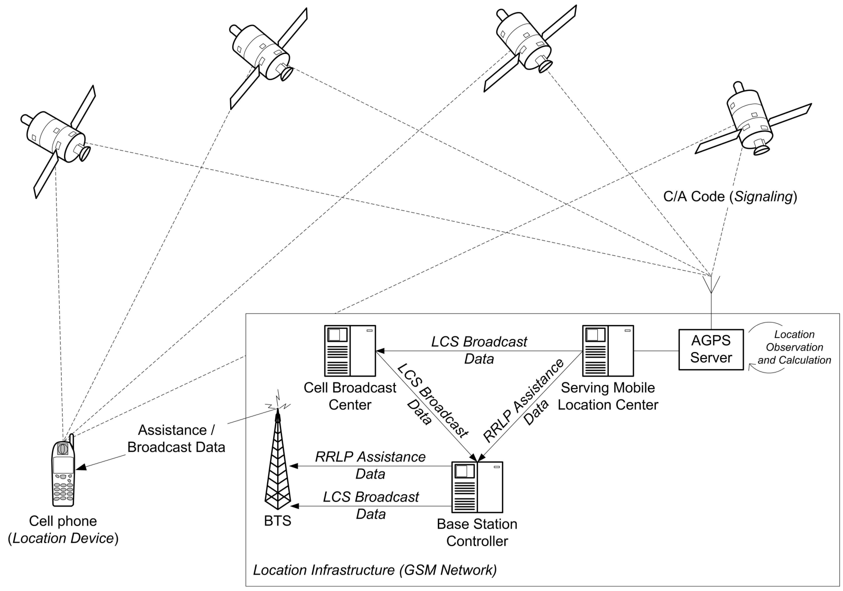


Figure 2.15: AGPS Assistance Data

2.3.5.3 Assistance Data Broadcasts in GSM

GSM augments its A-GPS and E-OTD location systems using assistance data broadcasts to the MS via the Cell Broadcast Channel (CBCH) and Short Message Service Cell Broadcast (SMSCB) [43]. The Serving Mobile Location Center (SMLC) creates a LCS broadcast message containing the assistance data as well as parameters indicating the target BTS and the time at which it is to be broadcast [42]. The assistance data is obtained from various sources within the infrastructure and communicated to the SMLC. This communication can be represented by the first (mode-3) communication from the 3rd party to the infrastructure (SMLC). The SMLC sends the message to the Cell Broadcast Center (CBC). The CBC transfers the message to the BTS, and from the BTS to the MS as detailed in The Technical Realization of Broadcast Services ETSI Standard [45]. This communication can be represented by the second (mode-3) communication from the location infrastructure (CBC/BTS) to the location device (MS).

2.4 Properties of Trusted Location Systems

In this section we describe the properties location systems should exhibit in order to be trusted. Location system primitives are used for the purpose of trust requirements

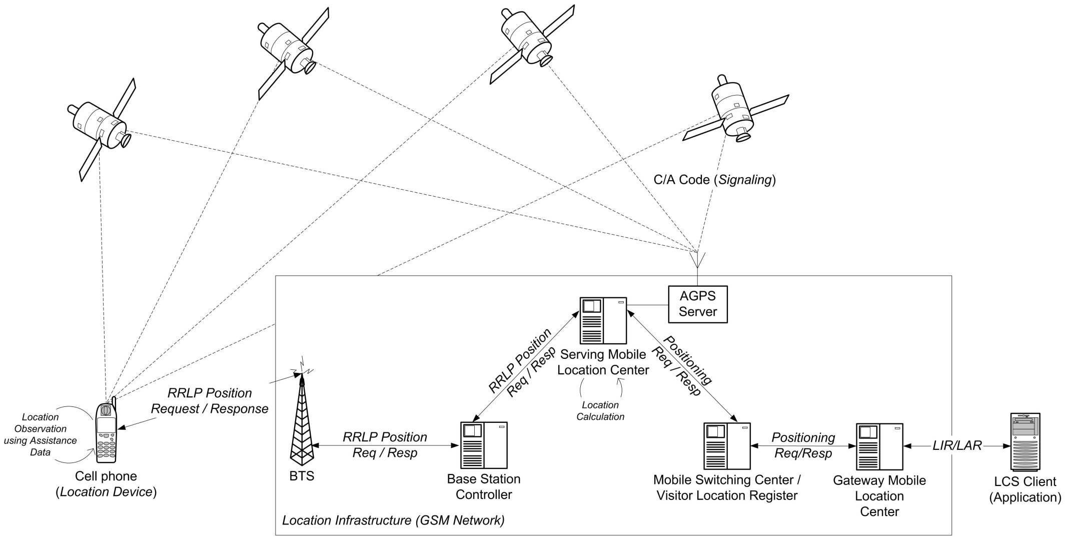

C/A Code (*Signaling*)
AGPS Server
Serving Mobile Location Center
Location Calculation
RRLP Position Req / Resp
Positioning Req / Resp
RRLP Position Request / Response
Location Observation using Assistance Data
Cell phone (*Location Device*)
BTS
RRLP Position Req / Resp
Base Station Controller
Mobile Switching Center / Visitor Location Register
Positioning Req/Resp
Gateway Mobile Location Center
LIR/LAR
LCS Client (Application)
Location Infrastructure (GSM Network)

Figure 2.16: MS-Assisted AGPS

analysis and classification. The components and operations of the generalized models can be reduced to the following four primitives:

1. Location infrastructure;
2. Location devices;
3. Signaling; and
4. Communications.

Observations and calculations are operations that are embedded within location devices or infrastructure. As such, these operations are omitted from the above primitives. This is because vulnerabilities in either a location device or infrastructure result in vulnerable observations and calculations. In addition, applications are not included in these primitives, as applications are assumed to be trusted and this thesis is not concerned with the privacy of acquired location.

At a minimum, authentication and integrity must be provided where detailed below in order for a location system to be trusted. Properties other than authentication and integrity are ancillary but further enhance the trust of location systems especially in the context of critical infrastructure applications.

2.4.1 Location Infrastructure

The location infrastructure must exhibit the following properties in order for location measurements observed from the infrastructure to be trusted.

- *System Integrity.* Location system infrastructure has system integrity if faults in the infrastructure can be detected and the consistency of location data can be verified. Faults may include both intentional and environmental errors in signaling, observations, and computation.

2.4.2 Location Devices

If the endpoint of the location result is not the location device, then the location device must exhibit the following properties in order for a 3rd party application to trust the location data.

- *Device Integrity.* The location device has integrity if faults or tampering of the location device hardware or firmware can be detected. The integrity state of the device must be communicated to the receiving 3rd party in order for the 3rd party to trust the location data the device provides, and the observations or location calculations where applicable.

2.4.3 Signaling

The signaling must exhibit the following properties in order for location measurements calculated using observations derived from the signaling to be trusted.

- *Origin Authentication.* This type of authentication is concerned with authenticating the origin of data, implicitly providing data integrity. In location systems, origin authentication of signaling is required, such that an observer is able to validate the source of the signaling to ensure that it is not a false recreation by an adversary.

- *Signal Integrity.* Signal integrity is present when signal manipulation such as the delay and rebroadcast of signals, the injection of misleading information or signal interference, both intentional and non-intentional, can be detected.

2.4.4 Communications

The following properties of communications between location system components and an application endpoint where applicable must be exhibited, otherwise the location system cannot be trusted despite the security of the components.

- *Entity Authentication.* This type of authentication is concerned with the identification of the parties involved in communications. This is important in location systems for authenticating the identity of components in which data is communicated.

- *Communications Integrity.* This refers to the integrity of the data communicated between components of the location system. In order to assure data is not corrupted or intentionally altered, data manipulation must be detectable. Data manipulation includes insertion, deletion and substitution of data.

2.4.5 Ancillary Requirements

There are a number of requirements that are ancillary to the properties required for the trust of a location system, but provide properties that may be desirable in a location system.

- *Availability.* Availability of location systems is particularly pertinent to critical infrastructure, where highly available location services are required. Availability is typically accomplished through redundancy.

- *Survivability.* Linger et. al. in [69] defines survivability as the capability of a system to fulfill its mission, in a timely manner, in the presence of attacks, failures, or accidents. For a location system to be survivable it must not only exhibit resistance to attacks, but provide strategies for detecting attacks and evaluating damage and maintaining essential services. Recovery is also an important property of survivability, involving the restoration and or maintenance of essential services. Survivability is particularly pertinent to location systems used in critical infrastructure.

- *Privacy.* A location system supports privacy if data communicated between location system components and a given destination are confidential from all but those who are authorized. Privacy may also be implemented in the signaling of certain types of location systems in order to deny access to signaling for certain users.

- *Non-repudiation.* Non repudiation prevents an entity from denying previous commitments or actions. A location system supports non-repudiation if the location of location device cannot be fraudulently denied given the location data sent to an application.

2.5 A Taxonomy of Attacks Against Location Systems

In this section we present a taxonomy of identified attacks based on primitives derived from the generalized models, defined in § 2.4. These primitives form the basis of our taxonomy, which categorizes the attacks that can occur in each location system primitive (Figure 2.17). These four categories are:

1. Attacks on location infrastructure;
2. Attacks on a location device;
3. Attack on signaling; and
4. Attacks on communications.

Each identified attack is discussed below with the potential consequences it may have on the trust of a location system.

1. *Physical Disruption.* Physical attacks that can be performed on a location device or location infrastructure include:

    - *Removal of power.* This will potentially cause denial of service for a location device, and prevent location data from being communicated to an application. Power outages may affect infrastructure, however it is most likely that infrastructure has redundant power backup.
    - *Blockage of antenna.* While antenna blockage is unlikely for infrastructure, blockage of the location device's antenna can be an effective method to prevent location acquisition and communications of location to an application.
    - *Theft and Physical damage*

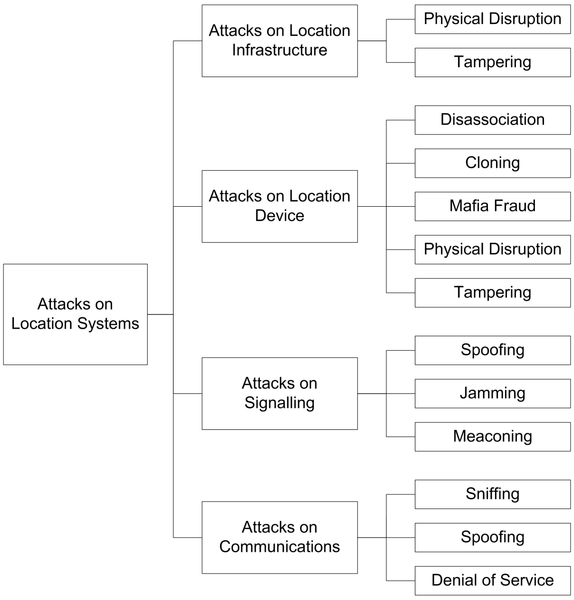


Figure 2.17: Classification of Attacks Against Location Systems

2. *Tampering.* This includes attacks on observation and calculation functions performed within the infrastructure or device. This type of attack could be performed on hardware or firmware/software causing the systems to behave improperly;

3. *Disassociation.* This is where the location device is physically removed and placed in an alternate location in order to cheat the system;

4. *Cloning.* This is where a location device is duplicated undetected, such that an adversary's location is seen to be in the location of the device that was cloned;

5. *Mafia Fraud.* This is where the device acts as a "Mafia agent" and relays all information between the participants in a communication exchange, causing misidentification for example. This attack requires the location device be compromised;

6. *Spoofing.* This involves interception, alteration and/or retransmission of a signal or data in such a way as to mislead the recipient;

7. *Jamming.* This is the deliberate radiation or reradiation of electromagnetic energy for the purpose of disrupting electronic devices and causing denial of service. Typically the transmitting power of a jamming device must exceed the power of the signal;

8. *Meaconing.* This involves receiving radio signals and rebroadcasting them on the same frequency to confuse navigation;

9. *Sniffing.* This is the unauthorized monitoring of information over a communications link; and

10. *Denial of Service against Communications.* This is where a malicious attack results in partial or total deprivation of a service. Denial of service is affective in preventing communication of location data to an application, and need not be performed in the vicinity of the location device. Depending on the communications technology used, there are numerous types of attacks possible ranging from attacks on the physical communications medium to communication protocol attacks.

2.6 Classification of Existing Location Systems

This section will present a vulnerability analysis of a number of common location technologies and two proposed location systems for security services against attacks detailed in the taxonomy (§ 2.5). The location systems are reduced to the primitives defined in § 2.4, and a vulnerability assessment conducted on each primitive in the location system.

We make use of evaluation levels to assess the ability of location system components to withstand direct attack. All the primitives derived from the generalized model are assumed to be security critical components, (i.e. those mechanisms whose failure would create a security weakness) and therefore must be assessed for each technology. The strength of each component of a location system shall be rated high risk, medium risk or improbable as follows:

1. *High Risk.* Attacks require few resources and could be performed by knowledgeable attackers;

2. *Medium Risk.* Attacks require moderate resources and could be performed by highly motivated attackers; and

3. *Improbable.* Attacks require significant resources and could be performed by attackers possessing a high level of expertise, where successful attacks are judged to be beyond normal practicality.

2.6.1 Global Positioning System

This subsection will discuss known vulnerabilities of GPS and classify known threats that exploit these vulnerabilities.

2.6.1.1 Location Infrastructure

The vulnerabilities of the GPS infrastructure in the context of the attacks identified in the taxonomy are detailed as follows.

1. **Physical Disruption *(Low Risk)*.** Physical disruption of a Satellite system is improbable due to the difficulty of accessing satellites in space. The GPS ground infrastructure is geographically distributed, making the feasibility of a physical attack on all control stations very improbable. While a physical attack on the satellites is improbable, it is stated by Adams in [1] that the US Space Command does not have an operational anti-satellite weapon. As such, it is feasible to attack the satellites using the methods detailed by Adams in [1].

   A U.S. adversary could place a crude anti-satellite system in orbit relatively cheaply through the use of ordinary meteorological sounding rockets that carry 50 to 100 pound payloads. Adams [1] stipulates that if a rocket could carry 40 pounds of steel buckshot available in most sporting goods stores, it could kick the pellets out into an appropriate orbit with an explosive charge, disabling any satellites they encountered.

   General Thomas Moorman is cited in [19] as identifying that current launch vehicles and their associated processes do not provide the responsiveness needed to rapidly replace or augment satellites. In addition, he states that the U.S launch infrastructure is vulnerable, inflexible and expensive. Physical disruption of space and ground-based augmentation systems are assumed to be a medium risk. This is because augmentation data is sourced from ground monitoring stations in both technologies. Where ground stations are distributed, an attack on a single ground station would result in a denial of service of augmentation data in the area it observes.

   The infrastructure of location systems used in critical applications should have a sufficient level of physical security to protect from the threats of tampering and disruption. Vulnerabilities in location infrastructure can be mitigated through the diversification of location acquisition technologies. While it is improbable that physical disruption will occur in satellite-based location infrastructure, diversification of technologies will increase survivability for critical applications reliant on accurate location. For example, the use of GPS-calibrated sensors such as Inertial Measurement Units (IMU) to monitor location in addition to the standard use of GPS would result in survivability for medium outages of GPS.

2. **Tampering:** ***(Low Risk)*** Tampering with the GPS infrastructure is improbable due to the inaccessibility of satellites and the high security of monitoring station installations. Tampering with the SBAS/GBAS infrastructure is also improbable due to the inaccessibility of satellites for SBAS and the assumption that there is high level of security at the monitoring station installations for both SBAS and GBAS.

2.6.1.2 Location Device

The vulnerabilities of GPS devices in the context of the attacks identified in the taxonomy are detailed as follows.

1. **Disassociation *(High Risk)*:** Disassociation of a GPS receiver is a high risk, as the removal of a receiver from a marine vessel for example, cannot easily be detected.

2. **Cloning *(Not Applicable)*:** The risk of cloning is not relevant to conventional GPS receivers, as they do not currently have any methods of authenticating themselves to an application, and as such, cloning would not be beneficial. If the GPS receiver was uniquely identifiable, the risk of cloning would be classified as high risk unless mitigated with some form of tamper-resistant module (e.g. smartcard), where removal of the tamper-resistant module and cloning of the receiver's identity would be intractable.

3. **Mafia fraud *(Medium Risk)*:** Mafia fraud attacks are medium risk, as a receiver must be compromised and must cooperate with another device.

4. **Physical Disruption *(High Risk)*:** There is a potentially high risk for disruption of GPS devices by removing the power supply or physically damaging the unit.

5. **Tampering *(High Risk)*:** An attack based on tampering with the receiver is also high risk, as the device is not typically physically secure, and there are no standardized mechanisms to authenticate the firmware integrity to an application.

An SBAS / GBAS receiver is typically integrated into or attached to a GPS receiver. The vulnerabilities of an SBAS/ GBAS receiver are same as the GPS receiver as detailed above.

2.6.1.3 Signaling

The vulnerabilities of GPS signaling in the context of the attacks identified in the taxonomy are detailed as follows.

1. **Spoofing *(Medium Risk)*:** Spoofing a GPS receiver involves generating signals and modulating navigation data over the signals that results in a legitimate location solution. The simulated signals can provide misleading information and create significant position, velocity and time errors. Spoofing of GPS signaling is considered as medium risk as GPS signal simulators are expensive but readily available. These simulators can reproduce GPS signals and navigation data, allowing the GPS signal to be easily spoofed. The simulation equipment required for a spoofing attack while expensive, can be rented relatively cheaply. A simulator could be plugged into the antenna of a GPS receiver on a truck transporting hazardous materials for example, giving the illusion that the truck is either moving or is stationary. This type of attack typically requires physical access or close proximity to the GPS receiver to be successful. To date there are no known occurrences of such activity.

2. **Jamming *(High Risk)*:** The risk of a jamming attack is considered high, as the C/A code transmitted on the L1 frequency is very weak (typically -130dBm at the antenna) and as such, easy to jam. GPS jammers generate noise on the L1 band and corrupt the original signal, making location estimation (and time synchronization) impossible, causing denial of service. The Report of the Defense Science Task Force on Tactical Air Warfare [25] states that current GPS receivers are vulnerable to jamming in acquisition mode at very long ranges from low-power jammers and will loose moderate range for reasonable jammer threats. This risk is quantified by Adams in [1] to the effect that a 100-watt jammer can affect a standard GPS receiver as far away as 600 miles (960 km) during initial GPS acquisition.

   In addition, it is stated that even when a GPS receiver has acquired the GPS signal and is using it for tracking, tracking could be interrupted within 28 miles (44.8 km) of the jammer. A 1-watt (cellular phone-size) jammer can be built from schematics that are readily available on the Internet (See Figure 2.18[3]), and can prevent a good quality civilian receiver from acquiring the C/A code from 37.5 miles (60 km). This is a significant threat for critical applications reliant on GPS.

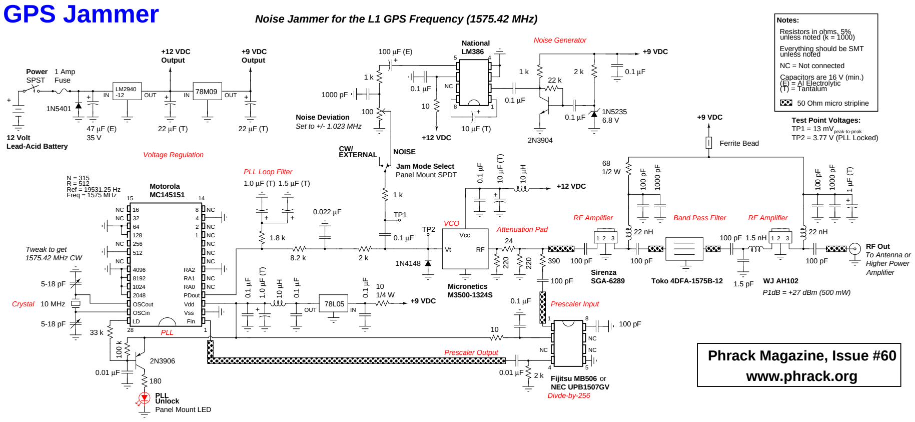


Figure 2.18: GPS Jammer Schematics

3. **Meaconing *(Low Risk)***: A meaconing attack, while potentially feasible, is considered to be improbable, as there is little evidence of successful low-cost technologies to perform the attack. The GPS signals can theoretically be captured and retransmitted in the same way an indoor GPS system does, with the exception that the signal is buffered, specific time delays on certain channels introduced, and the new signals are retransmitted confusing the GPS receiver. Meaconing is also potentially useful against space-based augmentation systems that provide a ranging signal such as the American Wide Area Augmentation System (WAAS) and European Geostationary Navigation Overlay Service (EGNOS).

2.6.1.4 Communications

There are a number of communication standards used by GPS and its augmentations. These are discussed below in the context of communications attacks identified in the taxonomy.

1. **Sniffing *(High Risk)***:

[3] See Phrack Magazine, Volume: 11, Issue: 60, Dec. 2002 for article on lost cost and portable GPS jammer. `https://www.phrack.com/`

- *NMEA 0183:* This is the defacto standard for marine navigation data communication of location data from a GPS device over a serial data link to an application device. The NMEA 0183 Interface Standard [74] defines electrical signal requirements, data transmission protocol and time, and specific sentence formats for a 4800-baud serial data bus, without providing any cryptographic message integrity. As a result, there is a high risk of sniffing because there is no cryptographic privacy protection.
- *RTCM-104:* This is the defacto standard for marine navigation data communication of GPS differential corrections (D-GPS), typically using marine radiobeacons[54]. This protocol does not make use of cryptography for integrity protection, and as such is vulnerable to the same attacks as NMEA. We address these vulnerabilities in Chapter 3.
- *RTCA:* Satellite Based Augmentation systems such as WAAS use the standard protocol RTCA for communications of D-GPS data via satellite. This is a standard protocol used in WAAS, EGNOS and MSAS. This protocol does not provide message integrity or encryption, and as such is vulnerable to the same attacks as RTCM.

2. **Spoofing *(High Risk)*:**

   - *NMEA 0183:* As there is no cryptographic integrity or privacy protection, NMEA messages can be easily spoofed. Simulator software available on the Internet facilitates one method of achieving the creation of fake NMEA messages.
   - *RTCM-104:* As there is no cryptographic integrity protection or authentication, RTCM messages are easily spoofed. The affects of a spoofing attack on a GPS receiver's reported position depends on the rejection characteristics of the GPS receiver. Inevitably some GPS receivers will accept virtually any correction data, regardless of how erroneous it is.
   - *RTCA:* Similarly with RTCM, the lack of cryptographic integrity protection or authentication allows RTCA to be easily spoofed.

3. **Denial of Service *(High Risk)*:**

   - *NMEA 0183:* Denial of service is considered as a high risk, but this depends on how the data is communicated to an application. The risk of denial of service for wireless transmission is high compared with the risk of denial of service on a cable, where a remote attack would not be possible. While denial of service at the communications layer is possible, it is easier to perform jamming on the signaling used for location acquisition to achieve the same result.

- *RTCM-104:* Denial of service is considered a low risk, as it would have little affect on navigation other than a potential reduction of accuracy if there were no other DGPS stations available. In addition, jamming a DGPS radiobeacon would take considerably more power than jamming the weak signal of GPS. If denial of service is required, it would then be more effective to jam the GPS signal rather than the DGPS signal.
- *RTCA:* Similarly to RTCM, the power required to jam RTCA data communications transmitted by satellite-based or ground-based augmentation systems would be considerably more than that required to jam the weak signal of GPS. As such, jamming would not be effective in denying a navigation solution. It may be a safety-critical issue if the RTCA transmissions are jammed during a precision airplane approach, at which time the pilots should be warned of failure.

2.6.2 Global System for Mobile Communications

This subsection will discuss known vulnerabilities of GSM and classify known threats that exploit these vulnerabilities.

2.6.2.1 Location Infrastructure

The vulnerabilities of the GSM infrastructure in the context of the attacks identified in the taxonomy are detailed as follows.

1. **Physical Disruption *(Medium Risk)*:** Physical disruption of GSM infrastructure is possible with medium to high risk, depending where the attack is performed. While Base Transceiver Stations (BTS) are vulnerable to physical disruption, the benefits of denial of service attacks against individual BTSs is limited due to the number and distributed nature of BTSs. If a Base Station Controller (BSC) is attacked, it will result in denial of service of all BTSs controlled by the BSC. An attack on a Mobile Services Switching Center (MSC), will result in denial of service to associated BSCs, and in turn the BTSs they control.

2. **Tampering *(Low Risk)*:** Tampering with the GSM infrastructure depends on the physical security of the sites. The physical security of an MSC is considerably higher than that of a BSC, which typically has better physical security than a BTS. For tampering to be affective without knowledge of a subscriber's location, it would require access to a BSC or MSC. It is assumed that tampering is improbable; however this assumes adequate physical security in all components of the infrastructure. If an adversary is able to get access to a BTS and tamper with the BTS software with the knowledge that a targeted subscriber is using the BTS, it is theoretically possible to send fictitious layer-3 Measurement Report messages to the BSC in order to prevent handover to another BTS. The adversary would then modify and transmit messages responsible for conveying location information to the BSC, such that the location of a given MS could be spoofed. The degree of spoofing is dependent on the granularity of location required. This type of attack is very unlikely to occur due to the engineering complexity and cost involved, as well as the limited ability to spoof an MS' location over large distances.

2.6.2.2 Location Device

The vulnerabilities of a GSM MS used for location, in the context of the attacks identified in the taxonomy, are detailed as follows.

1. **Disassociation *(High Risk)*:** Disassociation is a high risk, as the GSM Mobile Station (MS) can be removed or separated from the associated subject. This risk can be reduced to a medium risk using the association protocol we propose in Chapter 4.

2. **Cloning *(Medium Risk)*:** The risk of cloning is high due to the vulnerabilities of the COMP128 authentication algorithm and the ability to recover the secret key in approximately 8 hours [16].

3. **Mafia Fraud *(Low Risk)*:** There is improbable risk of mafia fraud for infrastructure observed / calculated location, as the infrastructure must be compromised. For location where observations or calculations are performed on the MS, there is medium risk as it requires the MS to be compromised and another MS to cooperate.

4. **Physical Disruption *(High Risk)*:** There is a high risk of disruption by the removal of the power supply or physical destruction.

5. **Tampering *(High Risk)*:** An attack based on tampering with the receiver is also high risk, as the device is typically not physically secure, and there are no standardized mechanisms to authenticate the firmware integrity to an application.

2.6.2.3 Signaling

The vulnerabilities of GSM signaling in the context of the attacks identified in the taxonomy are detailed as follows.

1. **Spoofing *(Low)*:**

   - *E-OTD:* Spoofing of signaling used to measure the E-OTD is considered improbable, as the OTD value is calculated from the time difference of the arrival of signal bursts from neighboring BTSs. An attack would require simulating the signals for at least one fake BTS in proximity of the MS. To spoof a significant distance would require the simulation of at least three neighboring BTSs.
   - *TOA:* Spoofing of signaling is considered improbable, as a single signal burst from an MS is used to measure TOA at 3 geographically separate BTSs. Spoofing the signal would have little or no affect on the location.
   - *TA:* Spoofing the signaling is considered improbable, as the signals would have to be simulated by a fake BTS, posing the problem of returning the TA measurement to the Serving Mobile Location Center (SMLC).

2. **Jamming *(High)*:**

   - *E-OTD:* Jamming is considered high risk as GSM jammers have become readily and cheaply available for purposes such as denying service to MSs in cinemas, hospitals, etc. The cell-phone size GSM jammer shown in Figure 2.19[4] is able to jam between 5-80 meters. Jamming is only practical where the location of the MS is known. It is possible to jam a BTS as shown in [91]. Jamming enough infrastructure to significantly affect location results is significantly more difficult and is considered low risk.
   - *TOA:* Jamming is considered a high threat for the same reasons as E-OTD.

[4] See `http://www.netline.co.il/index.html`

- *TA:* Jamming is considered a high threat for the same reasons as E-OTD.

3. **Meaconing *(Low)*:**

    - *E-OTD:* Meaconing is considered improbable, as the signal bursts would have to be buffered and retransmitted for at least one neighboring BTS, such that the signals are delayed when received at the MS antenna. This is a theoretical attack and would require equipment not commonly found, and most likely require significant engineering effort. In addition, this attack would require the attacker be in proximity of the MS. Meaconing would have a minimal affect in spoofing location of an MS.
    - *TOA:* Meaconing is considered improbable for similar reasons to those detailed in GSM E-OTD. The affect meaconing would have on the MS location would be minimal.
    - *TA:* Meaconing the signaling is considered improbable, as the TA measurement is critical for TDMA adaptive frame alignment processes, and is calculated in the BTS. Any value other than that representing the MS' propagation distance may result in a collision of signal bursts with an adjacent slot. As a result, it is very improbable that TA location obtained from the GSM infrastructure could be spoofed.

Figure 2.19: Netline Commercial GSM Jammer

2.6.2.4 Communications

The vulnerabilities of GSM communications in the context of the attacks identified in the taxonomy are detailed as follows[5].

1. **Sniffing *(High-Low Risk)*:**

   - *MS-BTS(Um):* The risk of sniffing transmissions is high due to encryption using the flawed A5 cipher. Barkan et al. in [8] propose a method of obtaining the A5 key in a few milliseconds, after which the transmissions can be sniffed.
   - *BTS-BSC(Abis interface):* There is a medium risk of sniffing where link-level encryption is not implemented over this link. In GSM there is no requirement for the use of encryption for communications between the BTS and BSC [47], and as such it may be possible to sniff data.
   - *Interfaces within infrastructure (A/Lb/Lc/Le/Lg/Lh/Lp/Ls):* These communications occur within the infrastructure over the A/Lb/Lc/Le/Lg/Lh/Lp/Ls interfaces [43]. It is assumed that these interfaces link systems that are within physically secured locations, and are physically secure themselves. It is therefore considered improbable that sniffing would occur due to the lack of accessibility.

2. **Spoofing *(Low Risk)*:**

   - *MS-BTS(Um):* Spoofing is considered medium risk, as not only does it involve attacking the A5 cipher, it also requires transmitting fictitious messages masqueraded as another user.
   - *BTS-BSC(Abis interface):* Spoofing is considered a low risk, as spoofing would require interception and modification of signaling on data links between BTS and BSCs. Where microwave transmitters are used in this interface, it may be possible to intercept a line-of-sight signal and retransmit modified versions of these messages. This would require the attacker have access to microwave transmitters and receivers, and that the microwave links are not encrypted. The GSM standard does not specify a requirement for microwave links to be encrypted [47]. Such an attack would require significant engineering and would be costly in terms of equipment required.

[5] Refer to Appendix F for an overview of the GSM architecture and the communication systems between various components of the GSM infrastructure.

- *Interfaces within infrastructure (A/Lb/Lc/Le/Lg/Lh/Lp/Ls):* It is assumed that these interfaces link systems that are within physically secured locations, and are physically secure themselves. It is therefore considered improbable that spoofing could occur due to the lack of accessibility.

3. **Denial of Service *(High-Low Risk)*:**

- *MS-BTS(Um):* There is a high risk of denial of service by jamming the MS or BTS based using commercial GSM jammers.
- *BTS-BSC(Abis interface):* The risk of a denial of service attack is deemed as medium, as it is possible to jam the microwave links when the BTS-BSC links are not cabled, to prevent transmission of location and assistance data.
- *Interfaces within infrastructure (A/Lb/Lc/Le/Lg/Lh/Lp/Ls):* It is assumed that these interfaces link systems that are within physically secured locations, and are physically secure themselves. It is therefore considered improbable that denial of service could occur due to the lack of accessibility.

2.6.3 Infrastructure-based IEEE 802.11 Wireless LAN Location

There are numerous proposals for calculating location from Wireless LAN signaling, typically using signal pathloss models. Location derived from 802.11 Wireless LAN has been classified as it is used in a proximity-based packet filtering proposal in Chapter 5, and is illustrated in Figure 2.20. The following classification is based on the mechanism used in Chapter 5.

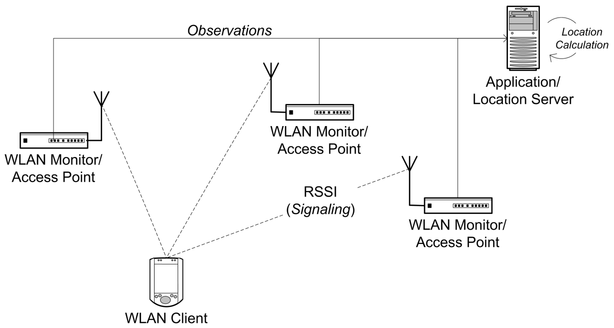

Observations
Location Calculation
Application/ Location Server
WLAN Monitor/ Access Point
WLAN Monitor/ Access Point
RSSI (Signaling)
WLAN Monitor/ Access Point
WLAN Client

Figure 2.20: Infrastructure-based IEEE 802.11 Wireless LAN Location

2.6.3.1 Location Infrastructure

The vulnerabilities of Wireless LAN infrastructure in the context of the attacks identified in the taxonomy are detailed as follows.

1. **Physical Disruption *(High Risk)*:** Wireless LAN access points continually transmit beacon frames which can be observed by freely available tools such as "Airfart"[6], and used with a pathloss model to locate access points in order to perform physical disruption. Refer to Chapter 5 for details of the pathloss model we use in our Wireless LAN location system, and a review of other commonly used models.

2. **Tampering *(High Risk)*:** Similarly with physical disruption of access points, tampering with access point power, antennas, etc. can be performed easily using the beacon frames transmitted to locate the access point.

2.6.3.2 Location Device

The vulnerabilities of Wireless LAN devices used for location, in the context of the attacks identified in the taxonomy, are detailed as follows.

1. **Disassociation *(High Risk)*:** Disassociation is possible through relocation of a WLAN card or its antenna from a client's terminal to an alternate location, such that the client is in a different location to where the location system determined the WLAN transmitter is.

2. **Cloning *(Not Applicable)*:** Cloning would only result in use of the same MAC address as another client. This can be easily achieved by spoofing the MAC address of a client. Other than this, cloning would not have any affect on location.

3. **Mafia fraud *(Not Applicable)*:** Mafia fraud is not possible as the infrastructure determines the location from signal pathloss.

4. **Physical Disruption *(Low Risk)*:** While physical disruption of a WLAN device is possible, it would result in denial of service to the communications of the client in addition to the location service. The location service is able to detect the location of a client as long as it is transmitting information. As such, we deem physical disruption to be a minimal threat.

[6] Refer to Airfart Project `http://airfart.sourceforge.net/`.

5. **Tampering *(High Risk)*:** Tampering with a WLAN card and its correct operations mode is trivial, as demonstrated in Chapter 6. WLAN cards are largely software-based and as such, it is not difficult to instigate malicious activity through modifications to the software or firmware of the WLAN card.

2.6.3.3 Signaling

The vulnerabilities of Wireless LAN signaling in the context of the attacks identified in the taxonomy are detailed as follows.

1. **Spoofing *(High Risk)*:** The signaling used for calculating the location is the same signaling used for WLAN communications. The received signal strength is measured, as it is used in the clear channel assessment algorithm to determine when a channel is free for transmission. When the signal strength indication (RSSI) is retrieved from a WLAN card, it is associated with a MAC address. It is the MAC address that is used in identifying the client that is being located. As the MAC address is easily changed, the location can be spoofed using this technique.

2. **Jamming *(High Risk)*:** Jamming of the WLAN signaling is trivial and can be done using the techniques described in Chapter 6.

3. **Meaconing *(Not Applicable)*** Meaconing has no effect on WLAN location, as the location is derived from signal pathloss not timing.

2.6.3.4 Communications

The vulnerabilities of Wireless LAN communications in the context of the attacks identified in the taxonomy are detailed as follows.

1. **Sniffing *(High/Medium Risk)*:** Communications, where observations are communicated to the location server over the WLAN, can be sniffed very easily due to vulnerabilities in the keying of the Wired Equivalent Privacy (WEP) encryption scheme. These vulnerabilities are discussed in detail in Chapter 6. Where communications is over physical wired, switched networks, the risk of sniffing is significantly reduced. Assuming an adversary can gain access to the network on which observations are being communicated, an attack on the ARP (Address Resolution Protocol) tables of an etherswitch can result in a node on the network

obtaining traffic not destined for itself. It would be most prudent to secure the transmissions using cryptography, such that the origin and integrity of observations can verified.

2. **Spoofing *(High/Low Risk)*:** Spoofing is a high risk where observation data is communicated over the WLAN. Where it is communicated over a wired network, an attack similar to that used for sniffing would be possible. The use of encryption for integrity and confidentiality would prevent spoofing of observations which would have the potential to spoof the location of a client.

3. **Denial of Service *(High/Low Risk)*:** Denial of service of WLAN communications can be accomplished through jamming where location observations are communicated using the WLAN. Where physical wiring is used, the risk of jamming the communications is minimal.

2.6.4 MacDoran GPS Authentication System Proposal

The first GPS-based location authentication system was developed by [24], the "Cyber Locator", using a patented method of location determination [70] to provide assurance of location integrity. This system uses raw GPS signals to derive a location signature, where both the client and server have the same view of satellites (Figures 2.21, 2.22). MacDoran in [70] states "The security afforded by the invention is actually enhanced by the limitations placed on the broadcast GPS signals known as Selective Availability (SA)". Before May 1, 2000 the Selective Availability policy was degrading the satellite pseudo ranging signals by dithering the navigation data and introducing an error in the clock (The Epsilon bias component), resulting in an approximate position calculation. For this reason it was improbable to predict the pseudo range and consequently the proposed location signature (LSS). After May 1, 2000 Selective Availability was no longer active. The only variation in the signal is due to atmospheric and ionospheric affects that have very few variations and do not change within hundreds kilometers. As such, we do not provide a classification of this architecture.

2.6.5 Comparison of Classified Location Systems

This subsection provides a comparison of vulnerabilities from the classified location technologies. The comparison is detailed in Table 2.1. The table represents risk as

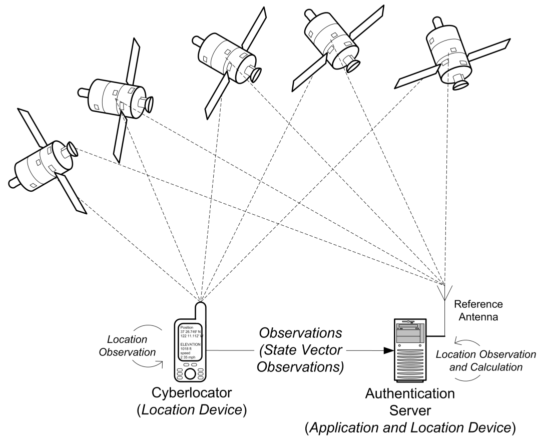


Figure 2.21: MacDoran Location Authentication Proposal

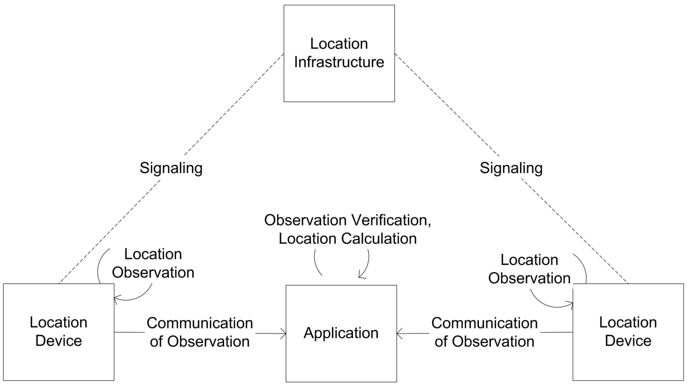


Figure 2.22: Combination of Location Device Observed and Calculated Models

(L)ow, (M)edium or (H)igh risk. The attacks for which the location systems are classified correspond with the taxonomy of attacks presented in § 2.5.

Table 2.1 illustrates the severity of vulnerabilities within a location system. The trust of an entire location system can be determined as the trust of the weakest component. For location systems where the location result does not terminate at the location device, but rather at an application, the protocols communicating the location information to the destination application must also be considered.

For GPS location systems, table 2.1 illustrates vulnerabilities of GPS location and the communications protocols, NMEA, RTCM and RTCA. In civilian GPS, the communication of location data from the GPS receiver to a remote application represents the highest risk for attack. Secure versions of these communication protocols should be developed to mitigate this risk. In Chapter 3, we propose a security scheme for

	Location System Threats														
	Intrastr		*Devices*					*Signaling*			*Comms*		
Location Technology	**PD**	**TA**	**DA**	**CL**	**MF**	**PD**	**TA**	**SP**	**JA**	**ME**	**SN**	**SP**	**DS**
GPS	L	L	H	-	M	H	H	M	H	L	▨	▨	▨
GPS - NMEA 1083	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	H	H	M
GPS - RTCM 104	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	H	H	H
GPS - RTCA	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	H	H	H
GSM	M	L	H	M	L	H	H	▨	▨	▨	▨	▨	▨
GSM - EOTD	▨	▨	▨	▨	▨	▨	▨	M	H	-	▨	▨	▨
GSM - TOA	▨	▨	▨	▨	▨	▨	▨	L	H	-	▨	▨	▨
GSM - TA	▨	▨	▨	▨	▨	▨	▨	L	H	-	▨	▨	▨
GSM - BTS-MSC (Um)	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	H	M	H
GSM - Abis	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	H	L	M
GSM - A/Lb/Lc/Le/Lg/Lh/Lp/Ls	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	L	L	L
IEEE 802.11 WLAN	H	H	H	-	-	L	H	H	H	-	H	H	H

Table 2.1: Location System Vulnerabilities

RTCM used by marine radio beacons.

Future satellite-based location systems, such as the emerging European Galileo project, will provide solutions to mitigate the vulnerabilities of civilian GPS signaling attacks such as spoofing, jamming and meaconing through the use of multiple frequencies, navigation message authentication and encrypted CDMA codes. This further emphasizes the requirement to develop secure tamper-resistant GPS receivers that mitigate device vulnerabilities such as disassociation through the use of trusted computing methods. Such GPS devices could be authenticated and ensure the location data obtained through a trusted location acquisition process could passed to a remote application via a secure protocol. Device tamper-resistance would facilitate the chaining of trust from the location infrastructure to the location device and to the remote application.

Most high-risk vulnerabilities in GSM-based location systems can be found in the location device and communications within the GSM network. Mitigation of GSM cell phone (location device) vulnerabilities will significantly increase the trust of location acquired using EOTD as well as TOA and TA. In Chapter 4 we propose a number of methods to achieve a greater degree of trust for GSM-based location acquisition, as well as a cell phone association protocol which increases the complexity of a disassociation attack.

Wireless LAN location systems suffer from high-risk vulnerabilities in infrastructure, location devices (Wireless LAN clients) and data communications. As all components of the location system are vulnerable and rated as high-risk, mitigation of any single component would have a negligible affect. Wireless LAN location systems require a new paradigm for location acquisition if wireless LAN-based location data is to be trusted.

2.7 Increasing Trust of Location Systems

Historically, military and civilian development of location technologies has been segregated. While much effort has been focused toward security of location in military applications, there has been little research in the development of secure location systems for use in critical civilian applications. We are only aware of mitigation measures directed toward the use of GPS.

There is little unclassified literature on GPS anti-spoofing techniques. Carroll et al. in [18] discusses a number of possible techniques for countering spoofing including the following:

1. Amplitude Discrimination;
2. Time-of-arrival discrimination;
3. Consistency of Navigation;
4. Polarization discrimination;
5. Angle-of-arrival discrimination; and
6. Cryptographic authentication.

Carroll et al. [18] notes that at present there are no practical mitigation methods available for spoofing attacks, and that a few potentially effective techniques would be too expensive for some civil application, in particular intelligent transportation systems.

The following subsections will review a number of currently available and emerging vulnerability mitigation strategies for GPS, both non-cryptographic and cryptographic.

2.7.1 Non-Cryptographic Signal Validation Techniques

Most non-cryptographic validation techniques were not designed to avoid intentional attacks, but rather to indicate malfunctions of satellites or the presence of physical disturbances. Non-cryptographic signal validation is based on signal and navigation checks that attempt to detect irregularities. Depending of the sophistication of the spoofer and the signal validation measures used, a spoofer may be successful in deceiving a receiver that does not have access to signal authentication.

A non-exhaustive list of signal and navigation checks discussed by Scott in [87] are detailed below.

Signal checks that can assist in detecting a spoofed signal:

- Monitoring of the Jamming-to-Noise power ratio (J/N) to check for energy levels that are above normal;
- Monitoring the Carrier-to-Noise-density ratio (C/N0) for consistency or unexpected C/N0 given the J/N; and
- Monitoring of the phase difference between antenna elements, to detect signals that come from the same direction.

The following navigation checks can be performed to assist in detecting a spoofed signal:

- Continuity checking for time and position;
- Use of a trusted internal clock to detect time drift of a spoofed signal given the likelihood spoofed signal is not synchronized with GPS time;
- Use of other navigation sensors such as Inertial Measurement Units (IMU) to check for inconsistencies and detect anomalies;
- Checking for large residual errors; and
- Use of Receiver Autonomous Integrity Monitoring (RAIM), or other integrity monitoring functions.

2.7.1.1 Integrity Monitoring Functions and Spoofing

Two integrity monitoring functions are discussed: Receiver Autonomous Integrity Monitoring (RAIM), and the WAAS/EGNOS GPS integrity monitoring services.

RAIM is a form of fault detection that allows a computerized GPS receiver to monitor the internal consistency of a set of redundant measurements in the GPS receiver to detect, and potentially remove, faulty measurements [94]. In particular, RAIM may assist in detection of spoofed differential corrections, as these would be inconsistent with its set of measurements. RAIM is good at detecting a failed satellite, but is still vulnerable to spoofing as far as a spoofer is able to simulate a healthy constellation of satellites. The primary application this technology was developed for aviation approach and landing augmentations.

GPS Augmentation systems can often be used to provide integrity and availability to the standard civilian GPS. WAAS and EGNOS both provide ranging signals that can be used by a GPS receiver as if there were a part of the GPS constellation. WAAS notifies all users within 6 seconds of a problem with any satellite in the GPS constellation, including the WAAS signal itself. This type of integrity indicates malfunctions and physical disturbances. It provides information on the health of the satellite constellation and augmentation system, but does not provide cryptographic integrity or authentication of the data source. The integrity data transmitted in RTCA messages includes information such as the carrier-to-noise-density ratio (C/N_0), which may assist in detecting a spoofer.

As RTCA messages are not cryptographically protected, integrity data can be easily spoofed. While trivial attacks against the transmission of integrity monitoring functions exist, they should not be used as a source of integrity data.

Similarly to space-based augmentation systems such as WAAS, ground DGPS radiobeacons provide an integrity service. For DGPS systems that provide correction data using RTCM messages, cryptographic integrity is not provided. The current RTCM protocol only provides a parity check for error detection and correction. In Chapter 3, we propose a security scheme for RTCM used by marine radio beacons to provide cryptographic authentication and integrity.

2.7.2 Cryptographic Signal Validation Techniques

This subsection discusses cryptographic validation techniques. These signal validation techniques specifically deal with intentional attacks against GPS. Existing military signal validation techniques and a number of proposed civil signal validation techniques are reviewed in this subsection.

2.7.2.1 Selective Availability Anti-Spoofing Module

The Selective Availability Anti-Spoofing Module (SAASM) is a tamper-resistant module used in Military GPS receivers to provide protection against attacks on signaling such as jamming and spoofing through the use of the P(Y) code, providing authenticated signaling (Figure 2.23[7]). This technology is currently restricted for use within the US Department of Defense. There is little literature about the SAASM, its cryptosystem and its protocols, however, Callaghan and Fruehauf in [17] provide a high level overview of the SAASM and Direct P(Y) signal acquisition.

Figure 2.23: L3 Xfactor™ Selective Availability Anti-Spoofing Module (SAASM)

Previous generation P(Y) code receivers required C/A code acquisition to obtain an accurate time before acquiring the P(Y) code, enabling alignment with the P(Y) code when keyed with the classified red key. The red key is used to decrypt encrypted CDMA codes. The red key requires physical transfer to the receiver by secure means. The security of the red key is critical, as compromise of the red key requires re-keying of all P(Y) code receivers.

New generation P(Y) code receivers use a SAASM, which provides direct P(Y) code acquisition without needing the C/A code, and allows for the use of unclassified black keys which can be electronically distributed as illustrated in Figure 2.24[8]. The black key is unclassified because it is the red key encrypted with the public key of the destination receiver. The black key is only decrypted within the tamper-resistant SAASM.

The SAASM module is manufactured by an approved manufacturer and sent to the Key Data Processor (KDP) Loading and Installation Facility (KLIF) which is where the cryptographic software is loaded onto the SAASM. The asymmetric key pair is generated and loaded onto the SAASM by the KLIF. The SAASM is then returned to the manufacturer where it is integrated into a GPS receiver (SAASM host equipment). Before it is released to the end user, the SAASM is re-registered with the final destination by the KLIF. This process results in simpler keying allowing the hardware and black key to be declassified.

[7] Interstate Electronics Corporation (L3) Xfactor™ SAASM with KDP II chipset. See `http://www.iechome.com`

[8] Sourced from: SAASM and Direct P(Y) Signal Acquisition [17]

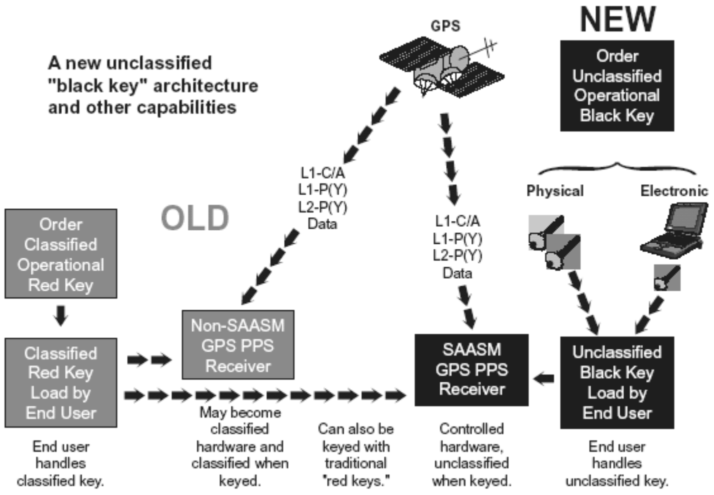


Figure 2.24: Old Red Key Versus New Black Key Keying Process

2.7.2.2 Navigation Message Authentication

Navigation message authentication is proposed by Scott in [87] for GPS-III. He proposes a digital signature scheme, whereby a new Type 7 Authentication message is used to authenticate Type 1 - 6 messages for the proposed L2C and L5 signals.

Galileo, the European satellite navigation system due for completion in 2008, will provide support for navigation message authentication on the Open Service (OS), a service projected for civilian use. The authentication system is briefly described in the Galilei Project Galileo Design Consolidation [56]. The navigation message authentication would utilize public key cryptography, and presumably include a signature to provide integrity and authenticity to the other navigation messages. It is mentioned that the public key would be distributed in a navigation message on all satellites. A high-level diagram of the Galileo navigation message authentication scheme is shown in Figure 2.25[9].

Navigation message authentication would significantly increase the difficulty of a spoofing attack. Such a scheme would require at minimum a new navigation message containing a digital signature of other navigation messages containing data including the ephemeris data. This would prevent an attacker from simulating navigation messages, significantly increasing the complexity of spoofing.

Spoofing could still theoretically be performed, however, an attacker would need to receive the legitimate signal in real time, obtain the navigation messages and modulate

[9] Sourced from: The Galilei Project: GALILEO Design Consolidation [56]

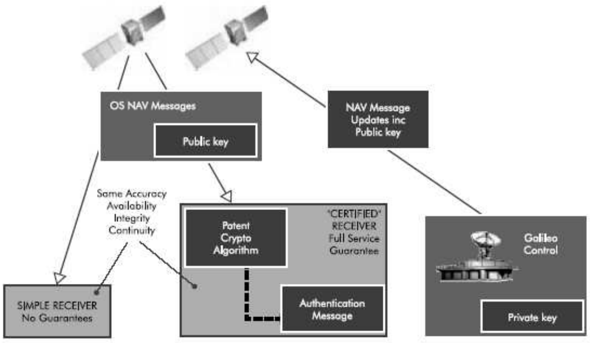

OS NAV Messages
Public key
NAV Message Updates inc Public key
Same Accuracy Availability Integrity Continuity
Patent Crypto Algorithm
"CERTIFIED" RECEIVER Full Service Guarantee
Galileo Control
SIMPLE RECEIVER No Guarantees
Authentication Message
Private key

Figure 2.25: Galileo Navigation Message Authentication

them over a simulated signal in order to spoof the location. As noted by Scott in [87], if the intended victim is already tracking legitimate signals, the time window of acceptance is very small, and would therefore require near real-time reception and remodulation over simulated signals. There is currently no commercial-off-the shelf simulator that has this functionality. Discussions with a number of engineers from two leading vendors of GPS simulators have indicated that a number of simulators are programmable and may be able to facilitate such functionality.

2.7.2.3 Signal Authentication

Next generation GNSS (GPS III and Galileo) will provide various levels of authenticated signaling to civil receivers. The European Galileo system is projected to provide the following services with signal authentication:

1. *Commercial Service (CS):* Transmitted on the E5, E6 and L1 frequencies. Access to CS signals is controlled though the use of encrypted CDMA codes and encrypted navigation data messages. All CS data is provided by service providers that interface with the Galileo Control Center; and

2. *Public Regulated Service (PRS):* Transmitted on the E5, E6 and L1 frequencies. Access to PRS signals is restricted though the use of encrypted CDMA codes and encrypted navigation data messages. This service is intended for government and military use.

Details of the cryptographic authentication schemes of these services have not been disclosed to date. It is assumed that they will operate in a similar manner to the P(Y) service and will utilize technology similar to the SAASM. Scott in [87] proposes two classes of a new type of spreading code authentication for GPS-III civilian applications:

1. *Public spreading code authentication:* Spread Spectrum Security Codes (SSSC) interleaved with the normal spreading codes are used to facilitate signal authentication. A receiver is able to authenticate the signal on receipt of a proposed Type 7 Authentication message, which is used to generate the SSSC sequence. The message is released several minutes after the sequence has already been transmitted. This allows the receiver to despread previously collected and stored samples. The signal is authenticated when the SSSC is detected at the correct power level; and

2. *Private spreading code authentication:* Similar to public spreading code authentication, except it uses a tamper-resistant Civil Anti-spoofing Security Module (CASM) that stores a protected red key, which is used in combination with the Type 7 Authentication message to generate an SSSC used specifically for private spreading code authentication. This SSSC is transmitted in the future with respect to the authentication message, significantly reducing the authentication delays experienced in public spreading code authentication.

2.8 Discussion

Trust can be increased through the mitigation of vulnerabilities in location systems. Where location devices are used in critical applications, it is imperative that the device can be trusted if it performs any observation or calculation functions. The use of tamper-resistance and trusted computing methods could facilitate a higher assurance of trust in location devices. In addition, the problems of disassociation may be remedied through the use of tamper-resistant mountings that protect the location device from disassociation.

Secure communication of both location data and augmentation data are critical to the security and reliability of critical applications. There is need for further research into secure protocols to replace the currently insecure NMEA, RTCA and RTCM protocols. Critical applications can utilize existing technologies such as GSM A-GPS for its security services such as encryption of the D-GPS corrections, providing a workable solution to the absence of a secure augmentation data service. The other GSM location mechanisms can provide a redundant location backup should there be a GPS outage.

The Galileo constellation is under development and will be operative from 2008. It will be composed of 27 active satellites + 3 spare satellites in Medium Earth Orbit. It has been projected on 6 bands: the E5a (1176.45MHz, the GPS L5) and E5b(1207.14MHz, the Glonass L3), the band E6 (1278.75MHz) and the GPS L1 (1575.42MHz) 8 MHz wider (extension E2 and E1) as detailed in [59]. The interoperability with GPS is realized by having two common frequencies in E5a/L5 and L1.

As Galileo is still under development there are no detailed specifications on signals and cryptography. There are few technical descriptions from the European Space Agency (ESA) documents, as detailed in [31] and [32]. Future work will involve investigating the Galileo proposals and exploring how these new services can be used in critical applications.

2.9 Summary

In this chapter, we have presented an informal framework for the analysis of trusted location systems. This framework contains a number of primitives that must be secure in order for a location system to be trusted. A taxonomy of attacks on these primitives was detailed, and used as a basis for a classification of a number of common location technologies, including the proposed Wireless LAN location system in Chapter 5. Lastly, this chapter discussed a number of technologies that are available to mitigate known vulnerabilities.

The following two Chapters present proposals for increasing the trust of GPS and GSM location systems.

Chapter 3

Broadcast Authentication and Integrity Augmentation for Trusted Differential GPS in Marine Navigation

3.1 Introduction

In this chapter we investigate attacks that can be mounted against marine Differential GPS radiobeacons in an attempt to spoof the location of a marine vessel, or facilitate denial of service of GPS-based navigation.

GPS does not provide instantaneous warning of system malfunctions, nor is a standard GPS receiver able to detect intentional malfunctions. Much effort has been directed at providing integrity for aviation applications such as precision and non-precision approaches. Receiver Autonomous Integrity Monitoring (RAIM) is a GPS integrity verification technology that has been developed for aviation applications. RAIM monitors the internal consistency of a set of redundant measurements in the GPS receiver to detect, and potentially remove, faulty measurements.

While RAIM technology could be used in marine applications, it appears as though maritime radiobeacons are predominantly used for integrity monitoring of GPS as well as enhancing positioning accuracy. Carroll et al. in [18] mentions that the International Marine Organization (IMO) is developing a marine RAIM standard, however this is likely to be very costly and as such, not adopted by smaller maritime operators who require access to constricted waterways.

Carroll et al. in [18] states than DGPS has already demonstrated the ability to greatly increase the safety of most maritime operations. It is a reliable navigation aid in poor weather and where constricted channels or where traffic congestion may increase the risk of an adverse event.

We are not aware of any previous work on DGPS vulnerabilities to date, except for a mention by Carroll et al. in [18], "A Vulnerability Assessment of the Transportation Infrastructure Relying on GPS", mandated by a Presidential Decision Directive to undertake a thorough evaluation of the US national transportation infrastructure. The report states that a spoofing attack on a DGPS radiobeacon is potentially feasible, but does not elaborate further. The US Department of Transportation Volpe Center as of February 2004, does not appear to have any active projects identifying or working on vulnerabilities in DGPS systems[1].

This chapter will present an overview of marine DGPS in § 3.2 and discuss a number of attacks that can theoretically be mounted against DGPS broadcast stations, in an attempt to spoof location or facilitate denial of service in § 3.3.

In addition, this chapter attempts to address these security issues through the proposal of an authentication and integrity scheme in § 3.4 that can assure the DGPS broadcasts originate from a trusted source, and that the broadcast data has not been tampered with.

This is facilitated through our proposed variation of the Time Efficient Stream Loss-tolerant Authentication (TESLA) protocol [80] (Refer to Appendix A) that provides the basis from which integrity and authenticity is provided to broadcast RTCM messages (Messages for differential GPS standardized by the Radio Technical Commission for Maritime Services (RTCM)). The security of the proposed scheme will be discussed, and finally efficiency and time/bandwidth costs of the proposed implementation of this scheme will be discussed in § 3.5.

Portions of this chapter have been published in the paper: A Broadcast Authentication and Integrity Augmentation for Trusted Differential GPS in Marine Navigation. In the Workshop on Cryptographic Algorithms and their Uses, Goldcoast, Australia, July 2004.

[1] Refer to `http://www.volpe.dot.gov/gps/projects.html` for active projects from the Volpe GPS group.

3.2 Marine Differential GPS Augmentation

The marine DGPS navigation service is an augmentation to GPS, providing pseudorange corrections and other ancillary information such as satellite health. This is designed to facilitate accurate navigation for critical harbor approach areas, as well as navigation through critical constricted waterways.

The following subsections will briefly introduce the concepts behind differential GPS, and discuss the implementation specific to marine radiobeacons.

3.2.1 Differential GPS

The accuracy of the Global Positioning System (GPS) is affected by numerous sources of errors.

1. *Satellite clocks.* Errors can be introduced from minute discrepancies that occur in the atomic clocks used on the satellites;

2. *Satellite ephemeris prediction errors.* The satellite orbit is constantly monitored, however slight orbit or “ephemeris” errors can occur in between monitoring intervals;

3. *Ionospheric delay.* Charged particles of the ionosphere may cause delays on the GPS signaling;

4. *Tropospheric delay.* The GPS signaling can be delayed by water vapor in the troposphere;

5. *Multipath errors.* Multiple copies of the signal can be received at a GPS receiver due to the effects of signal reflection, refraction, diffraction, etc. Some receivers use sophisticated signal rejection techniques to minimize this problem.

6. *Artificial errors.* Up till May 2000, the U.S. Department of Defense used selective availability to degrade location accuracy to civilian / non-authorized users. This was done by dithering the satellite clock and / or broadcasting erroneous orbital ephemeris data to create a pseudorange error.

The above errors can be reduced by applying corrections to the pseudoranges. A DGPS monitoring station within 200km of a GPS receiver will both observe the same pseudoranging signals and are prone to the same errors, given that both will have traveled though virtually the same part of the atmosphere. A DGPS monitoring station measures timing errors by comparing the observed time of arrival with the calculated time of arrival, given the known fixed position of the monitoring station. This error information is then transmitted to a DGPS receiver in the form of user differential range error (UDRE) scale factors and pseudorange corrections for visible satellites.

Typical errors that can be seen in GPS and corrected using DGPS are detailed in Table 3.1[2]. The table illustrates the affect DGPS corrections can have on errors in GPS. While receiver noise and multipath effects are not corrected by DGPS, they only account for an insignificant proportion of the errors introduced in standard GPS.

	Standard GPS	Differential GPS
Satellite Clocks	1.5	0
Orbit Errors	2.5	0
Ionosphere	5.0	0.4
Troposphere	0.5	0.2
Receiver Noise	0.3	0.3
Multipath	0.6	0.6

Table 3.1: Typical Error in Meters (per satellite) affecting GPS and DGPS Accuracy

3.2.2 Marine DGPS

Marine DGPS radiobeacons operate in the frequency range 283.5-325kHz, using Minimum Shift Keying (MSK) modulation to transmit the corrections. The corrections are transmitted using the RTCM-SC104 protocol [54]. They have ranges from 40 to 300 nautical miles. These are not commercial services, rather operated by government authorities[3].

The broadcast standard for the United States Coast Guard DGPS navigation service [99] details the requirements for DGPS monitoring stations and receivers for use in marine navigation as follows.

It is a requirement that DGPS monitoring stations ensure the integrity of the broadcast pseudorange corrections on the pseudorange level as well as on the positional level through the use of integrity monitors. Figure 3.1 illustrates the integrity monitoring process.

[2] Sourced from `http://www.trimble.com/gps/errorsources.html`.

[3] See `http://www.navcen.uscg.gov/` for DGPS sites within the USA, and `http://www.iala-aism.org/web/index.html` for a list of world-wide maritime DGPS sites

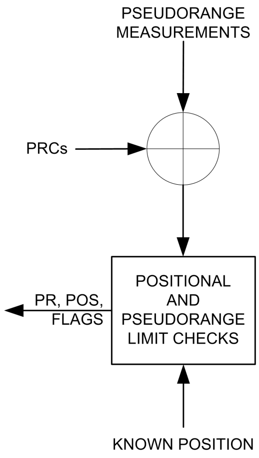


Figure 3.1: DGPS Integrity Monitoring Process

A DGPS receiver must alert the user of any out of tolerance or unhealthy conditions in the DGPS corrections. This is done through the following alarm mechanisms as specified in [99]:

- *Pseudorange Alarms.* These alarms are broadcast by setting the pseudo range corrections (PRC) field of the RTCM message header to 1000 0000 0000 0000 and the RRC field to 1000 0000. The user equipment should detect this setting and immediately stop applying any PRC derived information for that satellite until the alarm condition ends.

- *Position Alarms.* These alarms occur when either an insufficient constellation exists due to the lack of healthy pseudoranges or the failure of the pseudorange weighting or monitoring functions.

- *Unmonitored Alarm.* This alarm is raised when the corrections are not being monitored. In this case integrity in not provided.

The application of integrity messages is defined in [99] such that if an unhealthy or unmonitored condition exists as indicated by the message header of any message, it should be conveyed to the user equipment as a textual message. In addition, unhealthy or unmonitored conditions should cause a visual alarm to activate.

The RTCM messages that can be broadcast are detailed in Table 3.2. The most common messages transmitted are type 9 - GPS partial correction set. These messages are commonly transmitted with correction data for up to 3 satellites in each message, such that for 12 visible satellites, 4 type 9 messages would be broadcast.

No.	Message Name	Description
1	Differential GPS Corrections Fixed	This is the primary message type used to provide pseudorange corrections.
3	GPS Reference Station Parameters Fixed	This message contains the reference station's geographical location in Earth-Centered-Earth-Fixed (ECEF) coordinates. This message is typically transmitted every 3 to 30 minutes depending on the service provider.
5	GPS Constellation Health Fixed	This message contains information notifying users of satellite health.
6	GPS Null Frame Fixed	This message is used as a transmission filler for when there are no other messages transmitted.
7	DGPS Radiobeacon Almanac Fixed	This message provides a DGPS receiver with the information to select the optimum DGPS transmitter.
9	GPS Partial Correction Set Fixed	This message is of the same format as type 1, however unlike message type 1, it contains a smaller number of satellites. This message is transmitted at a much higher rate than the type 1 message, allowing a faster resynchronization.
16	GPS Special Message Fixed	This message is used to convey ASCII messages.

Table 3.2: RTCM Fixed Message Types

3.3 Attacks Against DGPS Transmissions

This section discusses possible attacks against marine DGPS radiobeacons with the intention of maliciously causing errors in the location calculation of a DGPS receiver, or denial of service.

While DGPS systems may facilitate integrity monitoring of the GPS constellation and the corrections that are being transmitted, it does not provide any origin authentication or integrity protection of these transmissions.

RTCM-SC104 messages include 6 bits of parity in each 30 bit word. While this provides protection against unintentional message corruption, it does not prevent an adversary from changing a message and recalculating the parity. Because there is no origin authentication of RTCM messages, it is very difficult to differentiate between a legitimate DGPS radiobeacon and a malicious one.

Similarly to GPS, DGPS is vulnerable to jamming and spoofing attacks, however there is very little publicly available information on these vulnerabilities. Carroll et al. in [18] mentions that spoofing attacks on DGPS broadcast stations are hypothetically possible, however does not elaborate on the types of attacks.

A number of possible attacks are summarized below, however they are hypothetical as we are not able to validate them. To a large extent, the ability to successfully perform an attack on a civilian receiver depends on its sophistication in rejecting erroneous correction data. We have identified two categories of attacks that can be performed against DGPS, which are discussed in the following subsections.

3.3.1 DGPS Spoofing

DGPS spoofing involves providing incorrect corrections to a DGPS user, with the intention to mislead them as to their location. The affects of a spoofing attack on a GPS receiver's position solution depends on the rejection characteristics of the GPS receiver. Inevitably some GPS receivers will accept virtually any correction data, regardless of how erroneous it is.

Spoofing GPS requires significant resources, such as a GPS simulator, to produce fake signals. While the Coarse/Acquisition (C/A) code is known and is relatively easy to generate, the signal structure and modulation technique are significantly more complex than DGPS. The signal structure of DGPS is known and trivial to recreate. RTCM message encoders are readily available on the Internet as part of projects such as DGPSIP[4].

Spoofing can possibly be achieved by providing misleading information in the following messages:

- *Type 1 - Differential GPS Corrections.* This message provides correction data for satellites in view of the DGPS reference station. It is used to provide psuedorange corrections (PRC) to a user for a GPS measurement at time t, such that: $PRC(t) = PRC(t_0) + RRC \cdot (t - t_0)$, where PRC is the pseudorange correction, RRC is the rate of change of the pseudorange correction, and t_0 is the modified *Z-COUNT*[5] from the message header. The pseudorange measured by the user, $PRM(t)$ can be corrected, such that: $PR(t) = PRM(t) + PRC(t)$. Misleading pseudorange and rate-range corrections can be sent to the user in order to spoof their location.

  Misleading information can be provided for the User Differential Range Error (UDRE) parameter, which is an estimate of the standard deviation of the dif-

[4] DGPSIP is an open-source platform for providing DGPS corrections using RTCM over IP. Refer to `http://www.wsrcc.com/wolfgang/gps/dgps-ip.html`.

[5] The modified *Z-COUNT* is a counter used for synchronization that increments every 0.6 seconds from 0 to 6000

ferential error as determined at the reference station. The UDRE can be used to weight the user's position solution. Providing misleading information in the header may also contribute in spoofing the position solution of a user.

- *Type 9 - GPS Partial Correction Set.* This message is similar to Type 1 messages, with the exception that corrections for a maximum of 3 satellites are transmitted instead of all visible satellites.

- *Type 5 - GPS Constellation Health.* This message is used to notify user equipment that an unhealthy satellite is deemed usable for DGPS. Civilian GPS receivers that detect spoofing though monitoring of carrier-to-noise-density ratio (C/N_0) for inconsistency or unexpected (C/N_0) from a DGPS monitoring site, could potentially be convinced that a spoofed GPS signal was consistent.

The broadcast standard for the US Coast Guard DGPS navigation service [99] states that an unhealthy broadcast should not be used under any circumstances. A healthy broadcast is one that is classified as healthy by its broadcast messages, is presently monitored, and the the PRC timeout (where age of PRC exceeds 30 seconds) is not exceeded for at least four satellites. It further states that the closest DGPS station should be chosen, where more than one broadcast is available, provided it is healthy. The closest DGPS station should even be used if its signal strength is low relative to other received signals.

A mitigating measure to increase survivability is to monitor the corrections from as many DGPS sites as possible, discarding broadcasts that are inconsistent with the majority of broadcasts. This should be possible due to the high level of coverage.

The denial of service techniques described in the following subsection may optimize the spoofing attacks described above, limiting the surrounding DGPS stations that may be used for providing a cross-check.

3.3.2 DGPS Denial of Service

There are two methods we have identified for achieving denial of service:

1. *Jamming the DGPS transmissions.* DGPS jamming involves causing denial of service though blocking Very High Frequency(VHF) transmissions at opportune times. This is typically done by the deliberate radiation or reradiation of electromagnetic energy, where the transmitting power of a jamming device must exceed the power of the signal. While jamming DGPS may require more power

than GPS, it should be trivial as the DGPS transmissions do not utilize modulation techniques that are resistant to jamming, such as fast frequency hopping or direct sequence spreading.

2. *Denial of service through spoofing health messages.* Based on the spoofing methods discussed in the previous subsection, it would be possible to cause denial of service by providing misleading information in the following RTCM messages:

    - *Type 5 - GPS Constellation Health.* This message could theoretically be used to convince the user equipment that a given set of satellites are unusable. This could be done by setting the LOSS OF SATTELITE WARNING parameter to 1, indicating to the user equipment that a change in the satellite's state to "unhealthy" is scheduled. In addition, the DATA HEALTH parameter could be set to 111, indicating that some or all of the satellite almanac data is bad. The ability of the user equipment to perform cross checks, determines the extent of denial of service that can be achieved.
    - *Type 7 - DGPS Radiobeacon Almanac.* This message is used to aid the user equipment in its choice of a DGPS transmitter. The parameters, RADIOBEACON RANGE, RADIOBEACON HEALTH, and STATION ID can be used to convey false information, such that the user equipment deems the neighboring beacons as unhealthy, or unsatisfactory for corrections due to the range. This message can also be used to optimize a spoofing attack, such that other DGPS transmitters that may be in range are disregarded.

Attacks that involve message manipulation, or broadcast of malicious messages are theoretically possible due to the absence of origin authentication and the absence of cryptographic integrity protection. In the following section, we describe a remedy to assist in overcoming this problem.

3.4 Proposed Message Authentication and Integrity Scheme for RTCM-SC104

There are a number of requirements for RTCM message transmissions that must be considered in developing an authentication and integrity augmentation for DGPS:

1. It is a requirement of DGPS receivers that no pseudorange corrections may be applied to the the user's navigations solution if its age exceeds 30 seconds [99];

2. Bandwidth is very constrained, as DGPS broadcast stations typically transmit at 100 or 200bps;

3. Ancillary messages should be limited to 17 words. This constraint is due to the requirement that type 16 - GPS Special Messages are not to exceed 5.1 seconds of transmission at 100bps (17 words) [99]. As such, messages used for the authentication and integrity scheme must be limited to at most this size;

4. The shorter the message, the greater the frequency of RTCM headers, significantly improving impulse noise performance. In addition, shorter correction messages provide lower latency and operate better at low data rates or in noisy environments; and

5. The computational power available to DGPS receivers may be limited in some circumstances. In larger vessels this would not be a consideration.

Traditional asymmetric signature mechanisms can be used to provide both authentication and cryptographic integrity protection. These mechanisms however, are too costly in both bandwidth, for the transmission of signatures and associated public key certificates, and in computation, for them to be used in DGPS. As such, light-weight broadcast authentication protocols were investigated.

As modification of RTCM messages would result in incompatibility with existing receivers, it was necessary to create additional messages used for supporting the authentication and integrity scheme, such that if the DGPS receiver does not support the scheme, the integrity and authentication messages are ignored. Providing an additional integrity message for each RTCM message transmitted would result in an inefficient use of the very limited bandwidth, and reduce the ability of DGPS stations to provide corrections without significant delay. As such, we proposed a modification to the TESLA protocol (Refer to Appendix A), which is detailed in the following subsections.

3.4.1 Initial Setup

A given DGPS monitor station is represented as A, and a DGPS receiver is represented as B. $y = F(x)$ is a secure one-way function, such that it is infeasible to calculate x from y.

The following initialization procedure is used to setup a hash chain, such that there is a hash value K_n for every 12 seconds of each hour for a total of 300 12 second intervals, and an additional value K_0 that can be distributed to clients.

1. A computes $K_{300} = F(s)$, where s is a random secret number chosen by A.
2. A computes K_0 by hashing K_{300} 300 times, such that $K_{299} = F(K_{300}), K_{298} = F(K_{299}), ...K_0 = F(K_1)$. The values $K_{299}...K_0$ are kept secret.
3. $A \rightarrow B : Sig_A(K_0), Cert_A$

It is assumed that the DGPS broadcast station's public key has been certified by a root DGPS authority, and the public key of the root authority has been installed on the DGPS receiver via an offline process. Hence the certificate corresponding to the private key of a DGPS broadcast station can be verified by a given DGPS receiver, and therefore can validate the signature used to authenticate K_0. The signature of K_0 for each new hash chain at the beginning of each hour is to be distributed 5 minutes before the commencement of the new chain in the proposed Type-59 message (Figure 3.7).

As $Sig_A(K_0)$ and $Cert_A$ are comparatively large and data bandwidth is limited, these message entities are fragmented over 12 message sequences to reduce the overhead on communications. A DGPS receiver must obtain the 12 messages in order to reassemble the public key certificate of the DGPS site, such that it can verify the integrity of DGPS transmissions. This process takes approximately 5 minutes at 200bps. The certificate format is detailed in Table 3.3 and is based on an X509 certificate, retaining only the necessary fields in order to minimize its size. The Subject field is the 10 bit broadcast station ID defined in RTCM. The size of the public key and signature are based on the use of 160 bit Elliptic Curve Cryptography (ECC), which provides equivalent security of 1024 bit RSA.

Issuer ID:	80 bits (10 characters)
Subject:	10 bits (BS-ID)
Valid from:	24 bits
Valid to:	24 bits
Public key: (A)	160 bits (ECC)
Signature: (K_0)	320 bits (ECC)

Table 3.3: Public Key Certificate $Cert_A$

3.4.2 The Broadcast Protocol

The protocol is synchronized in two ways:

- *Time synchronization.* The *Z-COUNT* field in the header of RTCM messages facilitates the time synchronization of the DGPS monitoring station and the DGPS receiver. The *Z-COUNT* field counts from 0 to 6000, incrementing every 0.6 seconds. We define protocol timeslots, such that each time slot has a duration of 12 seconds (a *Z-COUNT* increment of 20). The secret hash values K_{299} to K_0 are sequentially released in each timeslot.

- *Message sequence synchronization.* The *sequence (SEQ)* field in the header of RTCM messages counts from 0 to 7, facilitating message synchronization. We propose using this field to synchronize the protocol, such that each protocol sequence contains 8 messages, the first 7 messages of which are used for existing RTCM messages. The eighth message is an integrity message and is defined in § 3.5. An integrity message M_7, contains an iterated Message Authentication Code (MAC) calculated from the first 7 messages in the sequence and keyed by K_n, the hash value, K_{n+2}, 28 bits of $Sig_A(K_0)$ of the current hash chain, and 52 bits of $Cert_A$.

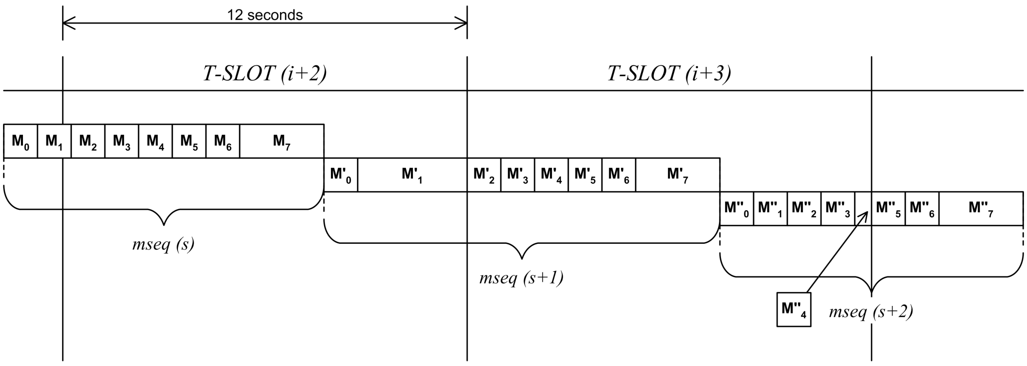


Figure 3.2: Example of Timeslots and Asynchronous Message Sequences

Figure 3.2 illustrates the protocol synchronization, and how message sequences are asynchronously transmitted. It can be observed from the diagram that the M_7 corresponds to a single timeslot, and as such that MAC of the message hashes in the sequence are keyed using a key generated from the hash corresponding to this timeslot. If M_7 is transmitted over two time slots, the timeslot at which the beginning of the message commenced is used for the keying of the MAC.

Figure 3.3 illustrates the generation of the integrity messages from the other messages in a given sequence, and the keys derived from the hash values in the hash chain.

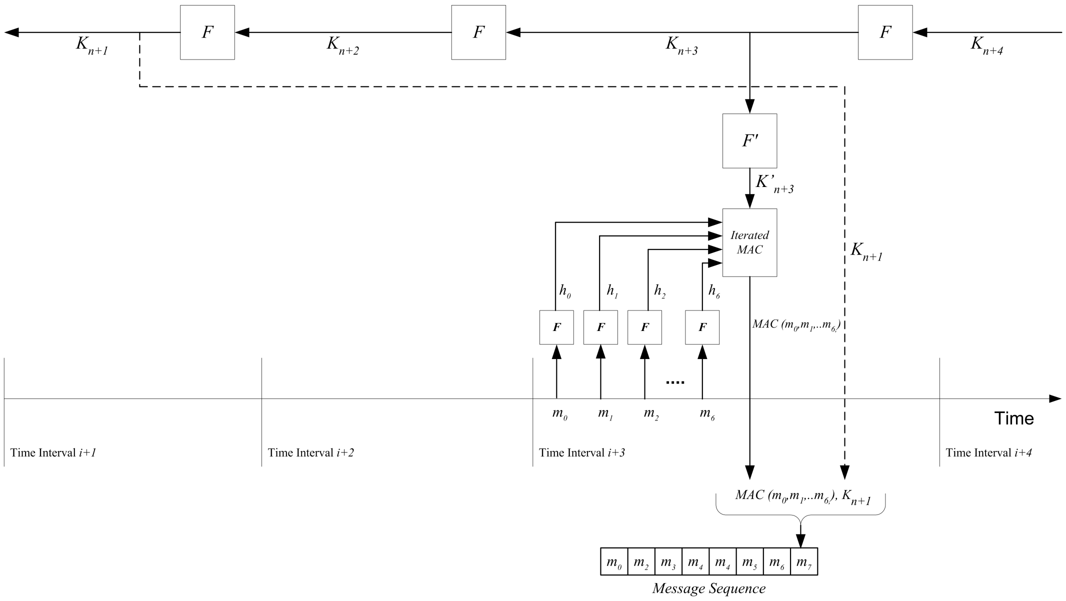


Figure 3.3: Broadcast Authentication Protocol with Integrity

In order to be backwards compatible with existing DGPS receivers and the RTCM protocol, integrity cannot be added to existing messages. Due to the limited communications bandwidth, it is not practical to transmit an integrity message for each RTCM message. Not only would this severely reduce effective communications bandwidth, it would significantly delay high rate partial GPS corrections (RTCM Message Type-9). To overcome this limitation, an iterated MAC is used, such that integrity is provided to a given message sequence rather than each individual message.

Figure 3.4[6] illustrates an iterated MAC (keyed hash function). The MAC is included in M_7 is generated based on the hashes of messages m_0 to m_6 in a given message sequence. This effectively means that an integrity failure on a message within a given sequence would result in integrity failure for all messages in the sequence.

3.4.3 Verification

This section details the verification process as performed on the user's DGPS receiver. We define the following notation such that:

[6] Diagram sourced from Handbook of Applied Cryptography [71], p332.

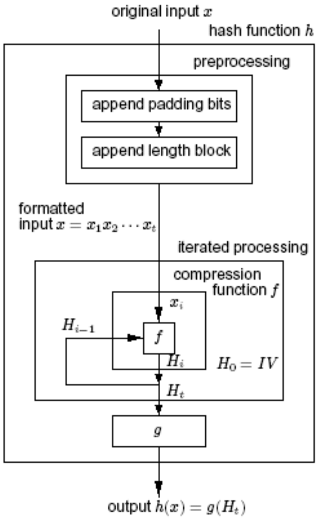


Figure 3.4: General Model for Iterative Hash Function

- *HASH* represents a hash value;
- $F(x)$ is a secure hash function, such that it is infeasible to calculate x from $F(x)$;
- $F'(x)$ is a secure key generation function;
- *T-SLOT* represents the timeslot used for hash distribution. It is calculated as a function of the RTCM header *Z-COUNT* field, such that the $\textit{T-SLOT} = INT\left(\frac{Z-COUNT}{20}\right)$; and
- *MAC* denotes a message authentication code calculated using an iterated keyed hash MAC algorithm.

The DGPS receiver maintains two sequences, a message sequence *mseq*, a buffer sequence *bseq*, and a set of hashvalues *hashvalues*:

- *mseq* denotes a sequence of message hashes. $mseq : \mathrm{seq}\, HASH$;
- *bseq* denotes a sequence of triplets containing a timeslot, the message sequence and a MAC. $bseq : \mathrm{seq}(\textit{T-SLOT}, mseq, MAC)$; and
- *hashvalues* denotes the function $hashvalues : HASH \mapsto \textit{T-SLOT}$, such that a hashvalue maps to a given timeslot.

The example below details the verification procedure for a DGPS receiver, starting to verify messages from timeslot $i + 2$, where i is the first timeslot for a the current hash chain, and the current timeslot is $i + 2$ (Refer to Figure 3.2).

In timeslot $i + 2$, messages $M_2, ..M_7, M'_0, M'_1$ are received by the DGPS receiver. First the hash of M_2 is concatenated to the sequence *mseq*. A triplet containing the timeslot of the last message in the sequence, M_7, *mseq*, and the MAC of the message sequence obtained from M_7, is concatenated to *bufseq*. On conclusion of a message sequence, and buffering of the message hashes, *mseq* is reset to an empty sequence.

If M_7 is received, the released hash value is verified in order to authenticate the source of the hash. The authentication is performed by hashing the released hash value until K_0 is verified.

Although *bufseq* represents a complete message sequence, it cannot be verified until hashvalue $K_{(i+5)}$ or later is released. For example, if there is no M_7 in a timeslot when $K_{(i+5)}$ is released, and the next M_7 contains $K_{(i+6)}$, the receiver can simply hash $K_{(i+6)}$ to produce $K_{(i+5)}$.

As with M_2, the other messages received in the timeslot are hashed and concatenated to the message buffer, *mseq*.

From the previous timeslot $(i + 1)$, $mseq = \langle F(M_0), F(M_1) \rangle$

Timeslot $(i + 2)$**:**

Received messages $M_2, M_3, M_4, M_5, M_6, M_7, M'_0, M'_1$

$$\begin{aligned} mseq' &= mseq \frown \langle F(M_2), F(M_3), F(M_4), F(M_5), F(M_6) \rangle \\ bufseq' &= bufseq \frown (\textit{T-SLOT}_{M_7}, mseq, MAC_{M_0..M_6}) \\ mseq' &= \langle\rangle \end{aligned}$$

B verifies $K_{(i+2)}$ by verifying

$$K_{(i+1)} = F(K_{(i+2)}), ...K_0 = F(K_1)$$

$$\begin{aligned} hashvalues' &= hashvalues \cup \{K_{(i+2)} \mapsto \textit{T-SLOT}_{M_7}\} \\ mseq' &= mseq \frown \langle F(M'_0), F(M'_1) \rangle \end{aligned}$$

As $K_{(i+5)}$ is not yet released, the first *mseq* in the *bufseq* cannot be verified yet. Timeslot $i + 3$ proceeds similarly to that of $i + 2$.

Timeslot $(i+3)$**:**

Received messages $M_2', M_3', M_4', M_5', M_6', M_7', M_0'', M_1'', M_2'', M_3'', M_4''$

$$\begin{aligned} mseq' &= mseq \frown \langle F(M_2'), F(M_3'), F(M_4'), F(M_5'), F(M_6') \rangle \\ bufseq' &= bufseq \frown (\textit{T-SLOT}_{M_7'}, mseq, MAC_{M_0'..M_6'}) \\ mseq' &= \langle\rangle \end{aligned}$$

B verifies $K_{(i+3)}$ by verifying

$$K_{(i+3)} = F(K_{(i+2)})$$

$$hashvalues' = hashvalues \cup \{K_{(i+3)} \mapsto \textit{T-SLOT}_{M_7'}\}$$

$\vdots$

In timeslot $(i+5)$, $K_{(i+5)}$ is released in M_7'''. After performing the buffering process as in timeslot $i+2$, the verification of message sequences for which the newly released hash value is required, are verified. The MAC of the message sequence, *mac*$_i$, is calculated and compared to the MAC stored in the buffered message sequence, $MAC_{M_0'..M_6'}$. The MAC is computed using the message hashes and a key generated from secure key generation function, F', which computes a key using the hash value returned for the time slot, $\textit{T-SLOT}_{M_7'}$ in the *hashvalues* set.

Timeslot $(i+5)$**:**

Received messages $M_4''', M_5''', M_6''', M_7'''$

(Buffering process of message hashes and corresponding timeslots proceeds as in timeslot (i + 2))

$$\begin{aligned} mseq_i &= head\ bufseq \\ &= (\textit{T-SLOT}_{M_7'}, \langle F(M_0'), F(M_1'), F(M_2'), F(M_3'), F(M_4'), F(M_5'), F(M_6') \rangle, \\ &\quad MAC_{M_0..M_6}) \end{aligned}$$

$$mac_i = MAC(F(M_0')..F(M_6'))_{F'(hashvalues(\textit{T-SLOT}_{M_7'}))}$$

if $mac_i = MAC_{M_0'..M_6'}$, messages $M_0'..M_6'$ have not been tampered with and originated from a certified DGPS broadcast source.

3.4.4 Discussion of Protocol Security

In this section we discuss the security of the modified TESLA protocol. Archer in [4] presents a mechanized correctness proof of the basic TESLA protocol [80] using TAME[7]. Archer in the paper also concludes that the degree of similarity of the proof of an analogous protocol to the proof of basic TESLA will depend on the degree of difference of this protocol from basic TESLA.

We assume that changes to the basic TESLA protocol do not modify the security properties of the proposed protocol. The proposed protocol does not vary significantly from the basic TESLA protocol except in its use of an iterated MAC for generation of a single integrity message, rather than a standard MAC being generated for each message transmitted.

A receiver can verify an integrity message in a given message sequence using the proposed protocol if:

- M_7 is successfully received;
- The succeeding sequences containing the hash used to generate the keys have been received;
- Applying the iterated MAC function to hashes of the received messages of a given sequence, using keys generated from the revealed hashes, yields a value equal to the MAC included in the integrity message, M_7, as detailed in § 3.4.2;
- M_i of a given message sequence arrives before M_{i+1} of the same sequence;
- The receiver has verified the DGPS monitoring site's public key certificate using the pre-installed public key certificate of the "DGPS certification authority"; and
- The receiver has verified that K_0 originated from a given DGPS monitoring site based on verification of the certificate containing the public key corresponding to the signature.

Since the proposed protocol does not deviate from the basic TESLA protocol except for the generation of the MAC, it can be assumed that the proposed protocol is secure. It is acknowledged that this assumption requires further work on proving the correctness of the protocol before the proposed protocol can be stated as provably secure.

[7] TAME (Timed Automata Modeling Environment) is a an interface to PVS, a verification system that supports a specification language integrated with support tools and a theorem prover.

3.5 Proposed Implementation

In this section, implementation details of the proposed authentication and integrity scheme are discussed.

The verification process could feasibly be operated in one of two modes:

1. *Sequential integrity validation.* Where the authentication and integrity scheme must validate the origin and integrity of a given message before allowing it to be used in calculating the position solution.

2. *Orthogonal integrity validation.* Where the integrity process operates orthogonally to the standard DGPS processing, such that when the origin or integrity of a message cannot be verified, the user is alarmed to the integrity problem.

Because the verification process takes at most 50 seconds at 200bps (Refer to § 3.5.1), it is not feasible to perform the verification sequentially, providing only integrity verified messages to the GPS receiver. This is because pseudorange corrections must not be older than 30 seconds [99].

As such, the verification logic operates orthogonally to the standard DGPS processing, such that integrity failures are conveyed to the user at most 50 seconds after they have been received. Instant integrity verification would induce an unsustainable overhead on communications, delaying the pseudorange corrections, such that their age would easily exceed the 30 second limit.

Similar to the application of integrity messages as defined in [99], an integrity failure detected using the proposed scheme will be conveyed to the user equipment as a textual message, as in health and monitor problems. In addition, message integrity failure should cause a visual alarm to activate.

An example "Trusted DGPS receiver" is illustrated in Figure 3.5, where the GPS receiver and DGPS receiver are separate. In the case that the DGPS receiver contains a GPS receiver, the Message Scheduler component is not needed. The trusted DGPS receiver consists of the following components:

1. *Message Scheduler.* This component is responsible for providing an RTCM message stream to both the Integrity Verification Processor, and the GPS receiver. The Message Scheduler removes the Integrity messages from the output RTCM data stream to the GPS receiver, and replaces them with 8 Type 6 GPS Null Frame messages. If an integrity failure is detected by the Integrity Verification Processor, the user is alarmed though a Type 16 - GPS Special Message. This

message is queued in the Message Scheduler to be sent during the period where the Type 6 GPS Null Frame messages are being sent.

2. *Integrity Verification Processor.* The Integrity Verification Processor performs the protocol functions, as detailed in § 3.4. A smart card or similar memory is used to store the DGPS Authority certificate. This certificate is used to verify the public key certificates broadcast from the DGPS transmitters, and hence authenticate the source of the broadcast hash, K_0, for each hour.

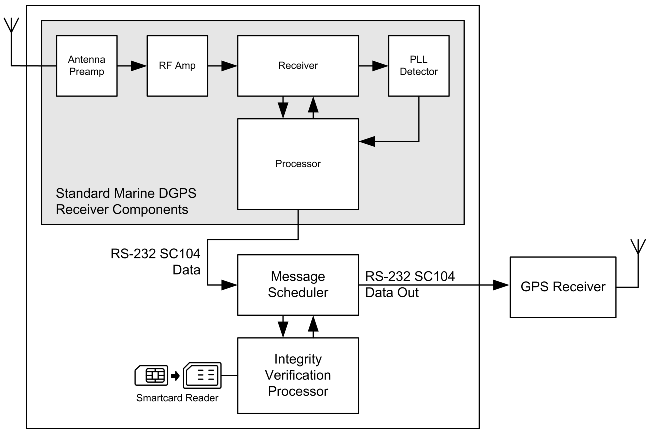


Figure 3.5: Example of Trusted DGPS Receiver

Two new RTCM messages are proposed, the Type-58 Integrity Message (Figure 3.6), and the Type-59 K_0 Signature Message (Figure 3.7). The message size is based on the use of Elliptic Curve Cryptography (ECC), which have considerably smaller key sizes than equivalent asymmetric algorithms such as RSA and DSA. The security of ECC for a 160 bit key size is comparable to 1024 bit RSA or DSA.

3.5.1 Efficiency of the Scheme

This subsection discusses the efficiency of the scheme in terms of bandwidth utilization and verification time. There is a trade-off between time to authentication / integrity alarm and bandwidth utilization. Message sequences can be reconfigured with larger numbers of messages, resulting in reduced consumption of bandwidth, however the time to alarm will correspondingly increase.

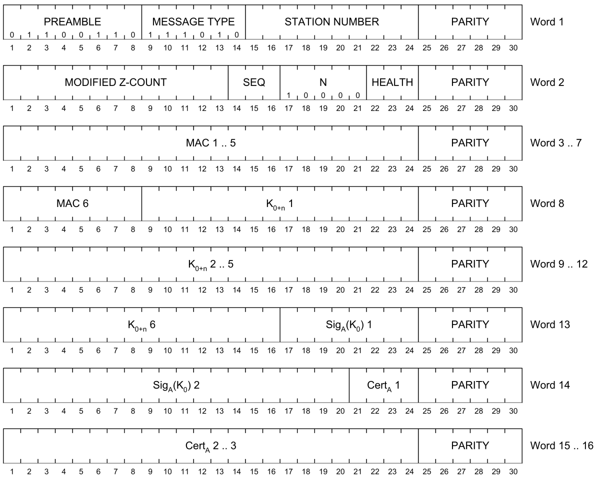


Figure 3.6: Proposed RTCM Type-58 Integrity Message

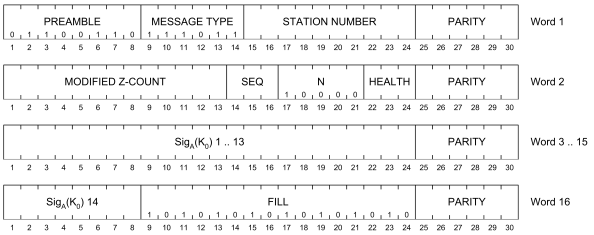


Figure 3.7: Proposed RTCM Type-59 K_0 Signature Message

3.5.1.1 Bandwidth Utilization

The following measures of bandwidth utilization have been determined using message sequences containing the smallest and largest pseudorange correction messages. The largest pseudorange correction message is the type 1 message containing corrections for the whole set of satellites in view of the monitor station (up to 12). The smallest pseudorange correction message is the type 9 message, containing the same format as the type 1 message, except in smaller groups of three satellites per message. Other message types are infrequent, and as such are not considered in this analysis. The minimum and maximum message sizes are described below.

- *Maximum PRC message size = 660 bits.* A type 1 message with 12 visible satellites occupies 22 30-bit words, 2 words being the message header, the remaining 20 words consisting of 24 bits of correction data and 6 bits of parity per word. The correction data for each satellite occupies 40 bits.

- *Minimum PRC message size = 210 bits.* A type 9 message contains a partial correction set for at most 3 satellites. This message contains 7 30-bits words, 2 words being the message header and 5 words consisting of satellite correction data as defined in the type 1 message.

Table 3.4 details the bandwidth utilization of the integrity scheme for message sequences containing minimum and maximum size PRC messages. In a message sequence of 8 messages, there are 7 standard RTCM messages and an integrity message (proposed type 58 integrity message), which introduces an overhead of 480 bits for each message sequence. For a message sequence of maximum size PRC messages, the integrity scheme requires 480 bits per 4620 bits. For a message sequence of minimum size PRC messages, the integrity scheme requires 480 bits per 1470 bits.

	Bandwidth Utilization (%)	
Protocol	**Min**	**Max**
Proposed Scheme	10.39	32.65
TESLA	72.73	228.57

Table 3.4: Comparison of Bandwidth Utilization

The proposed protocol scheme is also contrasted with the standard TESLA protocol in Table 3.4. In the TESLA protocol, each message contains a MAC and the key used to calculate the MAC of a message in a previous timeslot. As such, 440 bits must be added to the size of each message, resulting in an overhead of 480 bits per 660 bit message for a maximum size PRC message, and 480 bits per 210 bit message for minimum size PRC message.

3.5.1.2 Verification Time

This section discusses the performance of the scheme in terms of time to alarm of authentication or integrity failure.

The minimum and maximum size PRC messages, as discussed in § 3.5.1.1, are given below with their corresponding transmission times for 100 and 200 bits per second (bps).

- *Message sequence of minimum size PRC messages.*
  1950 bits (1470 + 480) = 19.5s at 100bps, or 9.75s at 200bps; and

- *Message sequence of maximum size PRC messages.*
  5100 bits (4620 + 480) = 51s at 100bps, or 25.5s at 200bps.

The hash value used to key the MAC in an integrity message, is released in another integrity message 2 timeslots after the initial integrity message is transmitted. At minimum the hash value could be released in:

$tx\_time(480 \text{ bits}) + 1 \text{ timeslot} + tx\_time(480 \text{ bits})$, where the initial integrity message is transmitted at *tx_time*(480 bits) seconds before the end of timeslot *i*, and the hash value is released in an integrity message transmitted at the beginning of timeslot $i + 2$.

- *Minimum hash value release time at 100bps.*
  4.8s + 12s + 4.8s = 21.6s

- *Minimum hash value release time at 200bps.*
  2.4s + 12s + 2.4s = 16.8s

Given the minimum hash value release times above, the time to receive the released hash value in message sequences consisting of minimum and maximum size PRC messages can be calculated. With sequences of minimum size PRC messages, the key could be obtained from an integrity message 2 message sequences after (39s) at 100bps, or 2 message sequences after at 200bps (19.5s). With message sequences consisting of maximum size PRC messages, the key would be released 1 message sequence at 100bps (51s) or 1 message sequence at 200bps (25.1s).

We define the time to alarm of integrity failure as the time it takes for the DGPS user to be alarmed to an integrity failure from the point which an integrity failure in a message occurs. This can be calculated as follows:

$hash\_value\_release\_time + r + c$, Where *r* is the remaining messages of a sequence which is at most is 7 messages. At maximum, this is 4440 bits (660*6 + 480) (44.4s at 100bps and 22.2s at 200bps) and at minimum 1740 bits (210*6 + 480) (17.4s at 100bps and 8.7s at 200bps).

Where *c* is the computation time required for integrity verification. The computation is negligible compared to the time delay imposed by the limited bandwidth, as hash functions are typically not computationally intensive. A computationally intensive public key operation would only be performed every hour for the verification of the newly distributed K_0.

Table 3.5 contrasts the time to alarm of integrity failure based on these numbers, the remaining messages of a sequence, and the computation time.

While time to alarm of integrity failure may be large in some cases with a bandwidth of 100bps, it should be noted that authentication failure will be alarmed at maximum 44.4s at 100bps and 22.2s at 200bps and at minimum 17.4s at 100bps and 8.7s at

	Max Time to Alarm of Integrity Failure (seconds)	
Data Rate	**Seq of Min Size PRC Msgs**	**Seq of Max Size PRC Msgs**
100bps	$39 + 17.4 + c \approx 56.4$	$51 + 44.4 + c \approx 95.4$
200bps	$19.5 + 8.7 + c \approx 28.2$	$25.1 + 22.2 + c \approx 47.3$

Table 3.5: Time to Alarm of Integrity Failure

200bps, based on the maximum number of messages remaining in a message sequence before an integrity message.

Depending on the the environment, the protocol can be configured for faster alarm times. For example, the current configuration may be adequate for large waterways, however for constricted waterways, it may be desirable to have significantly faster alarm times. Future research is required to determine optimal configuration parameters from a prototype implementation of the protocol.

3.6 Summary

Securing DGPS messages is critical to ensuring the integrity and source of marine navigation data. Disruption and tampering of DGPS messages can result in an increased possibility of an adverse event occurring. This is of particular concern in safety critical environments, where DGPS aims to provide enhanced safety for the transportation of hazardous cargo.

In this chapter we have presented a number of hypothetical attacks against DGPS radiobeacons which can be mitigated through the authentication and integrity scheme we proposed. The scheme not only facilitates backwards compatibility with DGPS receivers that do not support the scheme, but additionally provides assurances of integrity and authenticity whilst imposing a relatively low overhead on communications bandwidth and the latency of pseudorange corrections. Using this scheme, a DGPS receiver is able to use DGPS radiobeacons to provide a trusted location.

Proposed scheme provides solution to identified attacks through authentication and integrity scheme Scheme is flexible and can be optimized for particular the environment Provides increased trust for critical marine applications, as well as security services

Chapter 4

Proposal for GSM Tamper-resistant Location System

4.1 Introduction

In this chapter we propose a scheme for providing tamper-resistant location acquisition in GSM, and methods for integrating it into security services such as auditing and access control. We define a location metric to be a measurement of an object's current location at a point in time. This metric is used to determine whether or not the object will be granted access to a defined subject at that same point in time. Similar to biometrics, location determination technology is not 100% accurate in that there is inevitably a percentage of objects that are granted access falsely or conversely, denied access due to inaccuracies of the metric. Because the location information is being used for security purposes such as access control, the location determination technologies must exhibit some form of tamper-resistance, otherwise the metric may be deceptive in providing a false assurance of security.

Tamper-resistance of location measurements, as demonstrated in this chapter, can provide a reasonable level of assurance, such that attacking or spoofing is possible only with significant financial means and by highly motivated attackers. We are unaware of any previous work in location tamper-resistance with GSM.

This chapter is structured as follows. First the requirements of location tamper-resistance in GSM are presented in § 4.2. GSM measurements that comply with the requirements are reviewed in § 4.3 and § 4.4. A proposed method of obtaining high-assurance location measurements is described in § 4.5 and its practical use in a location system in § 4.6. A geographical data representation supporting trusted location areas is introduced in § 4.7 and applied to location-based access control and auditing in § 4.8.

A cellular phone association protocol is proposed in §4.9 to provide a reasonable level of assurance that a given user is associated with their cell phone. A security analysis of the protocol is given in § 4.10. An example Internet banking system that was implemented using the concepts presented in previous sections is presented in § 4.11.

Portions of this chapter have been published in the paper: Enhancing the Security of Internet Applications using Location: A New Model for Tamper-resistant GSM Location. In Proceedings of the Eighth IEEE International Symposium on Computers and Communications (ISCC 2003), Antalya, Turkey, July 2003.

4.2 Location Tamper-resistance in GSM

We define tampering of a location system as attacks against the signaling and observations of the signaling. Location tamper-resistance is dependent on the integrity of the signaling and the mode by which the location signaling is observed. Whilst the properties of trusted location systems detailed in § 2.4 suggests that all location systems can be trusted given fulfillment of the requirements defined, it is deemed unlikely Mobile Station (MS)-observed models of location acquisition can be trusted due to the cost and complexity of authenticating the firmware performing the observation functions. As such, we introduce a number of additional requirements and assumptions for a location acquisition method in GSM to exhibit tamper-resistance.

The use of infrastructure-based or third-party observers significantly reduces the likelihood that signal-tampering will occur. When a MS transmits an access burst on the Random Access Channel (RACH), observations as to the propagation delay can be measured. (The Location Infrastructure Observed and Calculated Model of location determination detailed in § 2.3.2)

For a GSM measurement to be tamper-resistant, it must comply with at least one of the following additional requirements:

1. *Using at least three observers.* The time an access burst is received at a given Location Measurement Unit (LMU) is observed and trilaterated with three or more LMUs to provide a location result. The LMUs must be time-synchronized, or the time-base difference between LMUs known. Increasing the number of

LMUs involved in a given observation not only increases the accuracy of the location result, but also increases the trust of the location result. This is because it becomes increasingly difficult to tamper with the signal timing observed by a large number of observers. In terms of attacks on the observer, increasing the number of geographically disparate LMUs increases the difficulty of an attacker tampering with LMUs in order to spoof the location. Erroneous results from LMUs can be detected and discarded where there are redundant measurements available (more than the minimum 3 observers). It should be noted that LMUs and the GSM core network equipment are assumed to be physically secure.

2. *Tamper-resistant signaling properties.* The properties of the signaling are such that tampering with the timing of the signal results in disconnection. In this case, the measurement must be correct for the MS to communicate successfully. An example of this is the timing advance used for TDMA adaptive frame alignment. Signaling tamper-resistance is particularly pertinent in situations where only a single observer is used. This is because the signaling can easily be tampered with in order to spoof the location. In addition, antenna extension attacks can be detected as tamper-resistant signaling usually requires round-trip-propagation to be calculated.

Both Time of Arrival (TOA) and Timing Advance (TA) based location determination techniques in GSM comply with these requirements and are discussed in the subsequent sections.

4.3 Time of Arrival as a Tamper-resistant Measurement

Time of Arrival (TOA) is the most tamper-resistant measurement available in GSM. This is due to the number of observers that receive a signal burst, such that regardless of when an access burst is transmitted or how much it is intentionally delayed, it is infeasible to delay or corrupt it in such a way as to affect reception of the access burst by numerous geographically dispersed LMUs. The more observers available for a particular measurement, the harder it is to spoof the location.

This location method, however, is vulnerable to antenna extension attacks. This is where the antenna is at location other than the MS, and signaling is relayed between the antenna and the MS. Only round-trip propagation measurements are able to detect the total propagation distance, which is increased in an antenna attack. The TOA location can be calibrated with the TA measurement, such that if the TA significantly exceeds the vector distance between the MS and serving BTS calculated from the TOA location, the TOA location can be rejected, and the TA used in its place. The TA reflects *at least* the minimum line-of-sight propagation distance of the MS to the serving BTS. The proposed verification procedure is as follows:

$v_{TOA} = \sqrt{(x_{BTS} - x_{TOA})^2 + (y_{BTS} - y_{TOA})^2}$
If $v_{TOA} \leq ((TA + \epsilon) \cdot 550) + 275$, *accept* TOA location, else *reject*.

Where:
x_{BTS}, y_{BTS} are the coordinates of the BTS;
x_{TOA}, y_{TOA} are the TOA location coordinates; and
ϵ is the TA error.

If $(v_{TOA}+m > (TA\cdot 550)+275)$, $\epsilon = 1$ else $\epsilon = 0$. ϵ is never to exceed 1. *m* is the margin of error allowable near the boundaries of the TA. The recommended value for *m* is a 5% error margin (≈ 27.5 meters). For example, an actual distance of 830m may result in a TA of 1 rather than 2 due to bit rounding. In this case $830+27.5 > (1\cdot 550)+275$), and a ϵ value of 1 will be used, resulting in an accepted location. Refer to § 4.4 for more details on the timing advance measurement.

Unfortunately, LMUs are not commonly deployed in GSM networks due to cost and the availability of other location determination mechanisms such as E-OTD, which require no additional infrastructure and result in similar if not better location accuracy. In Australia, Telstra, the network operator currently providing cellular location services, does not provide support for TOA. Only low resolution location based on the TA and Cell Global Identifier (CGI) are supported though Ericsson Mobile Positioning Center[1]. This has motivated us to investigate the Timing Advance measurement and its ability to facilitate tamper-resistance in particular detail.

4.4 Timing Advance as a Tamper-resistant Measurement

This section discusses the properties of the Timing Advance (TA) as a tamper-resistant measurement. GSM uses TDMA to share carrier frequencies with multiple users. Each

[1] For more information on location services supported by Telstra, refer to `http://www.telstra.com.au/datadevelopers/tools/services.cfm`, accessed January, 2004.

TDMA frame has 8 slots, providing each user with the carrier frequency for 0.577ms. A time-slot has 156.25 bit periods, the last 8.25 bit periods being the guard allowing a maximum of 8 bit periods of timing error before an adjacent slot is corrupted (Figure 4.1).

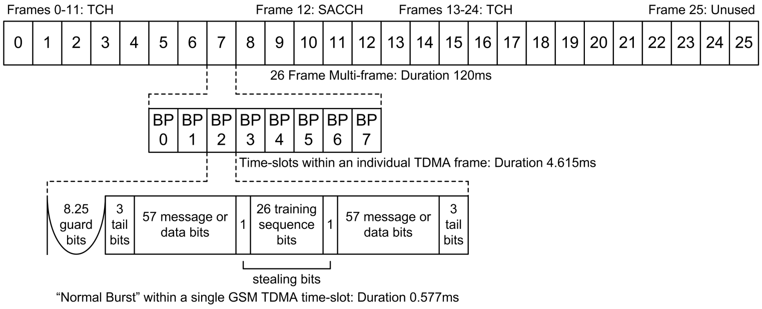


Figure 4.1: Time Division Multiple Access in GSM

The TA is used to adjust the timing to get the MS within the *(slot + guard)* if the phone is within approximately 35km of the BTS. The MS has to advance its timing by the round-trip propagation time, because its time is $x\,\mu$s off the BTS' time due to the outbound propagation delay, and then it has to allow for another $x\,\mu$s for the way back.

The TA can be used to calculate the distance of the MS from the BTS with a granularity of 550 meters, such that $d_{TA} = 3.70 \cdot 3 \cdot 10^8 \cdot \frac{1}{2} = 550m$, where d_{TA} is the distance per bit period of TA, where 1 bit period takes 3.70μs to propagate the round-trip of $BTS \rightarrow MS \rightarrow BTS$ at approximately the speed of light[2] [37]. The TA measurement is typically used to calculate the radius of an MS's location area:

$$(TA \cdot 550) - 275 \leq d \leq (TA \cdot 550) + 275$$

where d = calculated distance between MS and BTS.

The distance is $\pm$ 275 meters due to the TA being rounded to the nearest bit period, such that $0m \leq \text{TA}_0 \leq 275m$; $275m \leq \text{TA}_1 \leq 825m$; and so on.

The initial TA estimation is obtained via the *Immediate Assignment* layer 3 message as a result of the MS transmitting access bursts on the Random Access Channel (RACH), using a TA of 0. The BTS detects a burst transmission on the RACH and measures the delay of the signal relative to the expected signal from an MS at 0 distance. To keep track of the propagation delay, normal bursts sent by the MS are monitored by the BTS. Changes to the delay by more than 1 bit period result in the advancement or

[2] More accurately, the speed of 900 - 1800 MHz radio waves traveling in air medium.

retardation of the TA by 1 bit period, signaled on the Slow Associated Control Channel (SACCH). When a non-synchronized handover occurs, the MS transmits access bursts. The TA, as measured from the new BTS, is sent in a *Physical Information* message, which also signals the MS to stop sending access bursts.

The MS must time its transmissions according to signals received from the BTS, such that transmissions measured at the MS antenna are 468.75 (3 time-slots) $-$ *TA* bit periods behind the transmission received from the BTS [41]. The accuracy of this measurement is affected by the TA being rounded to the nearest bit period, and the tolerance of timings, which is a maximum of ± 1 bit period [41]. The effects of this may be apparent near TA boundaries.

The timing advance measurement can be obtained either from the MS or network operator and can be used to determine the distance an MS is from a given BTS. This measurement is only available from the serving cell when the MS is active, although when idle the MS may be paged without the subscriber noticing.

4.4.1 Experimentation in Spoofing TA Measurement

Experiments were conducted in an attempt to validate that the TA measurement is resistant to intentionally changing the timing of GSM signaling from an MS. It was envisaged that timing changes would result in disruption or disconnection.

The experiments took place using the following Base Stations observed in the vicinity of the MS:

- *Cell Identifier: 8887*[3] (Broadcast Channel (BCCH) Absolute Radio Frequency Channel Number (ARFCN) = 768; Base Station Identity Code (BSIC) = 74)
- *Cell Identifier: 8888* (BCCH ARFCN = 773; BSIC = 77)

An Ericsson TEMS phone[4] was used to conduct the experiments. Since the phone does not allow the manipulation of physical layer messages or messages transmitted in the Synchronization Channel (SCH), the MS was forced to perform a handover, such that the effect of changing timing of transmissions could be observed.

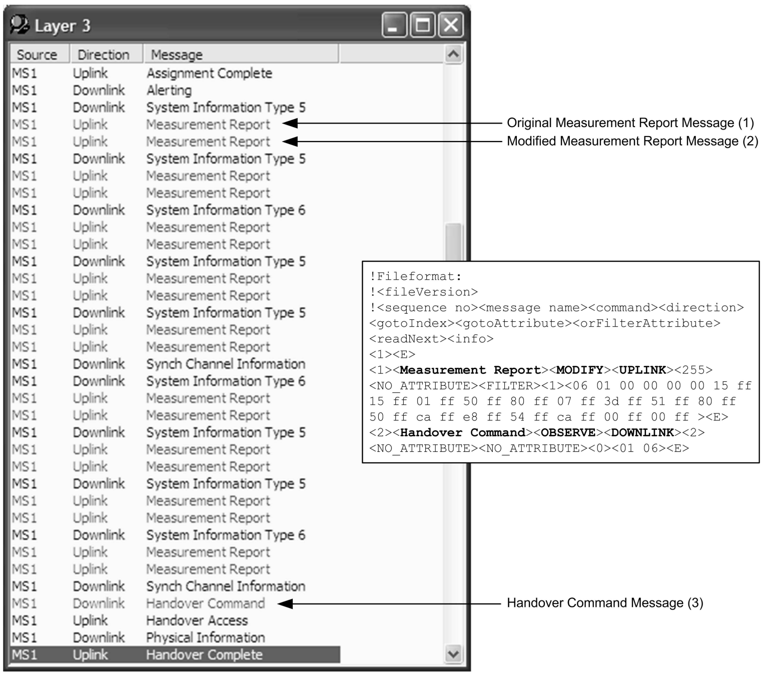
In order to force a handover to another BTS, it is necessary to modify the Measurement Report Messages that are sent to the Base Station Controller (BSC). If the serving

[3] The MCC, MNC and LAC are not given as we do no wish to identity the network on which the tests were conducted.

[4] A hardware/software solution designed to facilitate network troubleshooting and performance optimization using an Ericsson TEMS T28s(GSM900/1800) cellular phone with TEMS Investigation 1.2 software.

BTS received signal strength indication (RXLEV) is insufficient or significantly lower than the RXLEV of a neighboring BTS, the BSC may initiate a handover to the BTS which the MS observed with the highest RXLEV.

A Layer-3 modification script was written, such that *Measurement Report* messages were modified until a *Handover Command* was observed, as illustrated in Figure 4.2. The modification involved changing the serving BTS to an RXLEV of 1 (-110 dBm to -109 dBm), and ensuring the desired BTS has an RXLEV of 63 (the highest level of received signal strength, $<$ -48 dBm). This process is demonstrated in Figure 4.3.



Source	Direction	Message
MS1	Uplink	Assignment Complete
MS1	Downlink	Alerting
MS1	Downlink	System Information Type 5
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	System Information Type 5
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	System Information Type 6
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	System Information Type 5
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	System Information Type 5
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	Synch Channel Information
MS1	Downlink	System Information Type 6
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	System Information Type 5
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	System Information Type 5
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	System Information Type 6
MS1	Uplink	Measurement Report
MS1	Uplink	Measurement Report
MS1	Downlink	Synch Channel Information
MS1	Downlink	Handover Command
MS1	Uplink	Handover Access
MS1	Downlink	Physical Information
MS1	Uplink	Handover Complete

Original Measurement Report Message (1)

Modified Measurement Report Message (2)

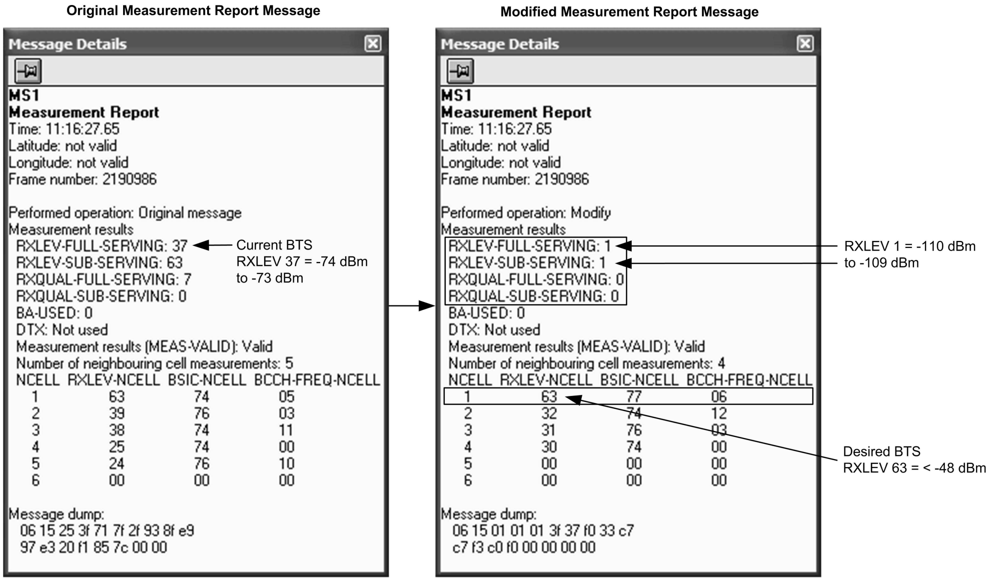
Handover Command Message (3)

```
!Fileformat:
!<fileVersion>
!<sequence no><message name><command><direction>
<gotoIndex><gotoAttribute><orFilterAttribute>
<readNext><info>
<1><E>
<1><Measurement Report><MODIFY><UPLINK><255>
<NO_ATTRIBUTE><FILTER><1><06 01 00 00 00 00 15 ff
15 ff 01 ff 50 ff 80 ff 07 ff 3d ff 51 ff 80 ff
50 ff ca ff e8 ff 54 ff ca ff 00 ff 00 ff ><E>
<2><Handover Command><OBSERVE><DOWNLINK><2>
<NO_ATTRIBUTE><NO_ATTRIBUTE><0><01 06><E>
```

Figure 4.2: Procedure to Force Handover to Desired BTS

This procedure was used to successfully force the MS to handover from the BTS with BCCH 768 to the BTS with BCCH 773. This would result in the MS transmitting access bursts on the RACH using a TA of 0, allowing BTS_{773} to calculate the round-trip propagation delay. Before the MS is able to transmit, it must wait for a Physical Information message that details the new TA. The test was designed to ascertain the effects of changing the signal transmission timing by modifying the Physical Information messages before they are applied by the MS.

A layer-3 modification script that modified the Physical Information messages was activated, such that Physical Information Messages were continually changed from a



Original Measurement Report Message

```
Message Details
MS1
Measurement Report
Time: 11:16:27.65
Latitude: not valid
Longitude: not valid
Frame number: 2190986

Performed operation: Original message
Measurement results
  RXLEV-FULL-SERVING: 37
  RXLEV-SUB-SERVING: 63
  RXQUAL-FULL-SERVING: 7
  RXQUAL-SUB-SERVING: 0
  BA-USED: 0
  DTX: Not used
  Measurement results (MEAS-VALID): Valid
  Number of neighbouring cell measurements: 5
  NCELL  RXLEV-NCELL  BSIC-NCELL  BCCH-FREQ-NCELL
     1        63          74           05
     2        39          76           03
     3        38          74           11
     4        25          74           00
     5        24          76           10
     6        00          00           00

Message dump:
   06 15 25 3f 71 7f 2f 93 8f e9
   97 e3 20 f1 85 7c 00 00
```

Current BTS
RXLEV 37 = -74 dBm
to -73 dBm

Modified Measurement Report Message

```
Message Details
MS1
Measurement Report
Time: 11:16:27.65
Latitude: not valid
Longitude: not valid
Frame number: 2190986

Performed operation: Modify
Measurement results
  RXLEV-FULL-SERVING: 1
  RXLEV-SUB-SERVING: 1
  RXQUAL-FULL-SERVING: 0
  RXQUAL-SUB-SERVING: 0
  BA-USED: 0
  DTX: Not used
  Measurement results (MEAS-VALID): Valid
  Number of neighbouring cell measurements: 4
  NCELL  RXLEV-NCELL  BSIC-NCELL  BCCH-FREQ-NCELL
     1        63          77           06
     2        32          74           12
     3        31          76           03
     4        30          74           00
     5        00          00           00
     6        00          00           00

Message dump:
   06 15 01 01 01 3f 37 f0 33 c7
   c7 f3 c0 f0 00 00 00 00
```

RXLEV 1 = -110 dBm
to -109 dBm

Desired BTS
RXLEV 63 = < -48 dBm

Figure 4.3: Modified Measurement Report Layer 3 Message

TA of 1 (825m) to a TA of 22 (12,375m)[5] until a Disconnect message was received. This is illustrated in Figure 4.4.

The result of this process was not as anticipated, as the TEMS cell phone/software did not continue to modify Physical Information Messages after the first modification. As can be seen in Figure 4.4, the MS appears to have been flooded with Physical Information Messages reflecting the correct TA, before it was disconnected.

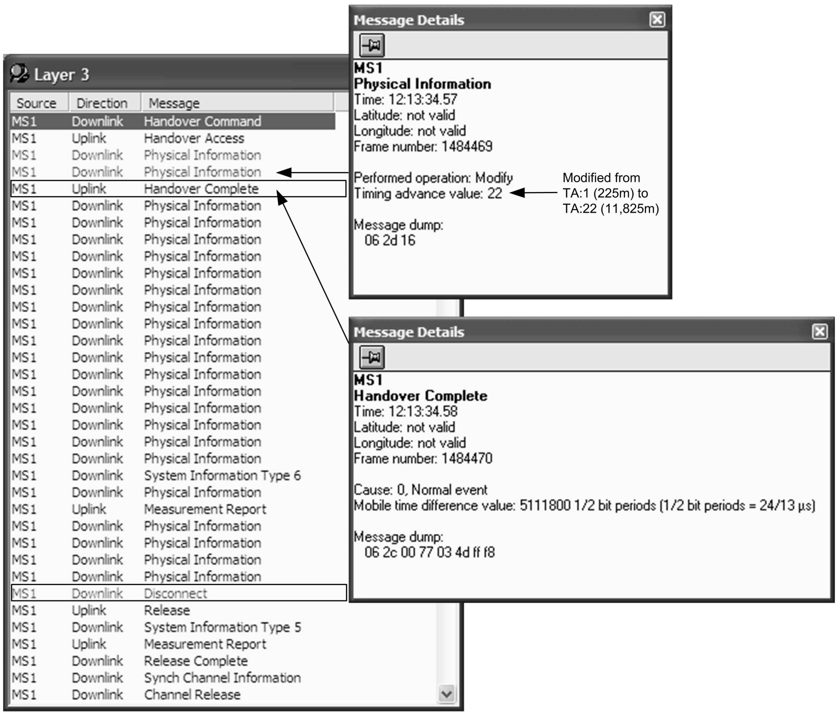
A voice call was active while this procedure took place. The voice call was audibly disrupted in a significant manner before the call resumed normally (with a TA of 1), or was disconnected. The tests were able to confirm beyond reasonable doubt that spoofing of TDMA signaling is unlikely and at the least would require significant cell phone modifications, and significant expertise to mount an attack.

4.4.2 Limitations of the Timing Advance measurement

From the behavior observed in § 4.4.1, it is assumed that tampering with the TA will cause significant disruption to communications or disconnection. Ideally, non-compliant cell phones should be disconnected by the BTS, but this is dependent on the vendor implementation of a BTS.

There is one known limitation of the TA as a location measurement. The MS must

[5] Sufficient TA difference to cause a collision with the adjacent time-slot.



Layer 3

Source	Direction	Message
MS1	Downlink	Handover Command
MS1	Uplink	Handover Access
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Uplink	Handover Complete
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	System Information Type 6
MS1	Downlink	Physical Information
MS1	Uplink	Measurement Report
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Physical Information
MS1	Downlink	Disconnect
MS1	Uplink	Release
MS1	Downlink	System Information Type 5
MS1	Uplink	Measurement Report
MS1	Downlink	Release Complete
MS1	Downlink	Synch Channel Information
MS1	Downlink	Channel Release

Message Details

MS1
Physical Information
Time: 12:13:34.57
Latitude: not valid
Longitude: not valid
Frame number: 1484469

Performed operation: Modify
Timing advance value: 22

Message dump:
06 2d 16

Modified from
TA:1 (225m) to
TA:22 (11,825m)

Message Details

MS1
Handover Complete
Time: 12:13:34.58
Latitude: not valid
Longitude: not valid
Frame number: 1484470

Cause: 0, Normal event
Mobile time difference value: 5111800 1/2 bit periods (1/2 bit periods = 24/13 µs)

Message dump:
06 2c 00 77 03 4d ff f8

Figure 4.4: Procedure to Modify Downlink TA Correction After Forced Handover

be trusted to time its transmissions according to signals received from the BTS, such that transmissions measured at the MS antenna are 468.75 (3 time-slots) - TA bit periods behind the transmission received from the BTS [41]. It is possible that MS could advance all its transmissions by $\frac{\text{dist. of MS to BTS}}{550}$ bit periods. Cellular phones/devices typically have GSM implemented in a single chip or chip-set, where the ability to control such parameters would not be available. Significant engineering skills and reasonable finances would be required to perform such an attack.

4.5 Proposed Method of Obtaining Location with a High Level of Assurance

The proposed location system uses the Cell Global Identifier (CGI) and the TA measurements for calculating the location. A number of constraints on the use of these measurements are defined in this section, such as the use of a different representation of the location area to ensure the location provided by the system encapsulates the entire area an MS could possibly be. This method provides a higher assurance of location than the commonly used GCID-based location, as a user is able to easily spoof their location by forcing a handover to the furtherest visible BTS. The TA in effect ensures such activity is detected.

One of the unique features of the TA is that it must be approximately correct in order for the MS to transmit in the correct slot. Because transmissions must be timed such that they arrive in the correct time-slot at the BTS, the delay of transmissions cannot exceed the guard period, otherwise the transmitted burst would collide with an adjacent burst. As such, the TA must be correct beyond the guard period otherwise transmission will be corrupted. A guard of 8.25 bit periods in effect means that the effective radius of any cell without timing advance can be a maximum of 4400m based on 550m per bit period. Locations within this area can be undetectably spoofed.

In terms of the measurement granularity provided by TA values, a TA measurement less than 9 cannot be assured. The TA can be spoofed for values less than 9 due to the size of the guards as detailed in § 4.4. A TA of 8 would indicate a distance between 3925m and 4675m ($(TA \cdot 550) \pm 275$, where $TA = 8$), and as such could still be spoofed. A TA of 9 indicates a minimum distance of 4675m, which is therefore the minimum distance before the TA is tamper-resistant. As such, the MS is assured of being between 0 and 4675m from the BTS, where the TA is less than 9. For TA values greater than 9 and less than 63, the MS is assured of being between 4675 and 34925m from the BTS. The value of 34925m is the maximum distance ($(63 \cdot 550) + 275$, where 63 is the maximum timing advance supported by standard GSM [41])

The propagation path cannot be assumed to be a direct line-of-sight as the signal may travel a longer path. What can be assumed is that the minimum timing advance that can be obtained is the TA representing the signal propagating the line-of-sight. This assumption holds true, as radio waves cannot exceed the speed of light in order to arrive at the BTS earlier.

This provides us with a fundamental grounding for tamper-resistance, in that an MS attempting to spoof its location from a distant location will have a timing advance indicative of at least the minimum line-of-sight distance between the BTS and the MS ± 1 bit period of tolerance. As such, a malicious user cannot spoof their location by use of a high gain antenna or other means of extending the antenna, whilst being up to 35km away from the desired BTS. The attacks based on antenna extension described by Gabber and Wool in [55] can therefore be detected by the TA measurement.

4.6 Practical Use of the TA in a Location System

In order to implement a location solution based on the proposed method, it is necessary to define a maximum allowable distance from a cell. For example:
Given a cell 0001, let the maximum allowable distance from the BTS be 4675m, such that the line-of-sight distance from the BTS must be less than 4675m.
$dist_{max} = 4675m$
$dist_{min} = 0$ ($dist_{min}$ must always = 0, as any value above 0 can be spoofed)

Example 1: $\mathrm{MS}_{\mathrm{test}}$ is within defined location area
$\mathrm{MS}_{\mathrm{test}}$: Serving Cell $= 0001;\ TA = 0$
$d_c = (TA \cdot 550) + 275 = 275m$
Test: $dist_{min} \leq d_c \leq dist_{max}$
$0 \leq 275 \leq 4675 = true$

Example 2: $\mathrm{MS}_{\mathrm{test}}$ is outside defined location area
$\mathrm{MS}_{\mathrm{test}}$: Serving Cell $= 0001;\ TA = 10$
$d_c = (TA \cdot 550) + 275 = 5775m$
Test: $dist_{min} \leq d_c \leq dist_{max}$
$0 \leq 5775 \leq 4675 = false$

It is possible that the MS in example 2 was legitimate but was adversely affected by the environment it was in. Under certain circumstances, the TA will be greater than the required minimum TA due to propagation effects in the environment of the MS. (Test results in Table 4.6 and Figures 4.5 and 4.6 illustrate this effect.) In this case a potentially legitimate user will be denied access. It is important to note that the TA calculation is based on the first arrived propagation path that has a significant RX level. The use of training sequence bits in signal bursts for equalization ensure that out of phase reflected signals are discarded. As such, the TA should represent the line-of-sight propagation path in most cases.

Table 4.6 details trials that were conducted and the resulting distances. The trials involved forcing handovers to each available BTS from three defined locations. One of these locations was in the central business district of Brisbane, the other two in suburban locations north of Brisbane. The technique used to facilitate forced handovers is detailed in § 4.4.1. Measurements were obtained from each BTS three times from at same location. The minimum observed distance is $(TA \cdot 550) - 275$ and the maximum observed distance is $(TA \cdot 550) + 275$. The measurement data collected from these trials is illustrated in Figures 4.5 and 4.6.

	Actual Dist. (m)	**Range of TA**	**Max Observ. Dist. Range (m)**	**Avg. Dist. (m)**
City 1	1035	2. . .3	1375. . .1925	1742
City 2	250	1. . .2	825. . .1375	1008
City 3	2436	5. . .5	3025. . .3025	3025
City 4	1233	3. . .3	1925. . .1925	1925
City 5	489	1. . .1	825. . .825	825
Suburb-A 1	1043	2. . .3	1375. . .1925	1650
Suburb-A 2	1006	2. . .3	1375. . .1925	1650
Suburb-A 3	2710	4. . .6	2475. . .3575†	3025
Suburb-B 1	901	1. . .3	825. . .1925†	1454

†In these results, the initial Timing Advance estimate contained a timing error less than 1 bit period. This caused the TA, after rounding, to be less than the line-of-sight distance. This was corrected shortly after the initial TA estimate.

Table 4.1: TA Measurement Trials.

These results have identified issues in the TA accuracy caused by bit rounding. This tends to be an issue close to the TA boundaries. It can be stipulated that the more time a given BTS has to calculate the TA, the more accurate it will be. The initial TA estimate tends to be less accurate than the TA of a MS in the same fixed location after a short period of time has elapsed. It is also possible within the 0m to 4675m range to have effects such as a TA being less than the line-of-sight distance. This is because the guard is sufficient within this distance to protect adjacent transmission from being corrupted.

Location at this accuracy can be used for numerous applications including credit card transaction audits for m-commerce applications or authorization to perform an Internet banking transfer. In this case, suburb-level granularity is acceptable. To increase the accuracy of the location system, it may be necessary to perform a forced handover onto a cell in the required area. This can be performed either by the network operator or the MS. A forced handover has been achieved with an Ericsson TEMS phone as detailed in § 4.4.1.

This location determination method requires no changes to current GSM infrastructure. A significant advantage of this implementation is that it only requires a GSM network with the appropriate location services implemented. All the information required for location calculation is already obtainable from any LCS98 compliant location system. Our prototype (Refer to § 4.11) has been developed using the LCS98 compliant Ericsson MPP (Mobile Positioning Protocol) [29].

Additional fields could be added to the MPP that quantify assurance of the measurements required and returned. For example, the initial TA estimation may be less

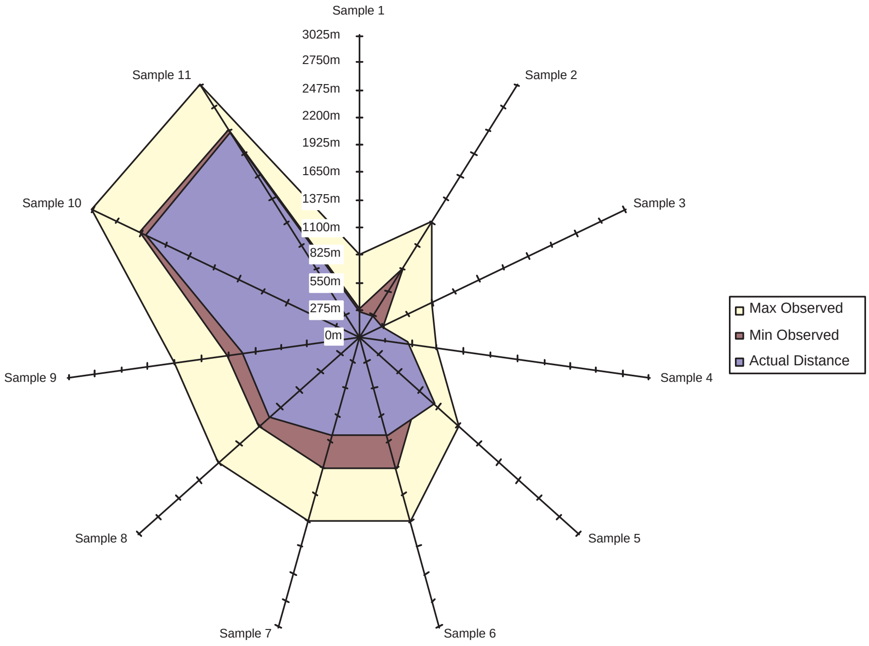


Figure 4.5: City Samples of TA Measurements

accurate than the TA obtained after channel establishment. Hence, the method of obtaining the location measurement could be adapted to accommodate the assurance level specified.

4.7 Geographical Data Representation of a Trusted Location Area

In both TOA and TA based location acquisition methods, location data is not represented as a point, rather a location area. It is important for security applications, that the location area represented encapsulates the entire area an MS could possibly be with a high degree of certainty.

The most flexible location area representation that provides support for the largest number of location areas (and location acquisition methods) is a polygon. An entire location area can be encapsulated within a polygon, such that location calculations using polygons are trivial. Access control applications can therefore perform fast and

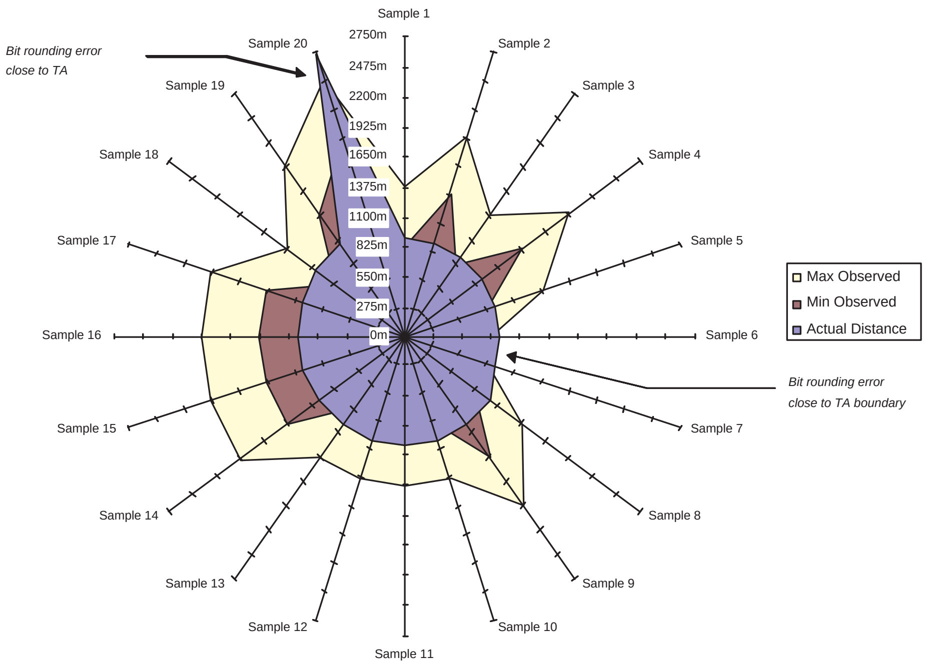


Figure 4.6: Suburb Samples of TA Measurements

efficient verification that a given polygon (representation of a user's location) is within a defined (polygon-based) location area.

Figure 4.7 illustrates a polygon representation of a location area determined from TOA measurements, and Figure 4.8 illustrates the representation of a location area determined from a TA measurement.

The Universal Geographical Area Description specification, as part of the LCS98 GSM specifications [46], specifies a polygon location area as an arbitrary shape described by an ordered series of points. The minimum number of points is 3, and the maximum supported by [46] is 15. The points are connected in the order given, such that the last point is connected to the first. A connecting line shall not cross another connecting line and two successive points must not be diametrically opposed.

While the polygon representation of a location area is ideal for security applications, there are a number of issues posed by its use with the The Mobile Positioning Protocol (MPP) version 5.0.

- *MPP does not support the specification of position area type in the location request.* The position area can only be influenced though the combination of

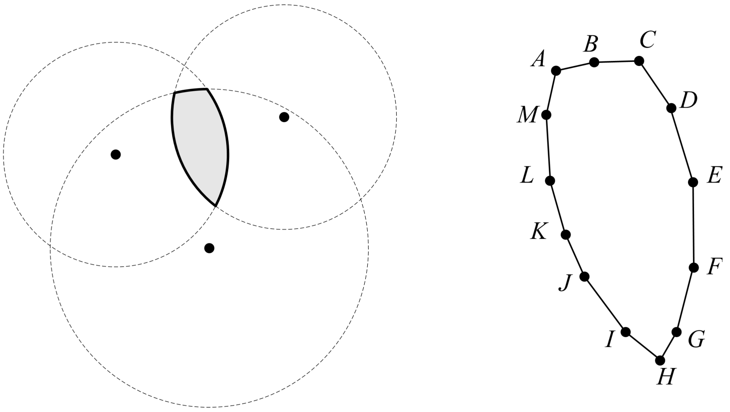


Figure 4.7: Representing a TOA Location Area as a Polygon

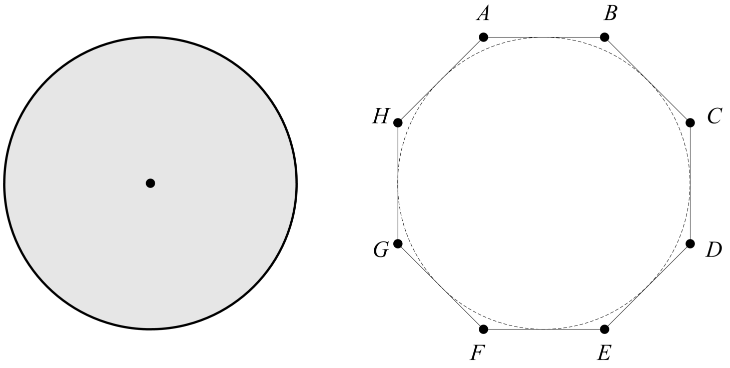


Figure 4.8: Representing a TA Location Area as a Polygon

Quality-of-service parameters *RESP_TIME* and *HORIZON_ACC*. By selecting 0 for both these parameters, it is most likely that the TA location acquisition method will be used.

- *MPP does not indicate the location acquisition method used.* As there is no provision for returning the location acquisition method in the location response, the TOA cannot be used as it cannot be determined whether it or another location acquisition technique such as E-OTD (which is easily spoofable and not suitable for security applications) with the same type of location area was used in the location calculation.

- *The format of a location area response for the TA measurement is undefined.* There are three possible location area representations[6] that could be used for a

[6] Refer to Appendix G for an overview of the location areas supported by GSM LSC98.

TA-based location area:

1. *Ellipsoid point with uncertainty circle.* This location area is characterized by a set of coordinates (an ellipsoid point) and a distance r, the radius of the circular area.
2. *Ellipsoid arc.* This location area improves on the accuracy of the previous location area by restricting the circle to a sector defined by two angles measured clockwise from the north. An additional uncertainty radius is introduced, such that r_1 describes the radius and r_2 describes the uncertainty[7].
3. *Polygon.* This location area is an arbitrary shape described by an ordered series of points (minimum 3, maximum 15), connected in the order given.

- *The method of calculating the polygon from the numerous GSM location acquisition techniques is undefined.* If a polygon representation of a TA-based location area is returned in a location response, an application cannot be sure that the polygon representation encapsulates the circular location area in a manner useful for security applications. This is because the method of deriving the polygon area from the circular area is undefined in the LCS98 specifications.

As the ellipsoid arc location area is the most commonly used representation of the TA-based location area, the following section will discuss access control and auditing methods using both polygons and circular location areas.

4.8 Location-based Auditing and Access Control using the Proposed System

This section discusses how location assurance mechanisms can be used in security services.

For audit applications, a polygon or ellipsoid arc (depending on the location result returned by a location server) can be stored as part of an audit trail. In the case that an ellipsoid arc is returned, only the coordinates of the BTS, radius r_1 and the uncertainty r_2 should be used to form a circular location area, such that it is certain with a high level of confidence that the MS of the subject is within this location area.

A possible attack on a location audit would involve forcing a handover to the furthest visible BTS, such that the represented circular location area is excessively large.

[7] r_2 is 550m for the TA measurement.

To limit the possibility of such an attack, there may optionally be an access control decision on the operation that is audited, such that there is a limit to the size of the radius (TA) where closer BTSs are available. An example of an audit application is a credit card payment protocol that includes the location of a MS in the transaction audit.

Where location is used to augment access control processes, ACLs (Access Control Lists) must contain the necessary location data to make authorization decisions. An ACL is associated with every object in a system and determines the permissions of a user or role on a given object by the membership of that user in the ACL.

In the proposed location-based access control system, ACLs would contain the object ID, user/role IDs, object permissions, and the location at which these permissions are valid. A single ACL may have many permissions, each permission with the option of being associated with a location constraint. A location constraint has an operation mode of allow or deny as denoted by *LAC_MODE* in Figure 4.9, such that either a given permission is only valid in the associated location areas, or is valid for all locations except the associated location areas. In this way, the access control logic of an application only grants permissions based on location constraints being met.

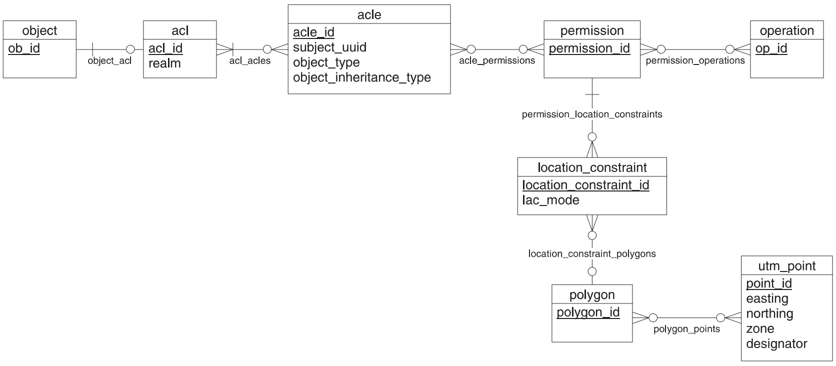


Figure 4.9: Conceptual Data Model of ACLs Supporting Location-based Access Control

Location areas associated with a permission are represented as polygons. These polygons define the area in which a subject is given access to an object. If a polygon is returned from the location server, the polygon can be tested to determine if it is within the defined location area. This is done by iteratively verifying that the set of points the polygon is composed of are not inside the defined location area. The Jordan curve theorem as detailed in Appendix E is used to perform this verification.

If the location server returns an ellipsoid arc, the following process is performed in order to determine if the location area is within a defined polygon location area. We propose a two-step verification approach using the radius r_1 and the uncertainty r_2 of the ellipsoid arc to form a circular location area:

1. Determine if the BTS coordinates of the circular location area are inside the polygon using the Jordan curve theorem (Refer to Appendix E).

2. If the BTS coordinates are within the polygon, determine if entire MS circular location area is within the polygon by checking that all of the points of the polygon $(x_{1..n}, y_{1..n})$ are outside the circular location area such that:
   if $r \geq \sqrt{(x_{1..n} - x_{\mathrm{loc}})^2 + (y_{1..n} - y_{\mathrm{loc}})^2}$, *point is not inside circular location area.*
   where $x_{\mathrm{loc}} = \text{BTS Easting}$, $y_{\mathrm{loc}} = \text{BTS Northing}$, $r = (550 \cdot \text{TA}) + 275$
   If true for all points of the polygon, it can then be assumed for most cases[8] that no lines of the polygon intersect with the circle, and as such, that the circular location area is within the polygon (Figure 4.10). The access control mechanism can either grant or deny access to a user depending on whether the location area is within the polygon.

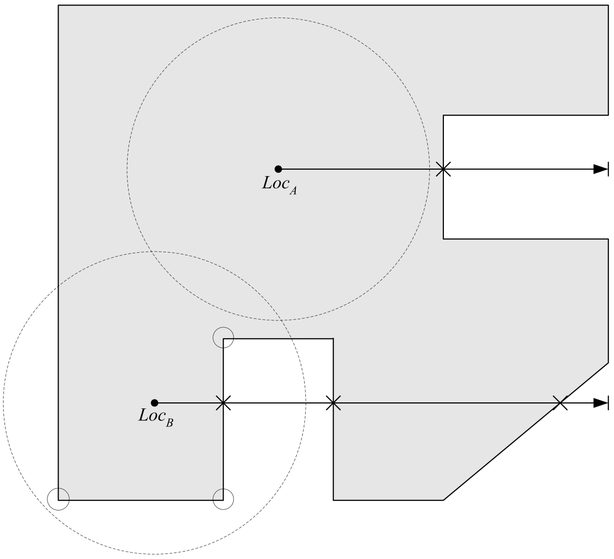


Figure 4.10: Testing a Circular Location Area Within a Polygon

[8] It is assumed that polygons used in the access control system are simple shapes that comply to the definition in § 4.7.

4.9 Cellular Phone Association Protocol

There are a number of prerequisites for the use of location services, the most important of which is the association of the location acquisition device to the corresponding user's session. Association of a phone to a web session is particularly important when using web services, as the session is typically not terminated at the phone, but rather at another terminal. This can be achieved by authenticating the location acquisition device, proving that it is in the possession of a legitimate user for a given user's session.

As the security of GSM affects the security of the association protocol, the GSM security protocol is first reviewed before introducing the proposed association protocol. The following subsections detail GSM authentication and session encryption, the proposed one-time password based association protocol, and a public key variation.

4.9.1 GSM Security Services

GSM provides the following security services [49]:

- *Subscriber identity (IMSI) authentication.* Provides protection against unauthorized network use and user impersonation;

- *Subscriber identity (IMSI) confidentiality.* Provides protection against tracing the location of a mobile subscriber by monitoring the signaling exchanges. The Temporary Mobile Subscriber Identity (TMSI) is used to identify a subscriber instead of the IMSI, which is stored securely in the SIM;

- *User data confidentiality on physical connections.* Ensures privacy of user information on traffic channels;

- *Connectionless user data confidentiality.* Ensures privacy of user information on signaling channels (including SMS messages); and

- *Signaling information element confidentiality.* Ensures privacy of user-related signaling elements.

GSM authentication is facilitated using a challenge-response protocol, using a pre-established long-term secret key K_i between the Subscriber Identity Module (SIM) and the Authentication Center (AuC). The authentication protocol is illustrated in Figure 4.11. In the following protocol description, R_{AuC} denotes a random number, *RAND*, generated by the AuC. $S\{R_{AuC}\}_{Ki}$ denotes a signed response, *SRES*, in which R_{AuC} is the data that is signed using secret key K_i. K_i denotes the individual subscriber secret key used for authentication, and K_c denotes the established session key for ciphering.

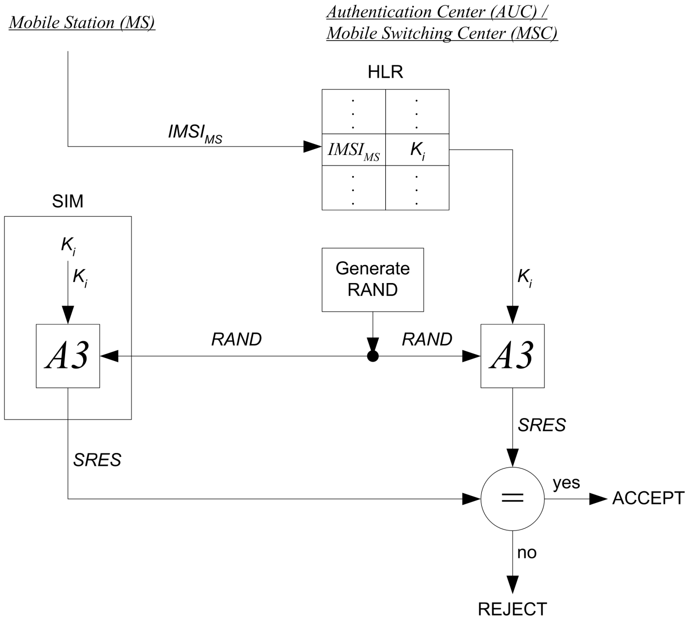


Figure 4.11: GSM Challenge-response Authentication

$$
\begin{array}{rcll r}
MS & \rightarrow & BSS/MSC/VLR: & IMSI_{MS} & (1)\\
BSS/MSC/VLR & \rightarrow & HLR/AuC: & IMSI_{MS} & (2)\\
BSS/MSC/VLR & \leftarrow & HLR/AuC: & IMSI_{MS}, Kc_{MS-AuC}(1..n), & \\
 & & & R_{AuC}(1..n), S\{R_{AuC}\}_{Ki}(1..n) & (3)\\
MS & \leftarrow & BSS/MSC/VLR: & R_{AuC}(1) & (4)\\
MS & \rightarrow & BSS/MSC/VLR: & S\{R_{AuC}(1)\}_{Ki} & (5)\\
MS & \leftarrow & BSS/MSC/VLR: & \{LAI, TMSI_{MS}\}_{Kc_{MS-AuC}(1)} & (6)
\end{array}
$$

The MS sends IMSI of its subscriber to the BSC/MSC, which performs a lookup in the VLR to determine whether a new authentication vector is required (1). Assuming first time authentication, the Visitor Location Register (VLR) sends the IMSI to the Home Location Register (HLR), which can provide an authentication vector, generated by the AuC (2). An authentication vector contains n random numbers $R_{AuC}(1..n)$, the signed responses corresponding to the random numbers $S\{R_{AuC}\}_{Ki}$ using secret key Ki, generated using the A3 algorithm, and a session key $Kc_{MS-AuC}(1..n)$ used for ciphering messages, generated using $R_{AuC}(1..n)$, Ki and the A8 key generation algorithm.

The Authentication vector is returned to the VLR (3), which via the BSS/MSC sends $R_{AuC}(1)$ to the MS (4). The MS computes the signed response $S\{R_{AuC}(1)\}_{Ki}$ using the *Ki* stored on its SIM card and the A3 algorithm, and returns it to the BSS/MSC (5). The response from the MS is passed to the VLR which verifies that the $S\{R_{AuC}(1)\}_{Ki}$ returned by the MS is the same as the $S\{R_{AuC}(1)\}_{Ki}$ in the authentication vector. If verification is successful, the VLR generates the TMSI and returns it and the LAI to the MS enciphered using $Kc_{MS-AuC}(1)$.

Figures 4.11 and 4.12 illustrate general authentication and ciphering procedures used in GSM. For re-authentication, location updating, and other procedures refer to [50].

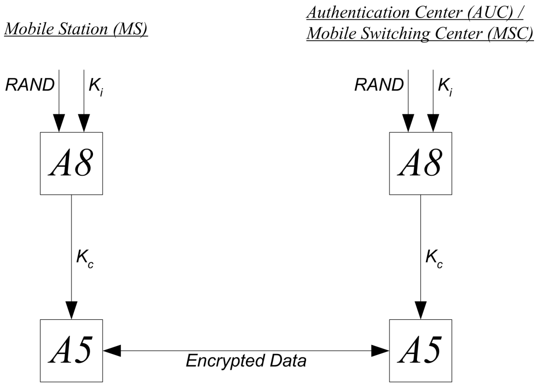


Figure 4.12: GSM Session Key Generation and Encryption

4.9.2 Association using One Time Passwords

In the proposed scheme, the covert channel is the SMS data bearer in GSM. The primary communications channel for a given transaction is via the web. A message is sent via this channel such that only the possessor of the cell phone, who has authenticated to the network operator, is able to receive this message. It is considered unlikely that an attacker is able to compromise the security of the primary channel, protected by Secure Sockets Layer and the localized covert channel, protected with GSM encryption.

The security of this scheme is afforded by the use of a covert channel to communicate a random challenge to the verifier, and a time constraint, such that it is unlikely that both communication channels are compromised within the time interval. Replay attacks are not likely, as a random number is used to generate the one-time password which expires within the defined time interval.

In the protocol below, r_A denotes a random number generated by A. $A \leftharpoonup B$ denotes a communication from B to A over a covert channel. $A \leftarrow B$ denotes a communication

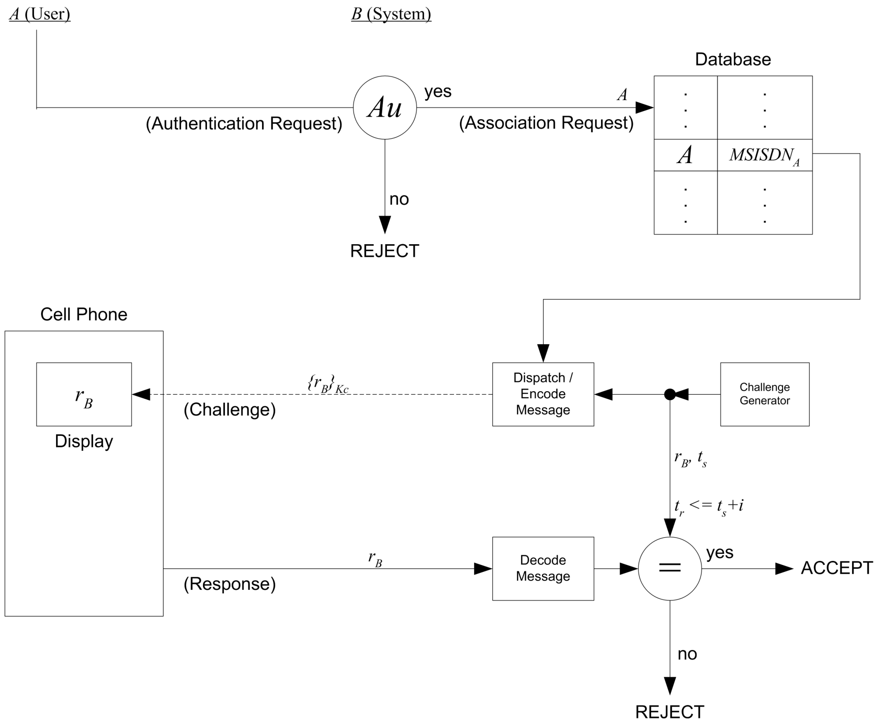


Figure 4.13: Cellular Phone Association Protocol

from B to A over the primary communications channel.

$$A \leftharpoonup B : \{r_B\}_{K_C} \qquad (1)$$
$$A \rightarrow B : r_B \qquad (2)$$

This scheme requires that a user authenticates to the service that requires trusted location before enacting the association protocol. This is required to prevent a malicious party from continuously requesting one-time passwords for users, flooding them with one-time passwords. The method of authentication is not specified and is outside the scope of the association protocol. It is assumed some form of mutual authentication takes place.

An example is a web service using SSL, for which a username/password or client SSL certificate are used to authenticate to the service. The user is able to authenticate the server using the server's SSL certificate. Once the user has been successfully authenticated, the web service sends a request to the server that implements the association protocol, such that a one-time password is dispatched to the Short Message Service Center (SMSC) for the MSISDN associated with the authenticated user. The one-time password is a truncated version of the random challenge, Base-64 encoded and dispatched as a *class 0* SMS message. A *class 0* SMS message appears immediately on the screen of an MS without any user interaction and without being stored on the user's SIM card or MS memory.

The user enters the characters of the one-time password that appear on the MS screen into the appropriate field on the web page of the web service session. The server implementing the association protocol verifies that the decoded random challenge is the same as the generated challenge, and that is was received within the defined time interval.

This protocol provides some assurance that the user who is using a web session is the same user that possesses the phone. When the location of this device is obtained, one can be reasonably satisfied that the located phone is the same phone that was authenticated.

4.9.3 Public Key Variation

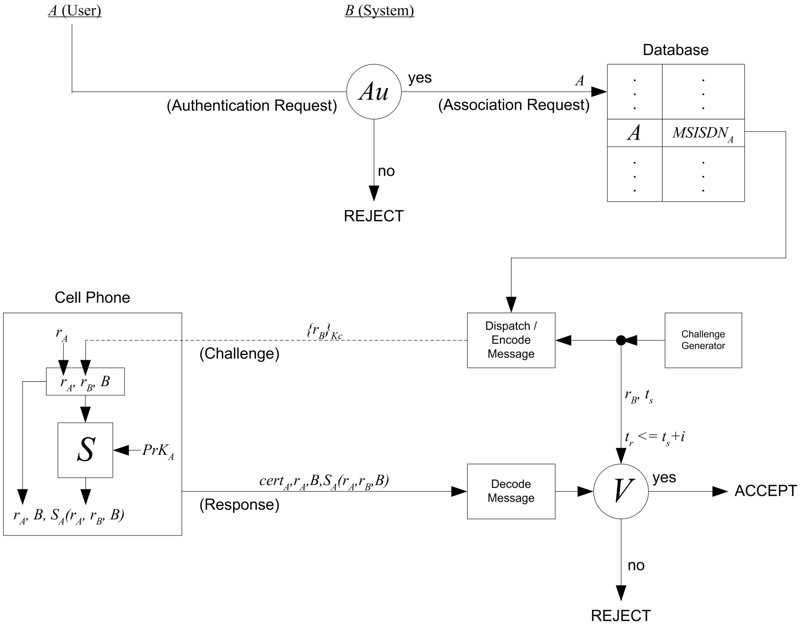


Figure 4.14: Cellular Phone Association Protocol - Public Key Variation

Messages in WAP may be sent either over an SMS or GRPS as a data bearer. This scheme requires that WAP push messages used in this protocol are sent using SMS as a data bearer, such that message (1) is sent over a covert channel.

This protocol is based on the ISO/IEC 9798-3 [62] unilateral authentication protocol with random numbers, using digital signatures. Unilateral authentication could be performed with random numbers in 2 messages. This protocol was chosen as a basis for the WAP-based association protocol for the following reasons:

1. Authentication protocols based on time-stamps are inappropriate due to the infeasibility of synchronizing cell phone clocks with the verifying system's clock;

2. As the client has already authenticated the web server based on SSL server-certificate authentication, only unilateral authentication is required;

3. The number of messages must be minimal due to the cost of communications; and

4. Signature operations are the only cryptographic operations supported by the Wireless Application Protocol (WAP) v2 WML Crypto Library [104].

In the protocol below, r_A denotes a random number generated by A. S_A denotes A's signature mechanism. $cert_A$ denotes the public-key certificate containing A's public key corresponding to the private key used to generate the signature. Where the public key is known to the verifier, the certificate can be omitted. $A \leftharpoonup B$ denotes a communication from B to A over a covert channel. $A \leftarrow B$ denotes a communication from B to A over the primary communications channel.

$$A \leftharpoonup B : \{r_B\}_{K_C} \qquad (1)$$
$$A \rightarrow B : cert_A, r_A, B, S_A(r_A, r_B, B) \qquad (2)$$

A random number is generated by B and sent over the covert channel to A (1). B caches t_s, the time at which r_B was sent to A.

A generates a random number r_A, and signs r_A, r_B and B's identifier. r_A is included in the signature in order to prevent chosen-text attacks.

B verifies that it corresponds to the cleartext identifier, that the signature is valid for r_B and that the signature of r_A is valid for the cleartext r_A. t_r is the time at which B receives (2) from A. i is defined as the time interval in which (2) must have been received and verified for association to be successful, such that $t_r \leq t_s + i$.

A WAP (Wireless Application Protocol) session has the advantage that the communications endpoint is in the phone. This scheme requires a cell phone that supports

WAP 2.0 / TLS (Transport Layer Security) tunneling, and a SWIM (Subscriber Wireless Identity Module).

TLS tunneling[102] is a new feature of WAP 2.0, which allows an end-to-end TLS connection to be established between the MS and the web server by tunneling the TLS protocol through the WAP gateway. Previous versions of WAP only provided support for WTLS(Wireless Transport Layer Security). WTLS only provides encryption to the WAP gateway, where a session is decrypted and re-encrypted to the SSL connection between the WAP gateway and a web server. The most evident problem with this protocol is that it places too much trust on the WAP gateway, and as such is not suitable for applications such as Internet banking.

A SWIM is a SIM and a WIM (Wireless Identity Module) integrated, containing an anonymous random public key pair, certified by the network operator. It is used in authentication for TLS/WTLS and application level security functions such as authentication and non-repudiation operations[103]. Unfortunately, we were unable to source a SWIM card for testing, and therefore cannot comment on the performance of this scheme.

Similarly to the one-time password-based protocol, the user authenticates to the service first. If authentication is successful, the web service sends a request to the server implementing the association protocol to perform association of a given user. A random challenge is generated and dispatched in a WAP push message to the MSISDN corresponding to the authenticated user. The WAP push message not only contains the random challenge, but a Wireless Markup Language (WML) Script that facilities the signature operations and the forwarding of the response to the server over the primary communications channel.

The MS user is prompted to sign the challenge, after which it is sent in a HTTP post to the service, which verifies the signature and that the specified time interval was not exceeded.

Refer to [104] for details on the WAP public key infrastructure model.

4.10 Security Analysis

The following subsections will discuss the security of the proposed one-time password association protocol, the public key variation, and GSM authentication and encryption, which the proposed protocols operate over.

4.10.1 Association Protocol Using One Time Passwords

A number of attacks on identification protocols are identified by Menezes et al. in [71]. These attacks are assessed for the one-time password protocol as follows.

1. *Impersonation.* This attack involves deception whereby an entity purports to be another. Impersonation of another user is avoided through the GSM challenge-response authentication protocol to ensure the cell phone subscriber is not being impersonated, and the application service authentication in which the user has preregistered the MSISDN of the cell phone to be located. Refer to § 4.10.3 for attacks against GSM.

2. *Replay Attack.* This type of attack involves replaying information from a previous protocol execution. Replay attacks are avoided though the use of a covert channel, in which the challenge is dispatched to the cell phone over an encrypted channel. In addition, the challenge must be verified within a defined time interval, reducing the ability of an attacker to compromise both communications channels within the time interval.

4.10.2 Public Key Variation

The assessment of attacks detailed in § 4.10.1 are valid for the public key variation, as well as the following additional attack.

1. *Chosen-text Attack* This type of attack is where an adversary strategically chooses challenges in an attempt to extract information about the user's private key. The attack generally involves the client signing a chosen challenge. The public key variation of the protocol is vulnerable to this type of attack, as there is no method of authenticating the source of the WAP push message containing the challenge. The WML Script Crypto Library does not provide support for verification operations, only signature operations. As such, the user can be coerced into signing any text. It is anticipated that in future releases of the WML Crypto Library, support for other crypto operations will be provided, and as such support for origin authentication can be provided.

4.10.3 GSM Security

The security of GSM affects the ability of an adversary to compromise a covert channel over GSM. While vulnerabilities in GSM security would reduce the complexity of compromising a user's covert channel, the attacker would require knowledge of the user's location in order to mount an attack. There are a number of security vulnerabilities of GSM that are discussed below:

- *Eavesdropping authentication vectors and IMSIs.* An attacker who is able to eavesdrop on a signaling link within the network may be able to obtain authentication vectors and IMSIs, allowing an attacker to impersonate those users. GSM security makes the assumption that the network infrastructure is trusted, and as such there is no requirement for encryption between network components;

- *Authentication vector replay attacks.* As a visited network has no assurance of the freshness of an authentication vector received from the HLR, an attacker could replay an old vector;

- *No origin authentication of authentication vectors received by a VLR.* A VLR has no means to validate the authenticity of received vectors, as such it may be possible for an attacker to invent authentication vectors, provide them to the VLR, and hence access a service impersonating a user or using a fictitiously created user;

- *Eavesdropping of IMSI.* As shown in the GSM first-time authentication protocol detailed in § 4.9.1, the IMSI must be provided in the clear when there is no TMSI available to use. In this case, it is possible to learn the MS location of the subscriber;

- *No authentication of network components or signaling integrity protection.* Due to the lack of integrity protection and origin authentication of network components, it is possible for an attacker to impersonate a BTS, such that it can explicitly demand MSs send their IMSI. In this case location privacy is thwarted also;

- *Attacks against COMP128 that result in the release of Ki.* The secret key K_i can be recovered due to vulnerabilities of the COMP128 algorithm. One reported attack [16] would require approximately 500,000 queries to the SIM, which could recover the secret key in approximately 8 hours. Rao et. al. in [81] introduce a class of side-channel attack performed against COMP128, with the ability to retrieve K_i in as few as 8 chosen plaintexts in less than a minute. An over-the-air attack is theoretically possible if an attacker impersonates a BTS. Using a fake

BTS that exploits GSM authentication protocols, it could be possible to mount an attack on COMP128; and

- *Weaknesses in A5 algorithm.* Attacks against the A5 algorithm facilitate recovery of *Kc* and hence an attacker to access the data encrypted between the MS and BTS. Biryukov et. al. in [12] describe an attack against A5/1 (the strong version of the A5 encryption algorithm) in which K_c can be computed in approximately 1 second based on the output of the A5/1 algorithm for the first 5 minutes of a conversation. Barkan et al. in [8] proposes a method of obtaining K_c in a few milliseconds when A5/2 (the "export" (weakened) algorithm) is used. 4 frames of A5/2 encrypted voice are sufficient to recover the key, resulting in an almost instantaneous attack. Barkan et al. in [8] also describes an active man-in-the-middle attack that would force the MS to use the weaker A5/2 algorithm for sufficiently long to retrieve 4 frames of encrypted voice.

Whilst these attacks exploit significant security flaws that defeat the GSM authentication and encryption mechanisms, it is unlikely that these attacks would be effective against the proposed association protocols. An attack against COMP128 requires physical access to the subscriber's SIM. Assuming an over-the-air attack is possible, an attacker must have knowledge of the subscriber's location in order to clone the SIM. Network operators are able to detect the presence of two cell phones with the same IMSI. Typically, this indicates that a SIM has been cloned, and prompts the network operator to disable access for that particular IMSI.

Attacks against A5 are limited in their ability to compromise the association protocols, as the information sent over the GSM covert channel are time limited and associated with a given TCP session. An attacker would need to initiate a TCP session to the intended destination (e.g. a bank), and be in the vicinity of the user in order to capture the one-time password. This firstly requires successful preauthentication before the one-time password is dispatched. Assuming the preauthentication can be defeated and the A5 attack based on a active man-in-the-middle attack is possible (the described attack by Barkan et al.[8] is only theoretical), an attacker could potentially gain access to the service. This is not possible with the public key version of the association protocol, as the user must willingly perform a public key operation on their phone, the response of which is used for authentication.

A more probable attack could be realized through hijacking a user's TCP session rather than attacking the association protocols. Session hijacking is where an attacker implements a man-in-the-middle attack such that a user connects to a server via an adversary. End-to-end security protocols such as SSL provide little protection from session hijacking, as they rely on weakly bound public key certificates to identify servers and to establish security contexts for symmetric encryption. Most users fail to comprehend the digital trust management present in PKI. Although a user may be warned that the host certificate has changed, most users will accept the new credentials. As such, this type of attack works well in practice. The dsniff collection of tools[9] allows attackers to implement active man-in-the-middle attacks against redirected SSH and SSL sessions.

4.11 Application of GSM Location System to an Internet Banking System

An example application of the tamper-resistant location concepts presented in this chapter can be seen in the Internet banking system that was implemented.

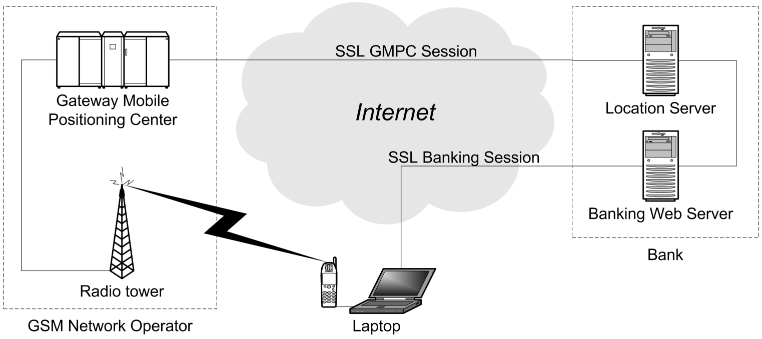


Figure 4.15: High-level Internet Banking System

The banking server was implemented using Java Servlets on Apache Tomcat[10]. The location server was implemented in Java, which made location requests (XML MPP Location Immediate Requests (LIR)) to the prototype Gateway Mobile Location Center (GMLC) over HTTP. The prototype GMLC used for testing is detailed in Appendix H.

The Internet banking system requires an initial authentication in the form of a unique user identifier, and a PIN. (Figure 4.16). Successful authentication of a banking user results in the initiation of the association protocol. A challenge is sent to the

[9] Refer to `http://naughty.monkey.org/~dugsong/dsniff/` for the dsniff tools.

[10] Refer to Apache Jakarta Project for information on Tomcat `http://jakarta.apache.org/tomcat/`.

user's cell phone, which must be entered into the "access code" field within the timeout period.

A user must also select the registered location that they are accessing the service from. These locations must be preregistered with the bank in an offline process, and may contain other constraints such as time of day Internet banking can be accessed from various locations. An example is a user's work location. A user may wish to only allow access to their Internet banking during the hours they are at work.

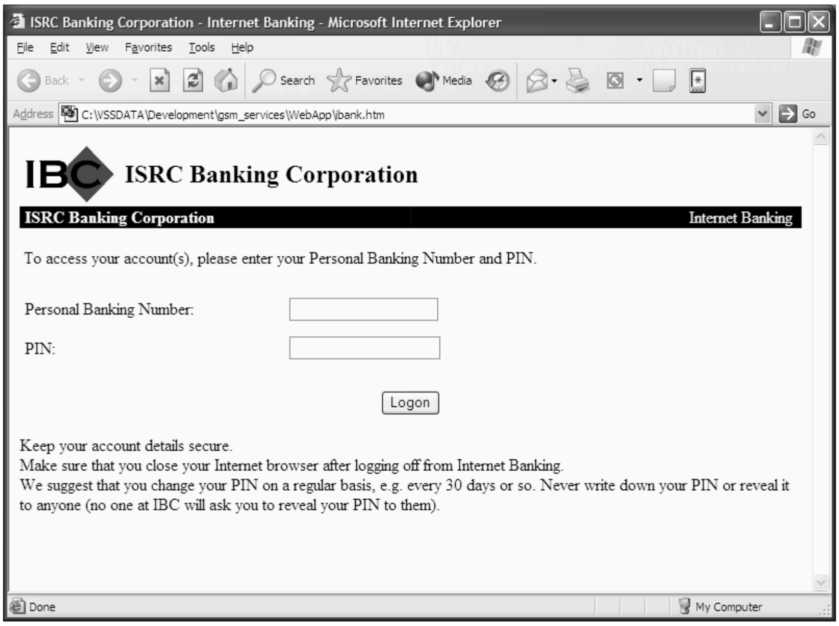
Once submitted, the authentication server validates that the phone is associated with the web session, completing the association protocol. The authentication server then proceeds to send a request for the user's location to the Gateway Mobile Location Center (GMLC). Using the methods detailed in § 4.7 and § 4.8, the authentication server determines if the user's location is within the registered location the user selected. Successful validation, authorized the user to access various Internet banking functionality.

Various granularities of access control can be specified, such that a user successfully associated and within a registered location area may be granted access to electronic international bank drafts, where as a user who was only able to authenticate using their unique user identity and PIN, is granted read only access to account balances. The location information can also be used for audit purposes, binding a transaction to a location.

A hypothetical WAP version of this service is illustrated in Figure 4.18. This was not implemented due to the lack of cellular phone support for WAP version 2.0 at the time.

4.12 Summary

This chapter has presented two methods of facilitating tamper-resistance of signaling for GSM location acquisition. Experimentation in testing the tamper-resistance of the timing advance was detailed and practical use of the timing advance in a location system was discussed. Recommendations for the representation of geographical data for trusted location systems were detailed and applied to an example auditing and access control system. In order to avoid disassociation attacks against the location systems, a cellular phone association protocol was proposed. A public key version of this protocol for WAP 2.0 enables cell phones was also proposed and demonstrated in terms of a web and WAP-based Internet banking service.


ISRC Banking Corporation - Internet Banking - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
Go
Logon
Done
My Computer

Address: C:\VSSDATA\Development\gsm_services\WebApp\ibank.htm

IBC ISRC Banking Corporation

ISRC Banking Corporation Internet Banking

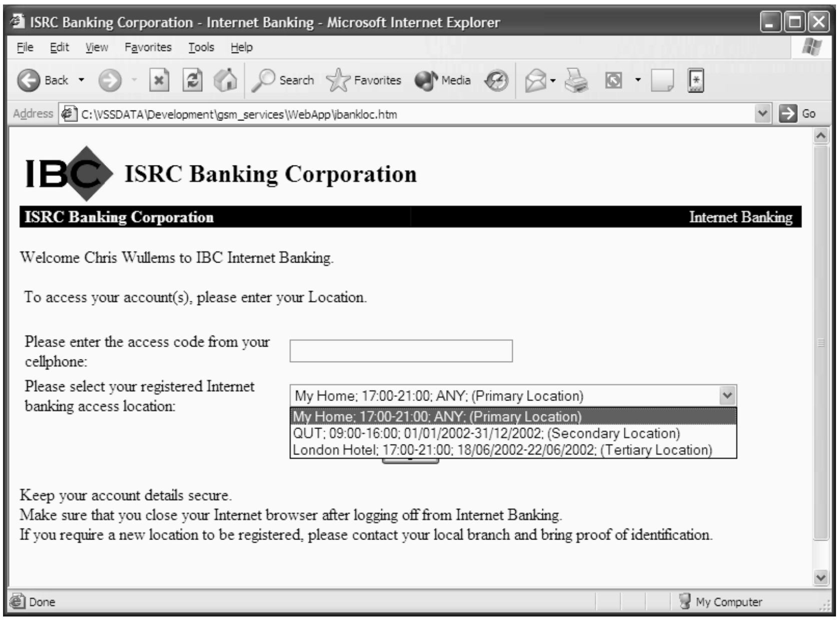
To access your account(s), please enter your Personal Banking Number and PIN.

Personal Banking Number:

PIN:

Keep your account details secure.
Make sure that you close your Internet browser after logging off from Internet Banking.
We suggest that you change your PIN on a regular basis, e.g. every 30 days or so. Never write down your PIN or reveal it to anyone (no one at IBC will ask you to reveal your PIN to them).

Figure 4.16: IBC Prototype Login


ISRC Banking Corporation - Internet Banking - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
Go
Done
My Computer

Address: C:\VSSDATA\Development\gsm_services\WebApp\ibankloc.htm

IBC ISRC Banking Corporation

ISRC Banking Corporation Internet Banking

Welcome Chris Wullems to IBC Internet Banking.

To access your account(s), please enter your Location.

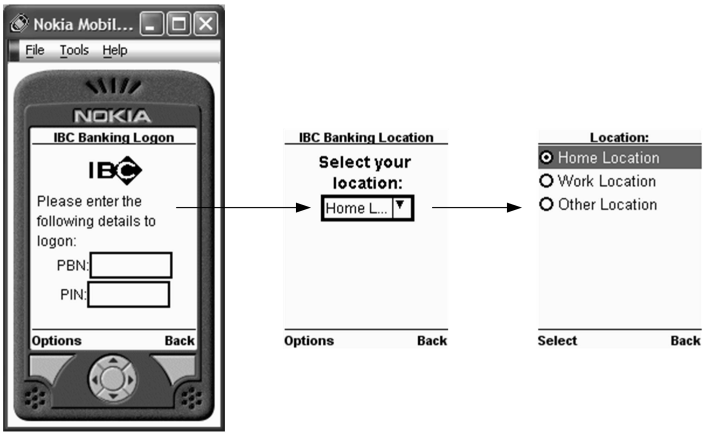
Please enter the access code from your cellphone:

Please select your registered Internet banking access location: My Home; 17:00-21:00; ANY; (Primary Location)

My Home; 17:00-21:00; ANY; (Primary Location)
QUT; 09:00-16:00; 01/01/2002-31/12/2002; (Secondary Location)
London Hotel; 17:00-21:00; 18/06/2002-22/06/2002; (Tertiary Location)

Keep your account details secure.
Make sure that you close your Internet browser after logging off from Internet Banking.
If you require a new location to be registered, please contact your local branch and bring proof of identification.

Figure 4.17: IBC Prototype Location Dialog



IBC Banking Logon

IBC

Please enter the following details to logon:

PBN:

PIN:

IBC Banking Location

Select your location:

Home L...

Location:

● Home Location

○ Work Location

○ Other Location

Figure 4.18: IBC WAP Interface

Chapter 5

Proximity-based Network Packet Filtering for IEEE 802.11 Wireless Devices

5.1 Introduction

In this chapter we investigate the use of context-awareness in access control processes at the IP layer. Location context-awareness in IP-layer access control for IEEE 802.11 Wireless LANs was investigated as wireless LANs have become increasingly popular in both home and office environments, and there is an increasing need to provide adequate security without unduly restricting functionality, especially for applications such as public access points. This chapter introduces a novel proximity-based packet filtering system which can be integrated into firewall/router devices, augmenting existing security mechanisms such as virtual private networks (VPNs), etc.

Proximity-based packet filtering is particularly useful in restricting the usage of a wireless network to unauthorized users. An example environment is a coffee shop or restaurant that provides wireless Internet access to its customers. With the use of proximity-based packet filtering, it is possible to restrict business and residential users who would have otherwise been in range of the wireless LAN from accessing the network resources without being in the coffee shop. The packet filtering system would require that users are within a predefined location area.

As WEP encryption is easily defeated by publicly available tools, the augmentation of proximity information into a packet filtering system increases the complexity of such an attack, requiring both software and hardware attacks.

This chapter details the proposed method of obtaining wireless LAN signal data from the monitor point, facilitating transparent location proximity acquisition using standard 802.11 equipment. The pathloss model used for calculating distances from signal data is detailed with a performance analysis illustrating the accuracy of the pathloss model in an indoor office environment (§ 5.4). The system architecture is presented in § 5.5, illustrating the augmentation of proximity-based information into a packet filtering system. Benchmark results rating the performance of the implemented prototype are also presented.

Portions of this chapter have been published in the paper: Proximity-based Network Packet Filtering for IEEE 802.11 Wireless Devices. In Proceeding of IASIS International Conference on Applied Computing, Lisbon, Portugal, March 2004.

5.2 Emerging Work in 802.11 Location and Ubiquitous Network Security

The primary contribution in this chapter is the integration of wireless LAN proximity context data into network security services such as firewall packet filtering. While there does not appear to be any previous work in ubiquitous network security services, there are three significant contributions to location acquisition in IEEE 802.11 wireless LAN. Bahl and Padmanabhan in [7] propose a location system that combines empirical measurements with signal propagation modeling to derive the location of a wireless device. Their proposal involves two approaches, an empirical method using a database of signal strength data, and the use of a signal propagation model that takes into account the attenuation of walls. The empirical method yielded better results using the Nearest Neighbor(s) in Signal Space algorithm (NNSS), however required an offline dataset.

This method is extended by Saha et al. in [85], using more sophisticated classifiers which take into account the distribution of data; Nearest Neighbor Classifier, Back-Propagation Neural Network, and Classification by Histogram Matching. This resulted in more accurate location results.

Kishan et al. in [67] proposed a location acquisition that utilized signal strengths from three access points, and performed a triangulation, based on laterations, to determine the location of a client. The proposed system measured the distance of the wireless device from each access point and performed a calculation to determine the circular area of influence. The location of the device is then approximated to be the intersection of these areas.

The determination of user location in a wireless LAN network for use in context-aware applications was explored by Jason et al. [89], and Asim et al. in [88]. It was identified that the use of location within a wireless LAN for context-aware applications offers almost limitless usage. Both papers explore the different schemes for obtaining user location for a context aware Portable Help Desk application. This context aware system allows for its user to determine the location of other users on campus as well as information about them.

Our approach significantly differs from these contributions in that we obtain the signal data of transmitting client devices from the infrastructure, where as these proposals observe the signals of access points from client devices. This is considerably better in terms of security and usage for access control, as the location server does not have to trust the client device to provide their location. This is then augmented into a packet filtering system for dropping access of devices that are not within the bounds of the wireless LAN. Our work explores the use of location determination for implementing a packet filtering firewall.

5.3 Enhancing Linux Packet-filtering Firewalls

A firewall serves as a single point of entry into a network, where traffic can be monitored and/or filtered. Only authorized requests are be allowed into the network, where such access control policies are enforced by the firewall.

There are two dominant types of firewalls: application proxies and packet filtering gateways. The common belief is that proxies are more secure; however, due to their nature of being restrictive and limited in performance, proxies have generally been adopted for out-bound traffic, rather than in-bound. The advantages of using packet filters, or even more sophisticated stateful packet filters, stems from the fact that they offer high-performance for in-bound traffic. For this reason, a packet filtering firewall was chosen for implementation in this system.

Having a packet filtering firewall allows finer granularity in control over communication packets. Incorporating a packet filter allows a firewall to determine if a wireless device has access to network resources. Extending this idea to allow user location information to determine the access control furthers the idea of a context aware firewall. The augmentation of context aware information to derive firewall rules allows for a dynamically changing access control list. Developing a context aware firewall system requires information such as user location and also the logic to update the access control list dynamically. The combination of netfilter and IP Tables, found in the Linux kernel, was chosen for its simplicity in design and also extensibility, thereby allowing the development of the proposed system.

The following subsections discuss the components and techniques used for location-aware packet filtering.

5.3.1 Netfilter

Netfilter is a framework for packet manipulation, found in the Linux kernel from version 2.3.xx development kernel and is still part of the overall firewall kernel design in the test 2.6.xx Linux kernels. The framework allows the mangling of networking packets outside of the Berkley Socket Interface [84]. IP Tables is an array of rules in memory that provides information to where packets from each hook should begin to traverse. Upon the registration of the table, userspace can read and replace contents within the table using the setsockopt() and getsockopt() functions. The combined use of netfilter and IP Tables provides the full firewalling capabilities of the Linux operating system, which easily facilitates future extensions.

5.3.1.1 Netfilter Framework

Each protocol defines “hooks” within the protocol stack, which the packets would traverse. At each of these “hook” points; the protocol calls the netfilter framework with the packet and hook number. Parts of the kernel can register to listen to these “hook” points, for each protocol. When a packet is passed to the netfilter framework, it will check if any kernel modules are registered for the protocol and the associated “hook”. At this point, the kernel module can examine the packet and decide on 1 of 3 possible actions to take:

1. Allow it to pass (*NF_ACCEPT*);

2. Forget the packet (*NF_STOLEN*); or

3. Queue the packet for userspace (*NF_QUEUE*).

The queued packets are then collected by the *IP_QUEUE* driver and sent up to user space for further processing. Figure 5.1 illustrates the 5 defined hooks for IPv4. There are 2 points of entry into the protocol stack; either a transmission packet arriving from the network or a packet created by the local system.

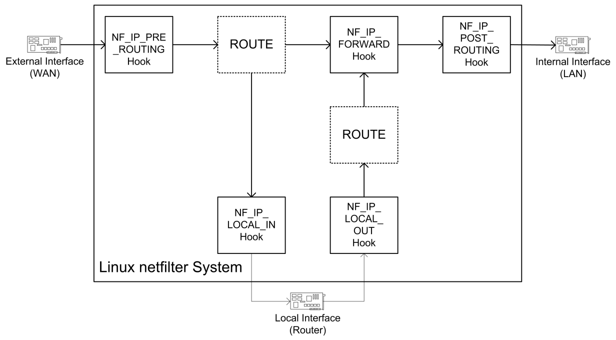


Figure 5.1: Linux Netfilter hooks

A packet arriving on the system via the network would first pass through a sanity check, before being passed to the netfilter framework's *NF_IP_PRE_ROUTING* hook, after which, the packet is passed through the local routing process. If the packet passes through without being filtered by the routing process (i.e. Not filtered because it is not possible to route), it will be passed on to 2 possible hooks:

1. If the packet was destined for the local system, it will be passed to the *NF_IP_LOCAL_IN* before it is passed to the local processes;

2. If the packet was destined for another interface, it will be passed from the local routing process to *NF_IP_POST_ROUTING* hook before it is transmitted down the line.

The *NF_IP_LOCAL_OUT* hook is called for packets that have been created on the local system. The packet is then passed through the local routing process before being placed on the protocol stack.

5.4 Proximity Measurements in 802.11 Wireless LAN

Calculating the location requires at minimum three monitor points and the use of a lateration calculations to determine the location of a device based on signal pathloss or signal propagation timing. Because of the inability of 802.11 devices to accurately measure the timing of signals in indoor environments, measurements are made using signal pathloss propagation models. There are three general types of propagation models: [75]

1. *Empirical propagation models.* These models are statistical, based on experimental data and are simple, efficient and particularly suitable for computing implementation, however are considerably less accurate than the other two models. There is no requirement for an environmental database, however, the propagation environment coefficient is very sensitive.

2. *Theoretical propagation models.* These models include ray-tracing models and Finite-Difference Time Domain (FDTD) models. Ray-tracing models are based on calculating all possible signal paths from the transmitter to the receiver; however they require extensive computational resources as well as extensive knowledge of the environment. FDTD models are based on numerical solution to Maxwell's equations of electromagnetic wave propagation. Similarly to ray-tracing, FDTD models require extensive computational resources and knowledge of the environment.

3. *Artificial neural network propagation models.* These models are based on neural networks with significantly better accuracy than empirical models, without the computational complexity of theoretical propagation models. These models require extensive training processes.

While theoretical and artificial neural network propagation models are more accurate than empirical models, an empirical propagation model was chosen because of its speed and minimal computational requirements, being particularly appropriate to packet filtering. Simplicity of implementation and no requirement for a database storing environment data were also significant factors in the choice of propagation model.

The following sections detail the methods used to obtain signal strength data from 802.11 devices and the pathloss model used to calculate proximity based on these measurements.

5.4.1 Signal Strength Acquisition in 802.11 Wireless LANs

Wireless LANs measure the received signal strength in a channel, as it is inherent to its medium access control operations. 802.11 devices support both Frequency Hopping Spread Spectrum (FHSS) and Direct Sequence Spread Spectrum (DSSS) at the physical layer for the transmission of data. Typically DSSS is used, as it is able to achieve higher data rates. Unlike CDMA which utilizes a set of spreading codes, the spreading architecture of 802.11 utilizes the same PN code for all 802.11 devices, resulting in a shared channel.

As wireless LAN channels are shared, a medium access control mechanism is used to share the channel with devices that are using the channel. The 802.11 standards (IEEE, 1999) define the Carrier Sense Multiple Access / Collision Avoidance (CSMA/CA) medium access control (MAC) protocol for the Distributed Coordination Function (DCF). This protocol operates with the physical layer which obtains the Received Signal Strength Indicator (RSSI), indicating the strength of a received signal. The RSSI is used by the Clear Channel Assessment (CCA) algorithm to determine when the channel is clear for transmission, based on a minimum threshold. As detailed in (IEEE, 1999), the RSSI measurement is a measure of observed energy by the physical sublayer, intended for use in a relative manner. Absolute accuracy of the RSSI reading is not specified.

The RSSI is represented as a 1 byte value, allowing for 256 different signal levels. How these levels are mapped to actual dBm values is dependent on a vendor's implementation. *RSSI_Max* is the specified maximum RSSI value a given vendor has chosen to implement. As such, RF energy is arbitrarily represented as integer values from 0 to *RSSI_Max*. The range of supported values typically spans -10 dBm to -113 dBm represented as 100 levels in common 802.11 devices.

A number of chipsets including the Prism II chipset support monitor-mode functionality. This allows the card to sniff all packets in a given channel. The RSSI value for each received packet can be extracted from the driver and used as part of a location calculation.

The wlan_ng[1] Linux device driver with a Prism II based 802.11 card was used for extraction of RSSI values. The driver is coupled with a utility that is able to initiate monitor mode. Open source wireless LAN tools are used to sniff packets from a specified channel. The tools are able to provide the RSSI of the given packet as a percentage of *RSSI_Max*. This value must be converted to dBm for use in a pathloss model.

As the RSSI is implemented differently by each vendor, it is necessary to find the conversion functions for the particular 802.11 device being used. A Prism II-based card was used for our experimentation. For Prism II-based cards, the signal power in dBm is linearly related to the RSSI. An RSSI-to-dBm constant is retrieved from the device via the driver by reading RID *(Resource Identifier): FC46, cnfDbmAdjust*. This can be used to determine the dBm from the RSSI by subtracting the value of *cnfDbmAdjust* from the RSSI value.

[1] Refer to The linux-wlan™ Project, `http://www.linux-wlan.org/`

5.4.2 Path Loss Model

In order to calculate the proximity in terms of meters, it is necessary to use a pathloss model which models the logarithmic decay of RF energy over a given distance. The pathloss model used in the architecture is given below:

$PL = 20log(a) + 10nlog(d) - G1 - G2 - 27.6$

PL is path loss (dB) = $txPwr(dBm) - rxPwr(dBm)$ a is the frequency of the channel (MHz); d is the distance (meters); $G1$ is transmitter antenna gain (dBi); $G2$ is receiver antenna gain (dBi); n is the environment coefficient that quantifies the environment; and 27.6 is a constant used to calculate distance in terms of meters.

The environmental coefficient n is derived through the sampling of pathloss $(txPwr - rxPwr)$ at known distances. By using known distances, it is possible to calculate n as an average of the environmental coefficients for each known distance.

In our office test environment, the environmental coefficient was calculated to be 2.37. This was based on samples taken in the office and is illustrated in Figure 5.2. This figure details the observed pathloss, the logarithmic curve based on the sample data and the logarithmic curve based on the pathloss model for the distances at which the samples were taken. The measurements were taken using a PRISM-based wireless LAN card using the methods detailed in § 5.4.1.

Figure 5.3 illustrates the distances calculated using the pathloss model and the observed pathloss versus the actual distances. The selected model is computationally fast and satisfactorily accurate for distances up to 13 meters for indoor environments. It begins to loose accuracy fairly quickly there after. For outdoor environments, the range is typically better due to line of sight conditions. The speed of the calculations is critical, given that the access control system must be able to react quickly to proximity changes.

Figure 5.4, which illustrates the office test environment, details the locations trilaterated from the above pathloss model and the observed pathloss from three monitor points versus the actual locations. Use of more sophisticated indoor propagation models that take into account obstructions such as walls, would yield a significantly better result. For proof of concept purposes, this location determination mechanism provided reasonably accurate location results. The access control architecture is detailed in the following sections.

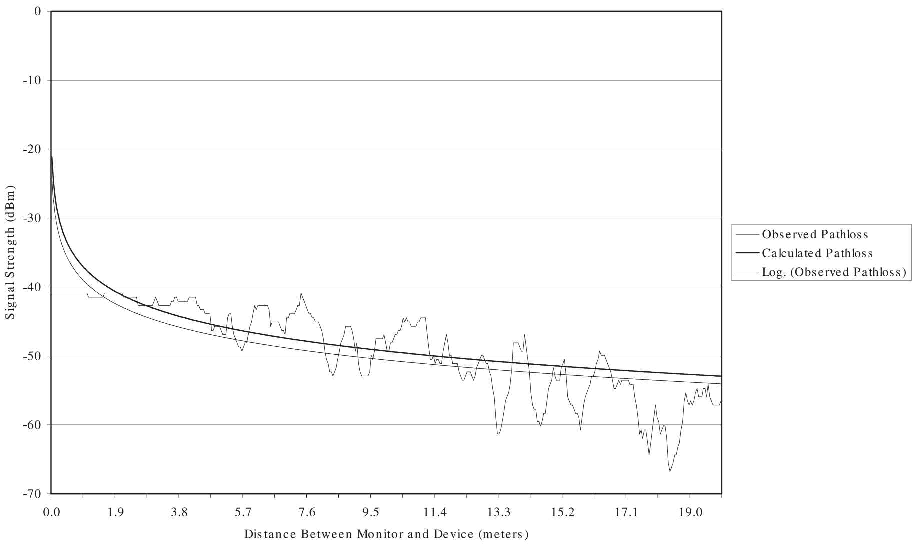


Figure 5.2: Signal Decay Over 20m in an Indoor Office Environment

5.5 System Architecture

The architecture is composed of a number of components as illustrated in Figure 5.5. The following subsections detail these architecture components.

5.5.1 Wireless LAN Monitor

Each wireless LAN monitor is used in the measuring of the signal strength of all devices participating in the wireless LAN network and is based on the RSSI value of each packet transmitted into the air.

Firstly, a monitor uses a wireless LAN card in promiscuous mode, allowing it to sniff all transmissions in the channel. This allows the monitor to extract the MAC address of each packet, and also read of the RSSI value, which is calculated as a percentage of *RSSI_Max*. This is formulated into a message, which is communicated from the wireless LAN monitor, to the context service with the following format:

```
<WLAN_MONITOR_ID> <INTERFACE_MAC_ADDRESS> <PERCENTAGE_OF_RSSI_VALUE>
```

Each monitor facilitates transparent signal strength acquisition without knowledge of client devices. As such, a client device does not require any specialized software, and any 802.11 device can be located using this scheme.

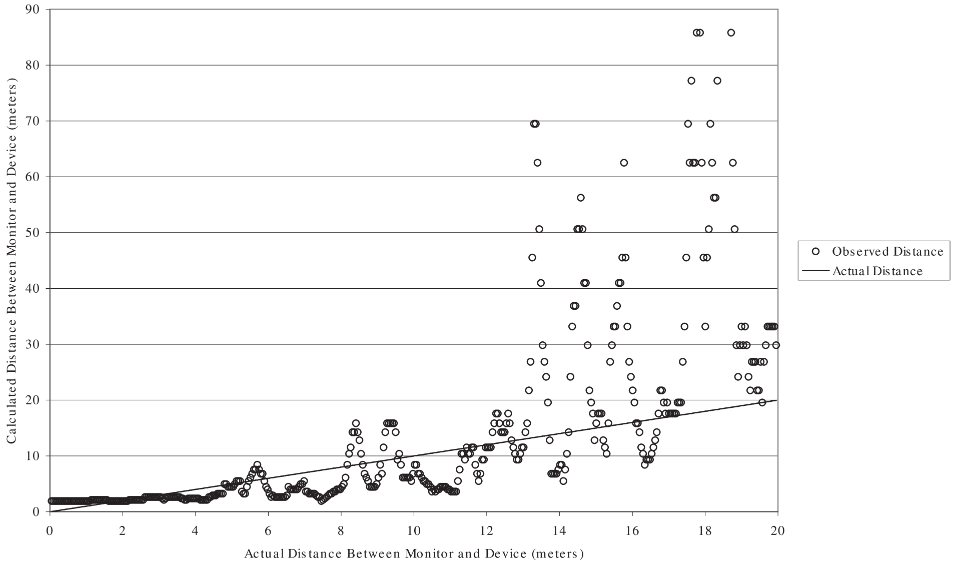


Figure 5.3: Actual Distances vs Pathloss Calculated Distances

5.5.2 Wireless LAN Firewall / Router

The Wireless LAN firewall / router consist of the Context Service (CS) and the Netfilter kernel module. The system has been designed such that the proximity calculation and access control processing is done independently of the kernel Netfilter packet processing. Orthogonal processing allows packets to flow without delay, as long as there is an entry in the CS access control list.

The CS supports three or more monitors, each of which must have a unique ID. While a greater number of monitors will result in a more accurate location, the latency for granting or denying access will increase, as the lateration processing becomes increasingly more time consuming[2].

Each monitor provides its ID, the observed device's MAC address and corresponding RSSI. On receipt of an update from a monitor, a cache containing the observed device's MAC address is updated for the corresponding monitor ID, with the observed RSSI. If there are not at least three fresh non-zero RSSI values, the location calculation will not be performed. Freshness is determined from the timestamp for a given Monitor's observed RSSI value, and a threshold defining the maximum age an RSSI value will be accepted.

Assuming these conditions are met, a set of fresh non-zero RSSI values are con-

[2] Exact figures are not provided as additional wireless equipment was not available to quantify the increase in latency.

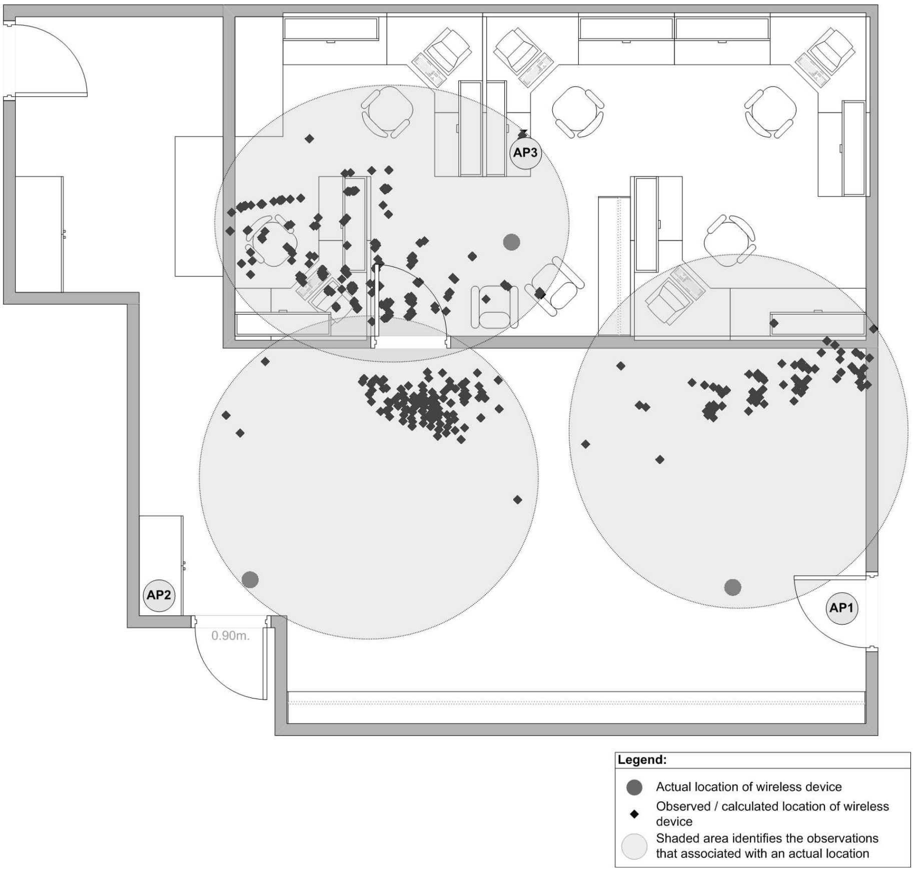

AP3
AP2
AP1
0.90m.
Legend:
● Actual location of wireless device
◆ Observed / calculated location of wireless device
○ Shaded area identifies the observations that associated with an actual location

Figure 5.4: Actual location vs location calculated from pathloss

verted to dBm (as detailed in § 5.4.1) and are laterated. Lateration involves finding the points of intersection of circles defined by the known locations of the monitor points, and whose radius is the distance derived from the pathloss model (as detailed in § 5.4.2). The result of this process is a single point which is forwarded to the Access Control Logic.

The Access Control Logic is responsible for determining whether a given point is within a specified location area. Access control policy is configured using polygons that bound the area in which client devices are granted or denied access. Polygons must have at minimum 3 points, and at maximum 15, such that the last point is connected to the first. Lines connecting the points are not to cross another. The access control logic determines if the location of the client, given as a point, is within the polygon using the Jordan Curve Theorem (Refer to Appendix E).

If access to a given client device is granted, it is added to an access control list, identified by its MAC address. The entry contains a timestamp, such that freshness of

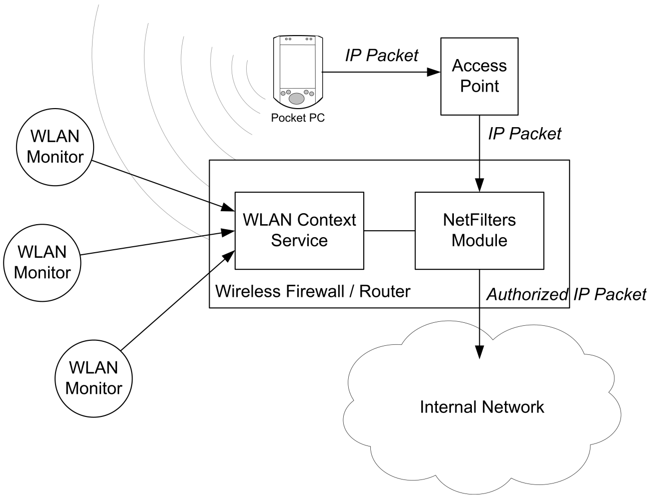


Figure 5.5: System Architecture

an access control decision is known. The Access Control Logic also supports explicit denials in its access control list.

When a packet transmitted by a given client device enters the Netfilter system, an access control decision is requested by the Netfilter kernel module for a given packet received through the Netfilter *NF_IP_PRE_ROUTING* hook (Figure 5.6). The source IP address of the incoming packet is sent to the CS in an access control request, for which a Boolean response is received resulting in the return of *NF_ACCEPT* or *NF_DROP* to the Netfilter subsystem. On receipt of an access control request, the Access Control Logic performs an ARP (Address Resolution Protocol) lookup to resolve the MAC address for the IP address detailed in the access control request. The Access Control Logic determines whether access is granted or denied by performing a lookup in the access control list, and verifying the freshness of the entry corresponding to the given MAC address. If an entry does not exist, access is denied by default.

5.6 Prototype Implementation

The architecture was implemented in a controlled environment for testing purposes. The setup included the following systems and equipment (Figure 5.8):

- 3 Wireless LAN Monitors: Pentium 4 class notebooks running Fedora Core 1 Linux. Each system utilizes a IEEE802.11b Wireless LAN Adapter with the PRISM2 chipset;

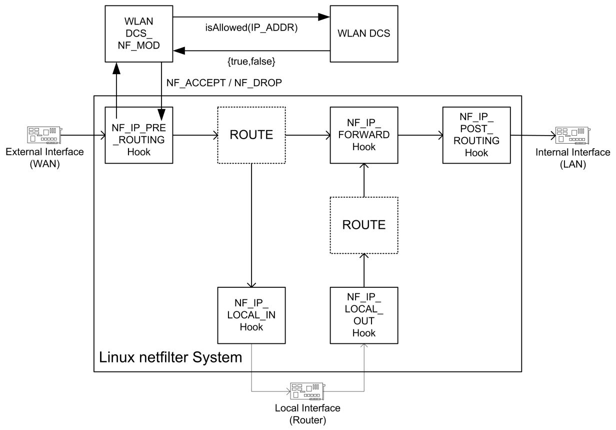


Figure 5.6: Context-aware Netfilter Module

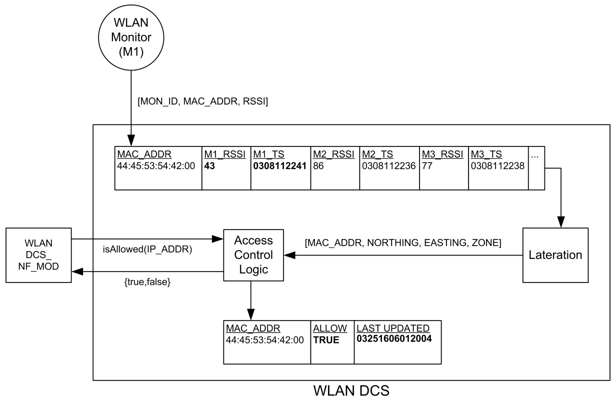


Figure 5.7: WLAN Dynamic Context Service

- WLAN Firewall/Router: Pentium 4 class desktop PC running Fedora Core 1 Linux. This system performed the packet filtering functionality and hosted the context service;

- ORiNOCO AP500, Wireless Access Point with an extended range antenna (2.5dBi); and

- NetGear DS108 Dual Speed Hub (10/100Mbps).

Timing functions were added to the kernel module and the context service to benchmark the performance of the packet filtering. The performance results are detailed in Table 5.1.

The effect of the access control delay is that the first few packets transmitted by a client device are dropped by the kernel module. This should not cause application

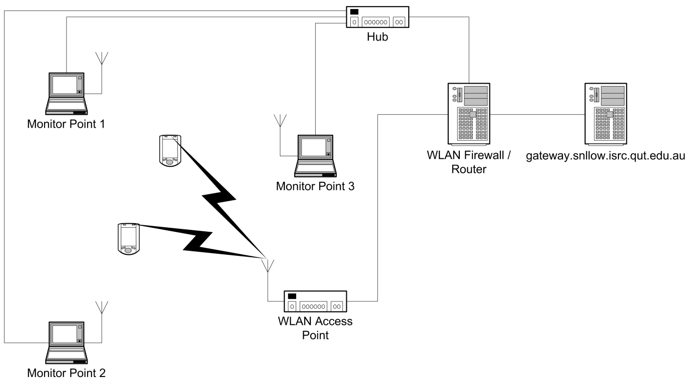


Figure 5.8: Network Diagram of Prototype Implementation

Process	**Average Time Per Packet (ms)**
Lateration	16.54
Packet sniffing and communication of RSSI for packet source MAC address to WLAN Context Service	61.38
Total packet delay	77.92

Table 5.1: Packet Filtering Access Control Delay

disruption, as TCP-based packets should be retransmitted according to TCP protocol functionality. After the delay in observing and calculating the location, assuming access is granted, packets are routed to the destination. If a client is to move outside an authorized area, this delay will also allow a small number of packets to be sent or received before access is denied.

5.7 Summary

In this chapter, we have proposed a novel proximity-based packet filtering architecture that uses the contextual information of a user's location. A method of trilateration was used for location acquisition, providing satisfactory results that demonstrate the proof of concept using extensions of the Linux netfilter architecture. Performance results of the signal propagation pathloss models utilized and the performance of the implemented prototype have been detailed. Improvements in the location performance can be made through the use more complex pathloss models or other location acquisition techniques identified in § 5.2.

Chapter 6

A Wireless LAN Denial of Service Attack

6.1 Introduction

Wireless LAN Location systems are typically used in Wireless LAN management, tracing of hackers, intrusion detection and can be used for context-aware applications. During the development of Wireless LAN location services, a vulnerability realizing a trivial denial of service attack was discovered. In this chapter, we discuss the attack and its ramifications.

Wireless LAN has been standardized by the IEEE in the 802.11 series of standards, detailing the requirements for Physical (PHY) and Medium Access Control (MAC) functionality. These standards have been widely accepted for supplementing wired infrastructure. It has become increasingly evident that the success and cost-effectiveness of commercial-off-the-shelf wireless LAN has lead to its widespread use in industries outside the traditional enterprise sector.

Sectors such as the industrial and transportation have more stringent requirements than enterprise in their requirements for availability, robustness, survivability. While the use of wireless LAN technology is significantly more cost effective, it was not designed to meet the requirements of sectors that form part of the national critical infrastructure.

This Chapter will introduce a trivial attack on the availability of direct-sequence spread spectrum Wireless LANs, demonstrating that it is not a suitable technology for safety-critical or critical-infrastructure applications.

6.1.1 Emerging IEEE 802.11 Applications

The specification by the IEEE of the 802.11 wireless LAN standards has led to the emergence of a wide range of affordable, and to a large extent interoperable, technology becoming available for wireless communications. The availability of affordable wireless communications technology has not evaded the attention of a number of sectors that have traditionally relied on expensive, proprietary radio technologies to meet their communications needs. Such sectors include transportation, process control, and telecommunications.

For example Alcatel, a large provider of control technologies to the transport sector, is including 802.11 based radios in their technology strategy [101] and is reportedly using COTS based IEEE 802.11 radios for trains in Las Vegas, Hong Kong, and Korea[1]. There also appears to be interest in the industrial networking arena for using commodity 802.11 data radios in a range of applications[2]; consideration of WLAN technology for use by the military[3]; public safety[4]; and the interworking of WLAN and 3G systems[5].

This emerging interest in adopting IEEE 802.11 technology for environments that have stringent performance and security requirements, especially in safety-critical control environments, is very concerning in lieu of the ease with which such communications can be disrupted, as described in § 6.4. While attacks on the confidentiality and integrity of WLAN communications can be expected to be resolved in current and future enhancements to the standards, attacks on availability as described in this chapter may not be so easily solved.

The structure of the chapter is as follows. First existing confidentiality, integrity, and availability attacks against IEEE 802.11 based WLANs are summarized in § 6.2. Aspects of the MAC and PHY layer protocols relevant to the newly described attack are reviewed in § 6.3 and the new attack is described in § 6.4. The new attack is analyzed in Section 6.5 and possible solutions to existing and this new attack are discussed in § 6.6.

[1] Refer to `http://www.tsd.org/cbtc/projects/`

[2] Refer to `http://ethernet.industrial-networking.com/wireless.htm`

[3] Refer to Joint Tactical Radio System `http://jitc.fhu.disa.mil/jtrs/`

[4] Refer to Advanced Network Technologies Division Communication & Networking Technologies for Public Safety `http://w3.antd.nist.gov/comm\_net\_ps.shtml`

[5] Refer to IEEE Communications, Volume: 41, Issue: 11, Nov. 2003 for articles on the integration of wireless LAN and 3G wireless.

Portions of this chapter have been published in the paper: A Trivial Denial of Service Attack on IEEE 802.11 Direct Sequence Spread Spectrum Wireless LANs. In the Third IEEE Wireless Telecommunications Symposium (WTS 2004), Pomona CA, USA, May 2004.

6.2 Existing Attacks Against IEEE 802.11

Since becoming standardized, a number of attacks, both theoretical and practical have been described against IEEE 802.11 networks. The attacks either focus on the confidentiality and integrity of wireless communications, or the availability of the wireless networking infrastructure.

Significant attacks against the security services provided by the Wired Equivalent Privacy (WEP) protocols are well documented [13, 53, 93].

A range of availability attacks, mainly directed at the management and MAC protocols used by IEEE 802.11 WLANs have also been identified [9, 51]. These attacks generally exploit the fact that many of the management messages in IEEE 802.11 are unauthenticated, rather they rely on correctly behaving MAC layer implementations. Some example attacks involve:

- Identity spoofing to permit deauthentication or disassociation of the victim node from the network (See Figure 6.1[6]);
- Exploiting power saving features; and
- Exploiting media access protocols through such activities as modifying backoff timers, and keeping network allocation vectors (NAV) at non zero values.

While each of these attacks is significant to varying degrees, it is our opinion that these attacks are most likely mitigated through the application of cryptographic techniques such as the authentication of messages[7].

6.3 Review of IEEE 802.11 Protocols

The IEEE 802.11 working group standards detail information about the physical layer (PHY) and Medium Access Control layer (MAC) protocols required for wireless lo-

[6] Image sourced from [61].

[7] IEEE 802.11i will address the security issues associated with WEP, but will not currently address some of the availability attacks described.

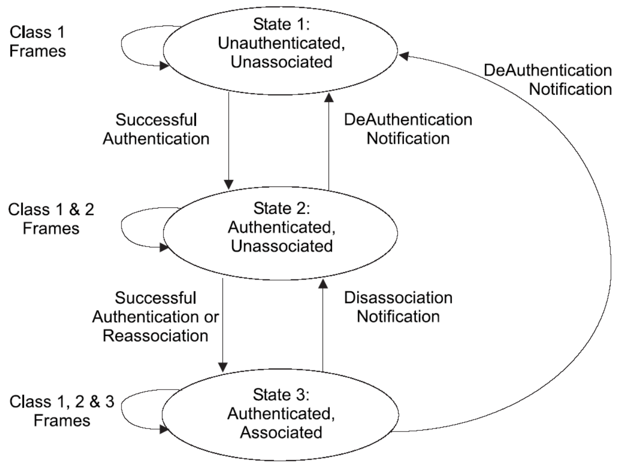


Figure 6.1: Authentication and Association States

cal area networking (WLAN). The PHY layer is further divided into the Physical Layer Convergence Procedure (PLCP) sublayer[8] and the Physical Medium Dependent (PMD) sublayer as shown in Figure 6.2[9].

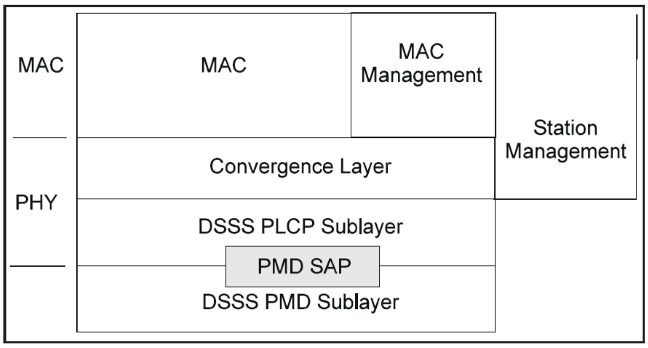


Figure 6.2: PHY / MAC Layers

The PHY layer provides the MAC layer with information about the availability of the underlying medium (carrier sense functions) and is involved with the reception and transmission of data.

To provide distributed, but coordinated access to the shared wireless medium, the IEEE 802.11 MAC protocols perform Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). This strategy minimizes the likelihood of two stations transmitting simultaneously, resulting in a collision and subsequent corruption of data,

[8] The use of this sublayer ensures that the MAC layer is not tightly coupled to a specific PMD

[9] Image sourced from [61].

while ensuring that the available bandwidth is effectively utilized. Fundamental to the functioning of CSMA/CA in the IEEE 802.11 MAC protocols is the Clear Channel Assessment (CCA) procedure performed by the PHY layer.

6.3.1 Clear Channel Assessment

The Clear Channel Assessment (CCA) is used by the MAC layer to determine (1) if the channel is clear for transmitting data, and (2) for determining when there is incoming data.

Evaluation of CCA is made by the PHY layer and the resulting assessment is communicated to the MAC layer via the *PHY-CCA.indicate* service primitive. This primitive can either be set to IDLE, when the channel is assessed to be clear, or BUSY when the channel is assessed to be in use[10].

The IEEE 802.11 series of standards define the following Clear Channel Assessment modes:

- *CCA Mode 1.* Energy above threshold. CCA shall report a busy medium upon detection of any energy above the ED threshold.

- *CCA Mode 2.* Carrier sense only. CCA shall report a busy medium only upon detection of a DSSS signal. This signal may be above or below the ED threshold.

- *CCA Mode 3.* Carrier sense with energy above threshold. CCA shall report a busy medium upon detection of a DSSS signal with energy above the ED threshold.

- *CCA Mode 4.* Carrier sense with timer. CCA shall start a timer whose duration is 3.65ms and report a busy medium only upon the detection of a high rate PHY signal. CCA shall report IDLE medium after the timer expires and no high rate PHY signal is detected. The 3.65ms timeout is the longest duration possible for a 5.5Mbit/s PSDU.

- *CCA Mode 5.* A combination of carrier sense and energy above threshold. CCA shall report busy at least while a high rate PPDU with energy above the ED threshold is being received at the antenna.

[10] The specific details of how this is achieved differ with each physical layer, but the process involves the detection of energy beyond some threshold (PMD_ED.indicate), or the acquiring of a valid code lock (PMD_CS.indicate)

6.3.2 Medium Access Control

The IEEE 802.11 standards specify both a centralized and distributed coordination function for controlling access to the shared transmission medium in WLANs operating in infrastructure mode (BSS / ESS). These are known as the Point Coordination Function (PCF) and the Distributed Coordination Function (DCF) respectively. The DCF is used to permit the ordered sharing of network resources in ad hoc, or IBSS based WLANs.

The PCF is centrally controlled and operates by a master (the access point) polling slaves (the stations) for any data they may have to transmit. This permits contention free access to the transmission medium and operates in a *speak only when spoken to* paradigm. The PCF mode of operation is not widely deployed, so is not discussed further.

The DCF as its name suggests provides distributed, but coordinated access to the shared medium. In coordinating access to shared medium there are two goals: (1) permit stations to commence transmission with minimal delay; and (2) prevent two or more stations from transmitting simultaneously to avoid collisions and data corruption. In order to meet these goals the 802.11 standard specifies a *listen before talking* paradigm for communications medium access, such that a station can only transmit while no other station is transmitting. This achieved by having the station perform a carrier sense operation, in the form of a clear channel assessment (CCA) prior to transmitting any data. To minimize the likelihood of two stations sensing the medium idle simultaneously, the medium must be idle for a period of time known as an inter frame space (IFS)[11] before transmission. Once the medium has been idle for at least the appropriate IFS time, a random backoff procedure is performed - further reducing the likelihood of two stations transmitting simultaneously (Figure 6.3[12]). The random backoff time is calculated as a function of the number of unsuccessful attempts to transmit an MPDU, such that as the offered load to the network increases, the larger the range of possible backoff time values becomes.

An additional IFS value is used by the DCF when frame transmission errors are detected. This value is known as the Extended IFS (EIFS) and it is the longest IFS value of all.

[11] These values are PHY specific and the use of different IFS values allows simplistic prioritization of traffic to be performed.

[12] Image sourced from [61].

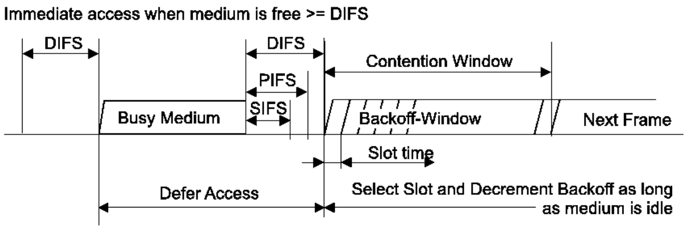

Immediate access when medium is free >= DIFS
DIFS
DIFS
PIFS
SIFS
Contention Window
Busy Medium
Backoff-Window
Next Frame
Slot time
Defer Access
Select Slot and Decrement Backoff as long as medium is idle

Figure 6.3: Inter Frame Spacing Relationships

6.3.3 Summary

The important points to note from this section, are that the CCA is used by the MAC layer to determine if the transmission medium is available for data transmission. Whenever the MAC layer receives a PHY-CCA.indicate(IDLE) it will wait for an IFS plus a random backoff time prior to transmitting. When the MAC layer receives a PHY-CCA.indicate(BUSY) service primitive from the PLCP layer it will defer from accessing the medium. The PHY layer uses energy detection, code detection, or some combination of the two to determine if the medium is busy and sets PHY-CCA.indicate accordingly.

6.4 A New Attack on IEEE 802.11

Here we propose a new attack on the IEEE 802.11 WLAN MAC protocol. The attack described permits a low powered, portable device such as a Compaq IPAQ, using a commonly available wireless networking card to disrupt wireless network communications over a significant range, for a significant period of time, in a manner that makes the identification and localization of the attacking node non-trivial.

The attack exploits the Clear Channel Assessment (CCA) procedure used by all standards compliant hardware and causes all stations within range, both clients and access points, to permanently defer transmission of data. This is achieved by stimulating the CCA in a manner that the channel is always assessed to be busy, thus preventing the transmission of any data over the wireless network.

Significant concerns are raised by this attack for a number of reasons, including:

- Attack can be mounted using standard hardware and commonly used drivers;
- Attack consumes limited resources on the attacking node, so is inexpensive to mount;

- Vulnerability being exploited will not be mitigated by emerging MAC layer security enhancements i.e. IEEE 802.11 TGi; and

- There is currently no defense against this type of attack for DSSS based[13] WLANs.

6.4.1 Attack Description

To facilitate correct MAC operation, the IEEE 802.11 standard mandates that a Station Management Entity (SME) will be present and able to interrogate layer specific status and control layer specific parameters. The two layers controlled via the SME are the MAC Layer Management Entity (MLME) and the PHY Layer Management Entity (PLME). Interactions among the management entities are depicted in Figure 6.4[14].

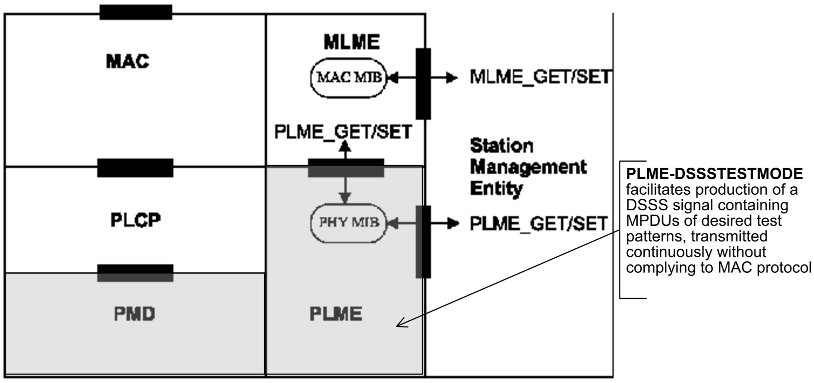

MAC
MLME
MAC MIB
MLME_GET/SET
PLME_GET/SET
Station Management Entity
PLCP
PHY MIB
PLME_GET/SET
PMD
PLME
PLME-DSSSTESTMODE facilitates production of a DSSS signal containing MPDUs of desired test patterns, transmitted continuously without complying to MAC protocol

Figure 6.4: Station Management Entities (SME)

The attack described here takes advantage of an optional PLME service primitive[15] that places the network card in a test mode of operation capable of continuously transmitting a specified bit pattern on a given channel.

Once the attacking node begins transmitting this pattern, all stations within range of the transmission, including AP's, will receive a PHY-CCA.indicate(BUSY) assessment of the channel state until the attacking node is disabled. This results in clients of the network perceiving the AP as out of range (Figure 6.5). The effect of the attack, as observed in our experimentation is almost instantaneous.

The following subsections detail the affects of the attack on PLCP receive and transmit procedures based on preliminary experimentation.

[13] These are the high rate WLANs

[14] Image sourced from [61].

[15] PLME-DSSSTESTMODE.request

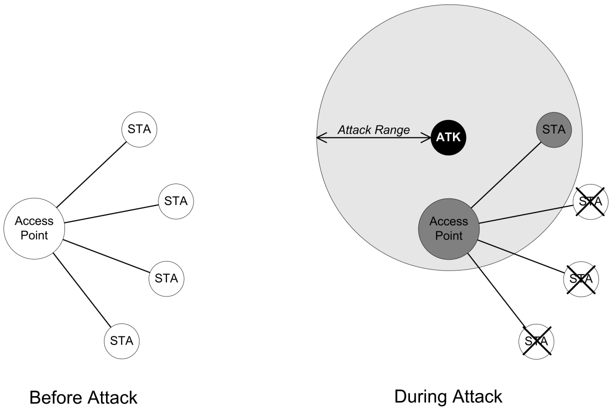


Figure 6.5: Attack on Infrastructure WLAN

6.4.2 Affects of Attack on the PLCP Receive Procedure

The affect of an attack against PLCP receive procedures is relevant where the attacking station is within range of a station, but out of range of the access point the station is associated with. In this case, beacons are successfully transmitted from the access point allowing stations to be associated and synchronized. The PLCP receive procedure is affected as follows for stations within range of an attacking node.

1. *The PLCP frame is transmitted by the attacking node while the media is idle.* In this case, the CCA algorithm detects the carrier and the sync pattern is successfully detected (Figure 6.7[16]). The PLCP frame sent by an attacking node is received, but fails the CRC check. Observations made during preliminary testing confirmed that PLCP frames received contained CRC errors. As the PLCP CRC test fails, CRC FAIL is returned and the state machine returns to RX_IDLE state as illustrated in Figure 6.7.

2. *The PLCP frame is transmitted by the attacking node while a beacon frame or a frame from another station is being transmitted.* The result of this is a collision, such that the sync pattern cannot be detected. The result is a loop where the CCA detects a carrier, sync pattern detection fails, resulting in a return to the state of RX_IDLE. Inevitably all beacon frames successfully transmitted will collide with frames transmitted by the attacker, resulting in disassociation from an access point in the attacked channel.

Observations made during preliminary testing confirm that both conditions cause the station port status to change from “Connected to IBSS” to “Out of Range (ESS)”.

[16] Image sourced from [61].

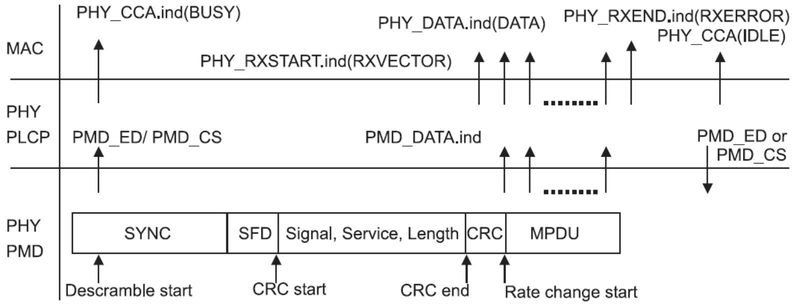


Figure 6.6: PLCP Receive Procedure

6.4.3 Affects of Attack Against the PLCP Transmit Procedure

When using the DCF to coordinate data transmissions in either infrastructure (BSS) or ad hoc (IBSS) mode the following transmit procedure is followed by the PLCP layer (See Figure 6.8[17]).

Prior to any transmission the PHY-CCA.indicate(IDLE) must be received by the MAC layer and appropriate IFS times must have elapsed. Once the channel is assessed to be clear, transmission of the PPDU is initiated by the MAC layer issuing the PHY_TXSTART.request(TXVECTOR) primitive. The PLCP layer then sets PMD parameters such as the antennae to use, transmission rate and transmission power level. Transmission of the data is then commenced via the PMD_TXSTART.request primitive. Transmission terminates when data transfer has completed, or the MAC layer issues a PHY-TXEND.request primitive.

The denial of service attack described in this chapter effects the PLCP transmit procedure directly and indirectly in the following ways.

1. *The attack occurs whilst a station is transmitting a frame.* In this case the transmission was initiated successfully, but part of the transmission is corrupted due to the initiation of an attack. An acknowledgment for the transmitted frame will not be received resulting in retransmission and an increase of the contention window[18].

   Acknowledgment failure or the receipt of PHY-RXEND.indication errors will result in the deferral of transmissions for a period greater than or equal to EIFS.

[17] Image sourced from [61].

[18] Backoff time is based on a random value between the minimum and maximum values of the contention window.

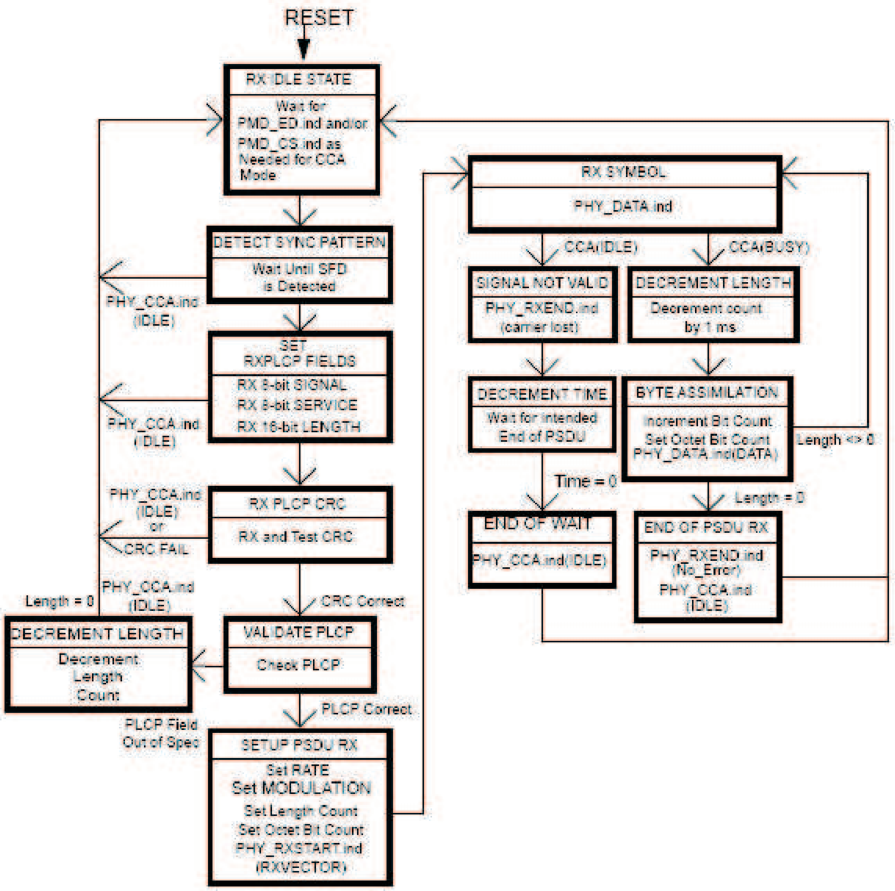


Figure 6.7: PLCP Receive State Machine

The EIFS is used to resynchronize the station to the actual medium state, such that reception of an error-free frame during EIFS, results in the resumption of normal medium access procedures using the DIFS and backoff timers. Because an error-free frame is unlikely to be received, the station could remain in this state for the duration of the attack.

In the event the station is able to resynchronize, subsequent retransmissions are not able to occur, as it is unlikely the station will receive a CCA.indicate(IDLE).

2. *The attack occurs whilst the station is idle.* The continuous transmission of data onto the shared media, by the attacking node, dramatically reduces the likelihood that any station will receive a CCA.indicate(IDLE) that is a prerequisite for any PLCP transmission to commence.

   As the media is sensed as busy during a backoff slot, the backoff procedure is

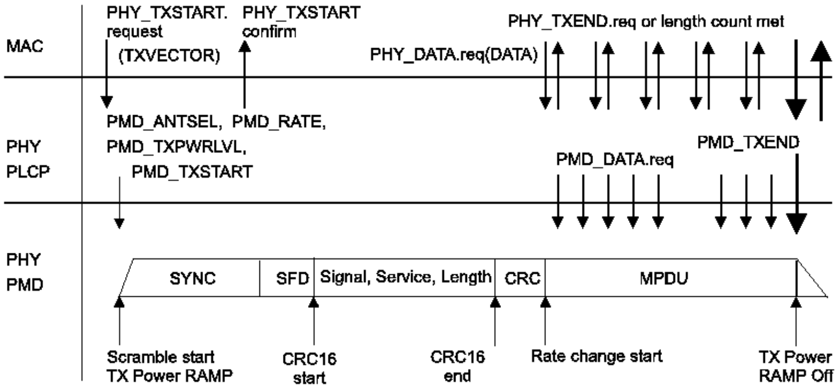


Figure 6.8: PLCP Transmit Procedure

suspended and the media must be sensed as idle for a period of time equal to the appropriate IFS[19].

Even if a station was able to assess CCA.indicate(IDLE), any transmissions from a station are likely to collide with those of the attacker resulting in the MAC layer backoff procedures being activated.

The affect of the attack on PLCP transmission procedure is illustrated in Figure 6.9.

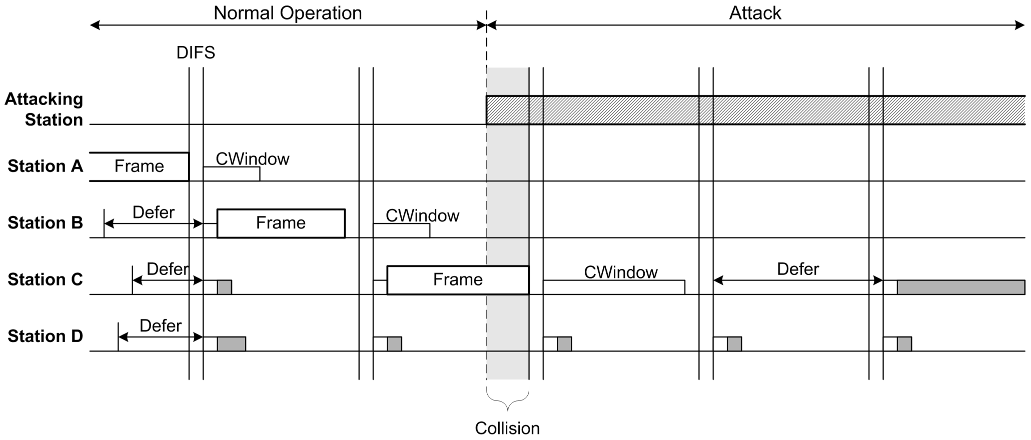


Figure 6.9: Backoff Procedure During Attack

[19] DIFS if no errors have been detected, EIFS if errors have been detected.

6.4.4 Attack Implementation

The attack was implemented on a laptop running Linux Fedora Core 1 and a Compaq iPAQ running Familiar Linux. Both implementations used the *linux-wlan-ng* drivers[20] and PRISM based wireless network interface cards.

The goal of the linux-wlan project is to develop a complete, standards based, wireless LAN system using the GNU/Linux operating system that provides a convenient interface to low level functionality of the PRISM based network cards. Low level configuration of the device is made possible via the user level application, *wlanctl-ng p2req_low_level* series of commands. These permit the manipulation of the management entities described in §6.4.1.

6.5 Analysis of the Attack

The following subsections detail the procedures used to create the attack, and how the attack was tested.

6.5.1 Accessing PLME-DSSSTESTMODE

As discussed in § 6.4, the attack described in this chapter depends on being able to access the test mode of operation via the PLME. Our testing revealed that this service primitive is implemented in a range of Intersil Prism2 based WLAN cards (Figure 6.10) and can be accessed programmatically via the wlan-ng drivers available for the Linux operating system. The network cards and drivers were tested on both desktop / laptop Intel PCs and a Compaq iPAQ 3750 running Familiar Linux.

6.5.2 Setting up the Attack

The following procedure is based on specific implementations of PHY sublayer functionality available through the wlan-ng driver for PRISM-based WLAN cards.

1. *Lock station to a given channel.* The station mode is set to WDS (specifies a repeater function in a wireless network), such that the channel is not changed back to an available channel it can associate to an access point.

[20] Refer to Absolute Value Systems `http://www.linux-wlan.org` for linux-wlan-ng driver implementation.

Figure 6.10: D-Link DWL-650 Prism2-based PCMCIA WLAN card used for testing attack

2. *Set transmission channel.* PHY services are used to change the current channel to the specified channel for the attack.

3. *Launch Attack.* The station's continuous transmit test mode is activated, such that the station radio is put in a continuous transmission state. MPDUs containing a test data pattern are continually transmitted until the test is stopped.

In order to quantify the affects of the attack, tests were conducted in a tempest-shielded room with 802.11b Wireless LAN equipment from a number of vendors. The attack program illustrated in Figure 6.11, was developed for the purpose of automated testing. The test results are presented in the following subsection.

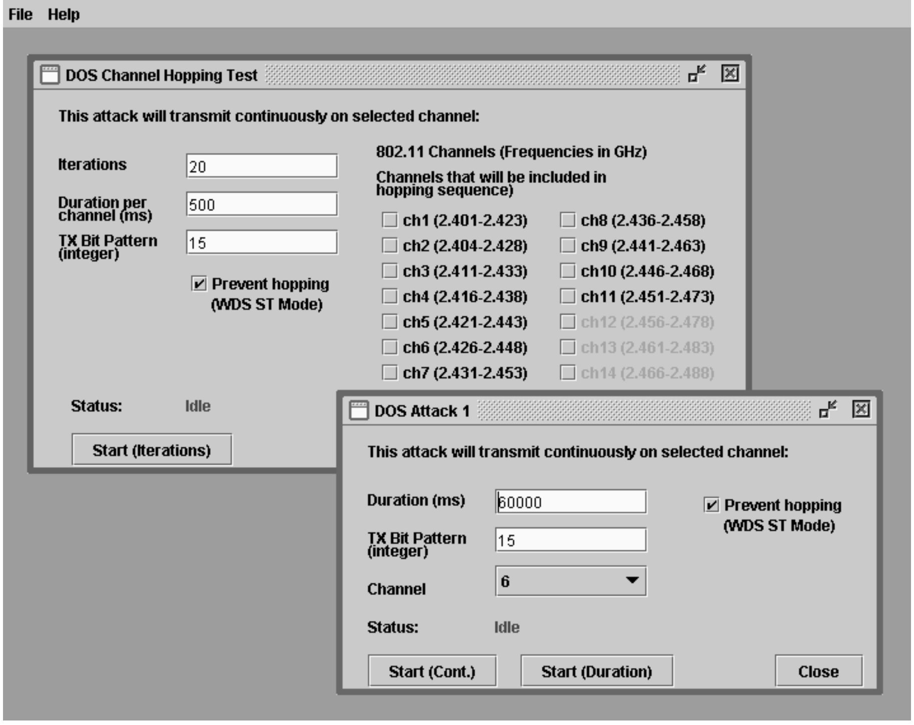
6.5.3 Attack Results

Testing was conducted in cooperation with the Australian Defense Signals Directorate on a limited number of PRISM, Orinoco and Aironet-based WLAN cards.

Experimentation confirmed that attacks performed within range of an access point significantly improve DOS results, as all associated stations are denied service. Ad hoc networks were found to be more resilient to attack than infrastructure networks as they are only affected by attackers within range. An attacker is able to use open source tools such as "Airfart"[21] to locate access points to optimize the attack.

In particular, attacks on infrastructure mode wireless LANs were highly effective, resulting in total denial of service for all wireless stations associated with the channel being attacked. The tests conducted made use of a PCMCIA wireless LAN card with

[21] See `http://airfart.sourceforge.net/`.



DOS Channel Hopping Test

This attack will transmit continuously on selected channel:

Iterations: 20

Duration per channel (ms): 500

TX Bit Pattern (integer): 15

☑ Prevent hopping (WDS ST Mode)

802.11 Channels (Frequencies in GHz)

Channels that will be included in hopping sequence)

□ ch1 (2.401-2.423)
□ ch2 (2.404-2.428)
□ ch3 (2.411-2.433)
□ ch4 (2.416-2.438)
□ ch5 (2.421-2.443)
□ ch6 (2.426-2.448)
□ ch7 (2.431-2.453)
□ ch8 (2.436-2.458)
□ ch9 (2.441-2.463)
□ ch10 (2.446-2.468)
□ ch11 (2.451-2.473)
□ ch12 (2.456-2.478)
□ ch13 (2.461-2.483)
□ ch14 (2.466-2.488)

Status: Idle

DOS Attack 1

This attack will transmit continuously on selected channel:

Duration (ms): 60000

TX Bit Pattern (integer): 15

Channel: 6

☑ Prevent hopping (WDS ST Mode)

Status: Idle

Figure 6.11: Automated Attack Tool

a transmission power of 23mW. The attack was virtually instantaneous with D-link, CISCO and Orinoco hardware. When a CISCO access point and only CISCO cards were used, they were slightly more resilient to the attack than the other cards. This could indicate that CISCO cards have a more robust implementation of CCA.

Figure 6.12 and 6.13 illustrate the power over the 2.4 Ghz FCC Wireless LAN frequency domain (channels 1-11). Figure 6.12 illustrates normal activity on channel 7 (a continuous transmission of data from one node to another in infrastructure mode), with the attack taking place on channel 3. As can be seen from the spectral graphs, it is unlikely that any CCA based on a power threshold (Mode 1) would detect a clear channel.

Figure 6.13 illustrates the attack taking place on channel 7 whilst data was being transmitted on the same channel. All data transmission immediately stopped, and the nodes involved in the data transfer were unable to detect presence of the wireless LAN.

Jamming on adjacent channels was also found to be effective, however this is dependent on proximity. There are some signal rejection characteristics for any 2 channels with less than 25Mhz separation. The adjacent channel is rejected as long as it is $<$ 35dB. We found that the attack power generally exceeds this threshold for adjacent

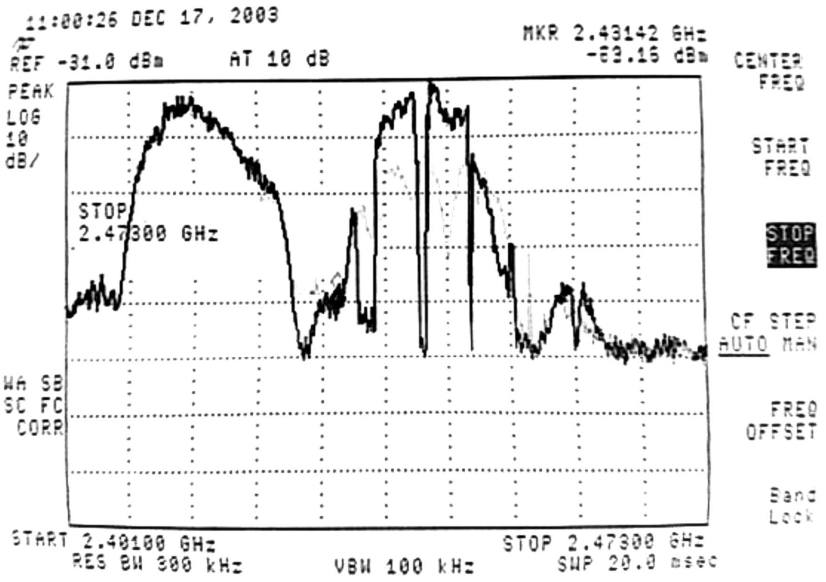


Figure 6.12: Attack on Channel 3, AP and Stations on Channel 7

channels when in sufficient proximity.

The wireless LAN standards [61] specify that transmitted spectral products shall be less than 30 dBr (dB relative to the SINx/x peak) for
$f_c - 22MHz < f < f_c - 11MHz$; and
$f_c + 11MHz < f < f_c + 22MHz$;
and shall be less than -50 dBr for
$f < f_c - 22MHz$; and
$f > f_c + 22MHz$.
where f_c is the channel center frequency (Figure 6.14[22]).

Overlapping and/or adjacent cells using different channels can operate simultaneously without interference if the distance between the center frequencies is at least 25 MHz. Adjacent channel rejection is defined between any two channels with $> 25MHz$ separation in each channel. The WLAN standards [61] define that adjacent channel rejection must be equal to or better than 35 dB.

Attacks initiated on adjacent channels can cause interference through raising the

[22] Image sourced from [61]

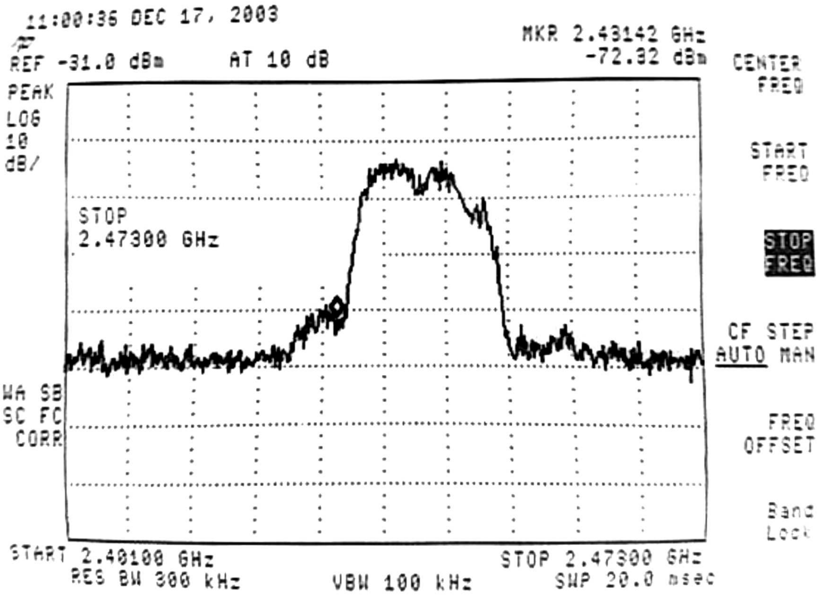


Figure 6.13: Attack, AP and Stations on Channel 7

noise floor of a given channel. Filtering is usually performed to minimize interference from adjacent channels, however Adjacent Channel Interference (ACI) generates side lobe energy that if stronger than the signal on a given channel, can dominate the channel's noise floor.

This effectively results in an increased noise floor that dominates the channel's signal-to-interference ratio (SIR)[23], reducing signal strength.

The effects of adjacent channel rejection and wireless LAN performance are discussed by [95], who states that WLAN RF receivers are designed with effective ACR for narrow band signals which are approximately 0.10 the bandwidth of 802.11 signals (Bluetooth, cordless phones, etc), however wide band ACI generates significant side band energy which falls into the pass band of an 802.11 receiver. In this case the increased size of the SIR will have a decisive effect on the data throughput of the WLAN.

The wireless LAN standards [61] define an optional channel disruption mitigation measure, *Channel Agility*, which allows a WLAN to move to different channels to

[23] Signal-to-interference is a ratio between the wanted signal power in the channel and the interference in the channel.

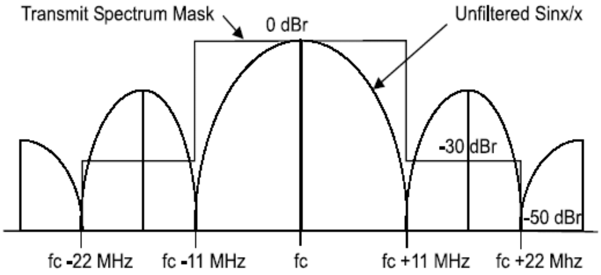


Figure 6.14: Transmit Mask

reduce the affects of interference caused by jamming and static channel assignments. None of the 802.11 hardware used in testing supported this option. This was confirmed though register lookups in the PRISM-based cards, and from spectral observation in the Orinoco and CISCO cards. The frequency agility option specifies two sets for channel hopping. The first set is the set of non-overlapping channels for the FCC domain, (1,6,11), the second is (1,3,5,7,9,11). (Figure 6.15[24])

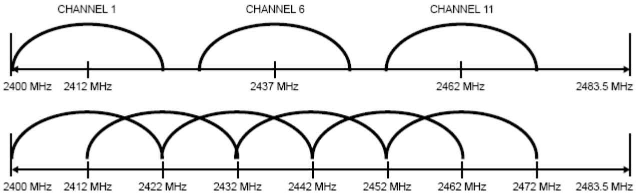


Figure 6.15: U.S. (FCC) Non-overlapping and Overlapping Channels

The first set would be easily disrupted by jamming the above 3 channels. If the second set is being used, it would require more jammers. While jamming (2,4,6,8,10) would seem logical, it would only increase the noise floor of the channels that we want to jam. In addition, channels 1 and 11 can successfully transmit an amount of data in some circumstances (as we observed with CISCO cards), as there is no cross channel interference from $<$ 2.412Ghz and $>$ 2.463Ghz. By jamming (1,3,5,7,9,11), not only is the noise floor increased for all adjacent channels, it facilitates a direct attack exploiting the CCA on these channels. We observed that this attack would result in DOS for the total band. Jamming every third channel was not sufficient for denial of service to the middle channels.

Testing of multiple attack nodes on the same channel was performed. This resulted

[24] Image sourced from [61]

in a significant increase of power, visible on the spectral analyzer. This would be effective in increasing the range of the attack for ad-hoc or for jamming adjacent channels, where a sufficient noise floor is required to facilitate denial of service. The primary attack described in this chapter, performed within the vicinity of an access point, would not significantly benefit from increased transmission power, except in the interests of performing the attack remotely.

The attack range could be significantly improved through the increase of transmission power and high-gain antennas. The maximum transmission power, as licensed by the Australian Communications Authority, for DSSS on the ISM band (2.4Ghz) is limited at 4 Watts EIRP[6].

Tool such as “airfart” and “Ethereal”[25] are able to detect attackers and their approximate location. Experimentation indicated that the jamming attack disables these tools.

6.6 Discussion

The IEEE WLAN standards are dynamically evolving with new features and capabilities constantly under consideration. This section considers the impact that these evolving features and capabilities are likely to have on the existing attacks against WLAN’s described in § 6.2 and the new availability attack described in § 6.4.

6.6.1 Existing Attacks

The attacks described in § 6.2 result from poorly designed security protocols and the inability of stations to distinguish between authentic or forged management frames. Significant efforts have been undertaken by the IEEE 802.11 TGi to rectify the security issues that plagued WEP and they have specified a comprehensive security framework for providing significant security improvements to existing and future wireless LAN standards. The key management framework detailed by IEEE 802.11i can be used to ensure that management frames are authenticated and reduce the risk that the attacks described can be successfully mounted.

[25] See `http://www.ethereal.com/`.

6.6.2 New Attack

Unfortunately, while the security efforts of IEEE 802.11i are capable of mitigating the security risks presented by existing attacks against 802.11 based WLANs, they will not have any impact on the new attack described in this chapter. Primarily because the 802.11i solutions are applicable at the MAC layer and the attack described operates at the PHY (PMD and PLCP) layer.

It is worth noting that the attack described in this chapter is dependent on the following: (1) access to an exposed low level interface capable of placing the attacking network card in *DSSSTESTMODE*; (2) the dependency of *DSSSTESTMODE* on the use of a DSSS PHY layer; and (3) the use of shared communications channel. The capability of mounting this attack using next generation WLAN cards based on the IEEE 802.11a specification is uncertain. Firstly, these cards operate with a different PHY layer, based on Orthogonal Frequency Division Multiplexing (OFDM) and there may not be access to a *DSSSTESTMODE* equivalent function in cards based on this standard.

While this chapter has focused on the effect of the new attack on networks using contention based MAC protocols such as DCF, the contention free MAC protocols such as PCF will be vulnerable too as they depend on the same CCA procedure as DCF[26]

A fundamental factor contributing to the effectiveness of this attack is the use of a shared communications channel. One strategy for making the attack more complex to mount would be to discard the shared communications paradigm in favor of dedicated communications channels based on dynamically negotiated spreading sequences. This would result in an attacker having to jam each channel independently or jam an entire frequency, increasing the power requirements to mount the attack and the risk of being detected and localized.

6.6.3 Possible Mitigation Measures

Intermediate solutions could look at proving alternate spreading sequences for safety critical environments, industrial environments, etc. In these environments, it is not necessary to have interoperability with commercial-off-the-shelf WLAN products. As such, spreading codes that are orthogonal to the published codes used in 802.11 could be used. This would assist in preventing trivial attacks such as those demonstrated in this chapter. The spreading code would be unique to a given wireless LAN and the

[26] The PCF is not widely implemented, so extensive testing using this mode of operation has not been performed.

nodes within it. Wireless LAN cards and access points would need to be modified via a firmware upgrade to support the new spreading sequence. As the sequence is kept secret, the engineering effort required to attack the wireless LAN would be significantly more than the trivial attacks that can be currently performed.

Future research needs to be conducted in developing highly available wireless networks that are suitable for safety-critical environments. Shared-media environments, such as 802.11, are clearly not appropriate for these environments due to the ease with which denial of service attacks can be performed. The use of cryptographically generated spreading codes in a CDMA-based network is a direction that could be investigated, in that it provides separate logical channels for each user, significantly increasing the difficulty in performing a DOS attack. There remains the problem of channel association. Existing CDMA technologies use a random-access channel to allocate a logical channel to a user. If the random-access channel is jammed, this effectively denies access to all wireless users requesting a logical radio channel. A significant research challenge is to find a method of establishing spreading codes between two entities, without the need for a random access channel, whilst maintaining scalability.

6.7 Summary

This chapter has presented a highly effective denial of service attack that can be mounted against IEEE 802.11 WLANs using a DSSS PHY layer. The attack is significant as it can be achieved using only commodity based hardware and software; has low power requirements; and can be executed with minimal chance of detection and localization.

A critical observation that can be made from this work is that the presence of engineering modes of operation, such as the *DSSSTESTMODE* interface exploited by this attack, in production hardware can present significant security concerns. These concerns are most pronounced in network environments that rely on the correct behavior of participating nodes for continued operation.

Until adequate strategies are in place to mitigate the significant threat of denial of service in current IEEE 802,11 DSSS WLAN technology, the application of this technology should be precluded from use in safety-critical environments which typically have stringent availability requirements[27].

[27] Refer to Appendix I for CERT advisories that have resulted from this research.

Chapter 7

Supporting Context-aware Access Control in Network Authentication and Authorization Architectures

7.1 Introduction

Ubiquitous computing is a rapidly growing research area, however its uses have been predominantly targeted at context-aware applications for smart spaces such as smart homes and workplaces. In this chapter we present an authorization architecture that uses the context-aware paradigm to augment existing and time-tested technologies such as Kerberos, facilitating fine-grained access control to network resources and effective enforcement of security policies.

Context-aware authorization has a number of requirements traditional access control mechanisms do not provide. First the authorization system must be able to dynamically grant and revoke permissions based on an access control policy and the continually changing context of a user. It cannot be assumed that a given set of authorization credentials will persist for the lifetime of a session. In order to support such access control policy, permissions must be centralized and a common representation of context data must be used.

The addition of context-awareness to access control systems significantly increases complexity, however it is a requirement of our architecture that administration of security policy must retain the same administrative efficiencies afforded by the use of role based access control.

The proposed architecture supports GSSAPI-based applications through the use of Kerberos. Common applications that currently support Kerberos include SMB file sharing, database servers (e.g. PostgreSQL), CVS and Java-based applications though JAAS[1]. The use of GSSAPI[2] allows existing network applications to be easily migrated to the proposed architecture.

The structure of this chapter is as follows. First access control mechanisms used for distributed access control are reviewed in § 7.2. Traditional architectures and emerging efforts in pervasive computing security are reviewed in § 7.3 and § 7.4. The goals of our architecture design are presented in § 7.5. § 7.6 details the formal specification of an object-oriented representation of RBAC, which is able to provide support context-awareness. The proposed architecture and its components are described in § 7.7, and the protocol used in the architecture is detailed in § 7.8. Implementation details including a description of the testing environment and several performance results of the prototyped architecture are detailed in § 7.9. A summary of the chapter is presented in § 7.10.

Portions of this chapter have been published in the paper: Towards Context-aware Security: An Authorization Architecture for Intranet Environments. In Proceedings of the First IEEE International Workshop on Pervasive Computing and Communications Security (PerSec 2004), Orlando FL, USA, March 2004.

7.2 A Review of Access Control Mechanisms

The following subsections perform a review of historical, current and emerging work in access control.

7.2.1 Access Control Background

Early access control models from the 1970s to early 1980s evolved from two fundamental types of access control:

1. *Discretionary Access Control (DAC).* Users are responsible for controlling access to objects they own. DAC has a number of shortcomings. First, DAC does not provide any real assurances on the flow of information in a system. Once a

[1] See Java Authentication and Authorization Service `http://java.sun.com/products/jaas/`

[2] See RFC 2743 - Generic Security Services Application Program Interface

user is in possession of an object, DAC does not impose any restrictions on its usage. Objects can be copied, such that access to a copied version is possible regardless of the access the original owner has set.

2. *Mandatory Access Control (MAC).* A system-enforced access control mechanism that uses clearances and sensitivity labels to enforce security policy. Access to information is only permitted where a subject's security clearance at least equivalent or at a higher level than the object's security clearance. MAC has a number of shortcomings. Information flow can pass through covert channels in prohibited ways, and MAC does not suit the requirements of industry and government organizations.

In the mid 1980s to mid 1990s, a number of alternate models appeared. Of particular significance was Role Based Access Control (RBAC), which still remains an active area of research today, with only few commercially available RBAC systems. Access control decisions made using RBAC are based on the roles that users have as part of an organization, and the access rights that are grouped with the roles. RBAC can be an effective method for minimizing administration, and developing security policies that compliment an organization's operations and information flows. RBAC is discussed in detail in the subsequent subsection.

7.2.2 Role-based Access Control

RBAC is designed to centrally manage user's privileges by providing layers of abstraction that are mapped to real users, operations and resources. User membership to a set of roles determines the permissions that are acquired in a given user session. One of the distinguishing and perhaps most useful features of RBAC is its support for role hierarchies. Permissions are managed in terms of abstractions such that operations and resources are given abstract names which are translated into real permissions on real systems. This reduces complexity and facilitates a context in which access control policies can be easily implemented.

The motivation for the use of RBAC in access control is principally concerned with the efficient management of access rights of large-scale systems. A number of benefits of RBAC have been discussed by Gallaher, O'Connor and Krop in [72], who identify a number of benefits RBAC has over other access control technologies. Significant benefits of RBAC include the simplification of system administration, the enhancement of organizational productivity, reduction in employee downtime, enhanced systems security and integrity, and simplified regulatory compliance.

There have been many efforts over the past years to define RBAC and work towards a unified RBAC standard. The first comprehensive framework for RBAC models was defined by Sandhu, Coyne, Feinstein and Youman in [83]. The framework consisted of four models of RBAC, that ranged from simple to complex:

1. *RBAC$_0$:* The most basic RBAC model, where users are associated with roles, and permissions are associated with roles;

2. *RBAC$_1$:* Builds on RBAC$_0$ by introducing role hierarchies;

3. *RBAC$_2$:* Builds on RBAC$_0$ by introducing constraints such as separation of duties; and

4. *RBAC$_3$:* Combines RBAC$_1$ and RBAC$_2$ such that constraints can be applied to a hierarchy of roles.

Since this initial family of models, there has been much research, which to a great extent has resulted in the establishment of a standard RBAC model, supported by the US National Institute of Standards and Technology. The proposed NIST standard for RBAC [52] categorizes access control management features into a number of cumulative functional packages:

1. *Core RBAC:* Core RBAC contains the following basic requirements that are essential for any form of RBAC to operate:

   (a) *User-role assignment is many-to-many*, such that a single user can be assigned to many roles and a role can have many users;

   (b) *Permission-role assignment is many-to-many*, such that a single permission can be assigned to many roles and a single role can have many permissions;

   (c) *User-role review*, such that the users assigned to a given role and the roles assigned to a given user can be determined;

   (d) *Permission-role review*, such that the permissions assigned to a given role and the permissions a role has can be determined. This feature is an advanced review requirement in that it is not mandatory in core-RBAC;

   (e) *User sessions*, such that the activation and deactivation of roles is allowed within a session; and

   (f) *Multiple roles*, such that a user is able to exercise the permissions of multiple roles.

2. *Hierarchical RBAC:* Hierarchical RBAC adds support for role hierarchies to core RBAC. A role hierarchy allows a role to inherit the permissions of their child-roles, and the child-role to acquire membership of their parent-roles. There are two types of hierarchies supported in hierarchical RBAC:

   (a) *Limited Hierarchy*, where only single inheritance of permissions and role membership is supported. This is typically facilitated through a simple tree structure; and

   (b) *General Hierarchy*, where multiple inheritance of permissions and role membership is supported.

3. *Static Separation of Duty (SSD) Relations:* Static separation of duties facilitate the enforcement of conflict of interest policies through the use of constraints on the assignment of users to roles. Consequently, a user is prevented from being assigned permissions of conflicting roles. These static constraints can be based on users, operations and objects. An example of such constraints are the requirements for prerequisite roles, role cardinality, and mutually exclusion of roles. This functional package supports the following SSD relations:

   (a) *SSD*, where constraints are placed on the assignment of users to roles.

   (b) *SSD in the presence of a hierarchy*, where inherited roles are considered in addition to assigned role for the enforcement of constraints.

4. *Dynamic Separation of Duty (DSD) Relations:* Dynamic separation of duties facilitates the enforcement of conflict of interest policies through the use of constraints on role activation within or across a user's sessions. DSD relations are similar to SSD relations, except that they operate within the context of role activation rather than user-role assignment as in SSD relations.

Of particular interest to our research is the adaptation of these models to context-aware environments.

Temporal RBAC (TRBAC) was introduced by Bertino et al. [11], which proposed the use of time-based constraints and activation dependencies for role activation. TRBAC provides support for periodic activation/deactivation of roles, and temporal dependencies which are expressed by means of role triggers. An example of temporal constraints can be observed in a bank teller role, who is assigned privileges on weekdays from 9am to 5pm. Temporal constraints limit role activation to these conditions.

Generalized RBAC (GRBAC) was introduced by Convington et al. in [20], to support context-awareness in a context-aware home environment. In this model of RBAC, two additional types of roles were introduced in addition to traditional subject roles, object and environment roles.

Object roles contain a membership of resources such as object types (video, audio, image), object creation dates, sensitivity, etc. Environmental roles contain a membership of environmental conditions, such as time and location. In the home environment, location roles may be defined such as “upstairs” and “downstairs”.

Access control decisions are then made based on a policy combining environment roles, object roles and subject roles. While GRBAC provides an improvement in the flexibility of security policy, it introduces complexity in administering access control policies and difficulty in ensuring they are conflict free.

7.2.3 Usage Control Models

Usage Control Models are a very recent addition access control models. The concept of Usage Control (UCON) was first presented by Park and Sandhu in [78], for the purpose of enabling finer-grained access control over the usage of digital objects in the domain of Digital Rights Management (DRM). The UCON model consists of three core components: subjects, objects, and rights, similar to those of traditional access control models. What distinguishes the UCON model from traditional access control models, is the three additional components: authorization rules, obligations, and conditions.

Authorization rules are a set of requirements that must be satisfied before granting a subject access to an object. These rules may be related to rights or obligations. Obligations are mandatory requirements that a subject must perform after obtaining or exercising rights on an object. Conditions are a set of decision factors that must be verified with authorization rules by authorization processes before granting access to an object. Both dynamic and static conditions are supported by this model. Access to objects is mediated by a client-side or server-side reference monitor, which performs the authorization processes for digital objects within its control domain.

Sandhu and Park in [86] further refine the concept of UCON, by providing a number of models and relating them to a broader range of applications outside the DRM domain. The “ABC” models[3] consist of eight components: subjects, subject attributes, objects, object attributes, rights, authorizations, obligations and conditions. Subject and object attributes are mutable properties that can be used during an access decision. The following “ABC” models have been defined:

1. $UCON_{preA}$: Pre-authorization, in which the access decision process is performed before access is granted;

2. $UCON_{onA}$: Ongoing authorization, in which the access decision checks attributes periodically based on time or events, to determine whether given requirements are still valid for access to be granted;

3. $UCON_{preB}$: Pre-obligations, such that defined obligations must be fulfilled before access is granted;

4. $UCON_{onB}$: Ongoing obligations, such that access decision process must periodically check that obligations are filled while rights are exercised;

5. $UCON_{preC}$: Pre-conditions, such that system or environmental conditions must be fulfilled before access is granted; and

6. $UCON_{onC}$: Ongoing conditions, such that access decision process must periodically check that the conditions are met while rights are exercised.

Each model can operate in a variety of modes that embody the behavior of attribute updates supported. These modes are where the attributes are (0) immutable; (1) pre-updated; (2) updated on an ongoing basis; and (3) post-updated.

While the above research [86] was published after we had developed a variation RBAC that facilitates context awareness, the research only discusses these ideas at a very high level, mostly targeted at DRM applications. Our work presents a framework, based on RBAC, that has been implemented to provide context-awareness to access control processes used in general network applications.

7.3 A Review of Network Authentication and Access Control Architectures

In this section we review existing network authentication and authorization architectures, investigate how they support access control, and where RBAC is supported, their

[3] “ABC” referring to Authorizations, Observations and Conditions.

level of compliance.

7.3.1 Kerberos

The Kerberos V5 protocol is predominantly a network authentication and key distribution protocol. It does however provide support for the inclusion of authorization data in ticket-granting and application tickets, although it is not implemented in the standard MIT distribution. The Microsoft Windows 2000 implementation of Kerberos V5 uses these protocol fields to achieve access control using proprietary extensions developed by Microsoft. These extensions are defined by Brezak in an Internet Draft[15]. The possible places for inclusion of access control data will be discussed in § 7.3.1.5.

Kerberos was developed by the Massachuttes Institute of Technology (MIT) for protection of network services provided by the Athena Project [73], version 5 of which became a standard [68].

Kerberos was designed to provide authentication, authorization and accounting services to a network environment of untrusted workstations. Kerberos provides a trusted 3rd-party service that can facilitate the mutual authentication of clients and services. The design goals of Kerberos are summarized below:

- *Authentication.* Kerberos is to provide support for:
    1. *Authentication to the Kerberos servers.* One-way authentication of principles to itself with the granularity of at least an individual user.
    2. *Network Authentication.* Mutual authentication of principles to each other with the granularity of at least an individual user and specific service.
- *Authorization.* Course-grained authorization can be implied by authentication. Kerberos is to provides support for this through a standard access control mechanism based on access control lists (ACL) containing authenticatable principle identifiers. A service may separately implement a different or finer-gained access control mechanism at its own discretion.
- *Accounting.* Kerberos supports the addition of integrated accounting services, given an authenticated principle.

The Kerberos design was based on the following assumptions [73]:

1. Workstations are insecure, with administrative management by individuals and no responsibility to a central administration.

2. Centrally operated servers are assumed to run under moderate physical security (e.g. servers in locked rooms)

3. Centrally operated servers such as the Kerberos authentication server are assumed to operate under considerable physical security.

4. The clocks of workstations, servers and Kerberos servers are assumed to be synchronized within a few minutes. Workstations are required to maintain their clock within the allowable margin. Time servers are used to facilitate time synchronization.

There are a number of limitations that must be considered in the use of Kerberos. The following subsections detail functional, environmental and protocol limitations.

7.3.1.1 Functional Limitations

Kerberos has a number of functional limitations. First, Kerberos does not provide centrally managed access control. As Kerberos is based on symmetric key cryptography, it does not scale well in large distributed environments over multiple realms. In addition, it does not provide non-repudiation services. These issues are common to all versions of Kerberos and are unlikely to be addressed in future versions.

7.3.1.2 Environmental Limitations

The environment Kerberos operates in can be critical to the security it provides. There are two significant environmental limitations:

1. *Key storage on workstations is not secure.* This includes both the long-term key and the session keys returned by the TGS. Compromising the long-term key is fatal to the security of Kerberos. A compromise of the session keys is also detrimental to the security of the system, although not as fatal as the compromise of the long-term key. Caching of session keys is an issue, as they may be compromised by a root user in a multi-user operating system. Even if session keys are cached, there is no guarantee that that memory is not paged to disk.

2. *Spoofing Login.* Bellovin and Meritt in [10] discuss how a login program can be replaced with a program that records user's passwords. This problem is due to the inherent trust of the client workstation. Due to the design of Kerberos, it is not easy to deploy countermeasures such as one-time passwords or

challenge-response authentication mechanisms, as the response to login is always encrypted with the long-term key derived from the user's password. There have been a number of smartcard-based proposals [64], [57], which remove the long-term key from the user's workstation. In addition, smartcards can be integrated into the public key extensions to Kerberos, PKINIT [98], such that a user's private key is never released to the workstation.

3. *Servers must be secure.* Compromise of the Kerberos servers is fatal to the security of the entire network, as the Kerberos servers hold all the long-term client and service keys.

7.3.1.3 Protocol Limitations

This subsection discusses known protocol limitations in Kerberos including password guessing attacks, chosen plaintext attacks, and replay attacks.

When a user wishes to obtain a Ticket Granting Ticket (TGT), a request is sent to the Application Service (AS). This request contains the username, Kerberos realm, and details of the requested ticket such as requested lifetime, services, etc. The AS returns a response containing the TGT. The TGT is encrypted with a key derived from the password. If the client has the correct password, they will be able to decrypt the response and obtain the TGT. One of the most serious vulnerabilities of Kerberos is password guessing attacks. Once a hacker obtains the password, they are able to decrypt the response and obtain a valid TGT.

Because the AS responds to all requests, it is particularly easy to compromise any user's key, by requesting a TGT on behalf of that user. The AS never actually authenticates the identity of a user, and as such is not aware of the authenticated state of a user. Pre-authentication based on timestamps is defined in Kerberos Version 5 [68], but is not mandated.

Wu in [106] describes an attack of this nature using the "service name" field of the response containing the TGT as the verifiable plaintext. This field is always "krbtgt". The authentication messages are captured using a tcpdump-like utility. This allows the field in the response, decrypted using a password generated by a password cracker, to be verified against the known field contents. Wu in [106] proposes a DES parity optimization method that significantly increases the performance of such an attack. The use of dictionary and hybrid-dictionary based password crackers also improve the performance of an attack.

This attack is also feasible on the most current versions of Kerberos without public key extensions. The Kerberos implementation in Windows 2000 is vulnerable to this type of attack as described by O'Dwyer in [77]. This similarly involves the acquisition of verifiable plaintext, which in the case of Windows 2000 is an ASCII timestamp. The format of the timestamp is known and is the basis for the offline attack. The details of how the the RC4 key is derived from the password is detailed in an Internet Draft [14]. Kerberos was designed such that the Kerberos servers did not keep track of the per-connection state of clients. In this way, the authenticator relies on the use of a timestamp to guard against replay attacks. Bellovin and Merritt in [10] discuss how a replay attack is possible within the lifetime of the authenticator.

The use of timestamps not only minimizes the state information that is needed by the Kerberos servers, it also allows the elimination of one message, or two for mutual authentication, from the protocol. Timestamps are used by Kerberos to ensure the freshness of the initial message from the client to the end-server. One of the problems with using timestamps instead of nonces is that loosely synchronized clocks are needed. Neuman and Subblebine in [76] discuss issues with the use of timestamps as nonces, such as the ability to replay a post-dated authenticator. The client clock can be synchronized to the time returned in the initial request from the Kerberos server. If the post-dated authenticator is sent before the time is synchronized, that authenticator can be replayed when the time incorrectly recorded in the authenticator is reached. The window time for post-dated authenticators is the maximum ticket lifetime.

Neuman and Subblebine in [76] suggest that this can be overcome by the client carefully checking its time against the time returned by the AS, and the storage of authenticators by end-servers in non-volatile storage if a post-dated authenticator is received.

Requiring the client to synchronize their clock to the time returned by the AS makes the assumption that the client can be trusted to synchronize their clocks and that the client software can be trusted to act correctly. Kerberos version 5 provides mechanisms where a server can require the use of traditional nonces instead of timestamps.

7.3.1.4 Version 5.0 Protocol

The notation in Appendix B is used to describe the Kerberos protocol. The Kerberos protocol is detailed as follows, and is illustrated in Figure 7.1.

1. *KRB-AS-REQ:* A client, *C*, sends an authentication request to the authentication service *AS*, with pre-authentication.

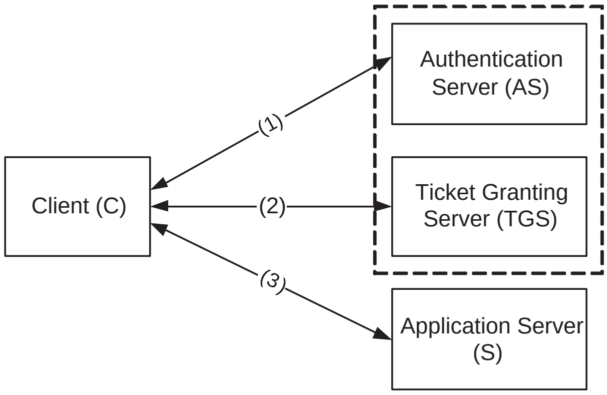


Figure 7.1: Kerberos Protocol

$C \rightarrow AS : \{TS_C\}_{K_A}, A, TGS, T_s, T_e, N_C$

KRB-AS-REP: The authentication service replies with the ticket granting ticket and key k_{C-TGS}, encrypted using key K_A.
$AS \rightarrow C : A, TGT, \{k_{C-TGS}, T_s, T_e, N_c, TGS, C\}_{K_A}$
$TGT = TGS, \{k_{C-TGS}, A, T_s, T_e, C\}_{K_{AS-TGS}}$

2. *KRB-TGS-REQ:* The client sends a request for a given service, including the ticket granting ticket to the ticket granting service. The request is encrypted using key k_{C-TGS} obtained from the authentication service.
$C \rightarrow TGS : \{A, TS'_C\}_{k_{C-TGS}}, TGT, S, T'_s, T'_e, N'_C$
$TGT = TGS, \{k_{C-TGS}, A, T_s, T_e, C\}_{K_{AS-TGS}}$

KRB-TGS-REP: The ticket granting service replies with the service ticket and key k_{C-S}, encrypted using key k_{C-TGS}.
$TGS \rightarrow C : A, STKT_S, \{k_{C-S}, N'_C, T'_s, T'_e, S, C\}_{k_{C-TGS}}$
$STKT_S = S, \{k_{C-S}, A, T'_s, T'_e, C\}_{K_{KDC-S}}$

3. *KRB-AP-REQ:* The client sends an initial request to the application service, including the corresponding service ticket obtained from the ticket granting service.
$C \rightarrow S : STKT_S, \{C, TS''_C, SN_{C-S}\}_{k_{C-S}}$

$STKT_S = S, \{k_{C-S}, A, T'_s, T'_e, C\}_{K_{KDC-S}}$

KRB-AP-REP: The application service replies to the request facilitating mutual authentication.
$S \rightarrow C : \{TS''_C, SN_{C-S}\}_{k_{C-S}}$

7.3.1.5 Placement of Authorization Data in Kerberos Messages

Kerberos provides a number of fields facilitating the inclusion of authorization data in its messages. Authorization data can be included in an authenticator field of a message, a TGT, or a Service Ticket. The authorization data element is detailed below:

```
AuthorizationData ::=    SEQUENCE OF SEQUENCE {
                         ad-type[0]              Int32,
                         ad-data[1]              OCTET STRING,
                         }
```

The authorization-data field contains the following elements for the inclusion of authorization data.

1. *AD-IF-RELEVANT:* Authorization elements encapsulated within this element are intended for interpretation by application servers that understand them. Application servers that do not understand the elements ignore this field. These authorization elements are effectively optional and as such facilitate interoperability among different implementations. In addition, the definition of the parent authorization-data field allows elements to be added by the bearer of a TGT and at the time service tickets are requested.

2. *AD-KDCIssued:* Authorization elements issued by the KDC can be encapsulated within this element. For KDC-issued elements, the elements are signed by the KDC (using an encrypted checksum). Privileges provided this way elevate the user's access from the default access provided to an application service. As such, if a signature is not present, the field is ignored and the user will not gain the privileges requested. This authorization data element is appropriate for supporting centrally managed privilege attributes, which can be included by the KDC in a cryptographically protected Privilege Attribute Certificate (PAC).

3. *AD-AND-OR:* Authorization elements included in this element are interpreted based on a condition-count element which defines whether to implement an "or"

operation or an “and” operation. Application servers that are not able to interpret this element must reject the ticket.

If Kerberos was to support RBAC, the requested role(s) could be included by the client in the authenticator of the KRB-AS-REQ message.

```
Authenticator ::=   [APPLICATION 1] SEQUENCE {
                    authenticator-vno[0]   INTEGER,
                    crealm[1]              Realm,
                    cname[2]               PrincipalName,
                    cksum[3]               Checksum OPTIONAL,
                    cusec[4]               Microseconds,
                    ctime[5]               KerberosTime,
                    subkey[6]              EncryptionKey OPTIONAL,
                    seq-number[7]          UInt32 OPTIONAL,
                    authorization-data[8]  AuthorizationData OPTIONAL
                    }
```

Any data to be included that may influence the privilege attributes returned by the KDC will be included as part of the *AD-IF-RELEVANT* element within the authorization-data field of the *KRB-TGT-REQ* message. The enc-authorization-data field is an encrypted encoding of the desired authorization data with either the specified subkey, or the session key in the TGT.

```
TGS-REQ ::=         [APPLICATION 12] KDC-REQ {
                    }

KDC-REQ ::=         SEQUENCE {
                    pvno[0]                INTEGER,
                    msg-type[1]            INTEGER,
                    padata[2]              SEQUENCE OF PA-DATA OPTIONAL,
                    req-body[3]            KDC-REQ-BODY
                    }

KDC-REQ-BODY ::=        SEQUENCE {
                        kdc-options[0]              KDCOptions,
                        sname[1]                    PrincipalName OPTIONAL,
                        from[2]                     KerberosTime OPTIONAL,
                        till[3]                     KerberosTime,
                        rtime[4]                    KerberosTime OPTIONAL,
                        nonce[5]                    INTEGER,
                        etype[6]                    SEQUENCE OF Int32,
                        addresses[7]                HostAddresses OPTIONAL,
                        enc-authorization-data[8]   EncryptedData AuthorizationData,
                        additional-tickets[9]       SEQUENCE OF Ticket OPTIONAL
                        }
```

The privilege attributes are to be included in the *AD-KDCIssued* element within the authorization-data field of the ticket returned.

```
Ticket ::=        [APPLICATION 1] SEQUENCE {
                  tkt-vno[0]           INTEGER,
                  realm[1]             Realm,
                  sname[2]             PrincipalName,
                  enc-part[3]          EncryptedData
                  }

EncTicketPart ::=     [APPLICATION 3] SEQUENCE {
                      flags[0]                   TicketFlags,
                      key[1]                     EncryptionKey,
                      crealm[2]                  Realm,
                      cname[3]                   PrincipalName
                      transited[4]               TransitedEncoding,
                      authtime[5]                KerberosTime,
                      starttime[6]               KerberosTime OPTIONAL,
                      endtime[7]                 KerberosTime,
                      renew-till[8]              KerberosTime OPTIONAL,
                      caddr[9]                   HostAddresses OPTIONAL,
                      authorization-data[10]     AuthorizationData OPTIONAL
                      }
```

7.3.1.6 Supporting Role-based Access Control in Kerberos

Role-based Access Control (RBAC) can be supported in Kerberos, but only in a very limited form. As demonstrated in the previous subsection, KDC-authorized roles can be included in the ticket-granting / service tickets. RBAC in this operation mode does not support permission-role review, and as such does not strictly meet the requirements for the core RBAC package of the proposed NIST standard [52]. Access control decisions in this mode of RBAC are made by application services based on the roles presented to the application service in the service ticket. The roles may be mapped to entries in ACLs within the application service, similar to that of group-based access control.

A significant issue with the binding of roles to ticket-granting tickets is that there is no support for revocation of privileges except through the expiry of a ticket. In addition, this prevents a user from changing their roles within a session.

7.3.2 Distributed Computing Environment

The Distributed Computing Environment (DCE) V1.2.2 [96], developed by the Open Software Foundation, augments the Kerberos V5 protocol to provide access control through the transfer of additional security attributes in Kerberos tickets. DCE authentication and authorization proceeds as follows:

The first and second interactions in DCE, as illustrated in Figure 7.2, are the same as Kerberos, except that the client obtains a ticket to the DCE Privilege Service (PS) instead of a destination application service as in Kerberos.

The client obtains a Privilege Ticket Granting Ticket (PTGT) from the PS (Third interaction in Figure 7.2), which contains an Extended Privilege Attribute Certificate (EPAC) seal (a cryptographic checksum of the EPAC). The EPAC contains additional attributes such as user and group identities.

The client subsequently requests a privilege ticket to the application service from the ticket granting service, providing the PTGT. The TGS responds with a privilege ticket to the requested application service, and interactions proceed as with standard Kerberos.

DCE provides an RPC interface from which a client's EPAC can be obtained and verified against the EPAC seal provided in the privilege ticket. DCE servers protect their resources using Access Control Lists (ACLs), such that an ACL entry corresponds to a user or group included in the EPAC.

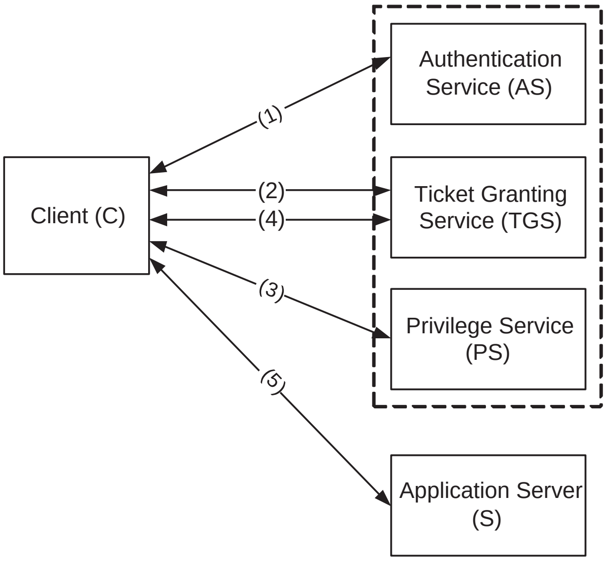


Figure 7.2: Distributed Computing Environment (DCE) Architecture

7.3.3 Secure European System for Applications in a Multi-vendor Environment (SESAME)

SESAME is an augmentation to Kerberos, supporting both standard Kerberos key distribution and support for public key extensions. SESAME additionally provides support for authorization data in Privilege Attribute Certificates (PACS), defined in Standard ECMA-219[30], which are used in SESAME's implemntation of RBAC, as well as support for group-based access control.

SESAME is built around a "push model" of privilege management, where privileges are pushed to the target application server by the client as shown in Figure 7.3. This push model is advantageous in supporting least privilege, where the target application is only aware of the privileges it needs to know.

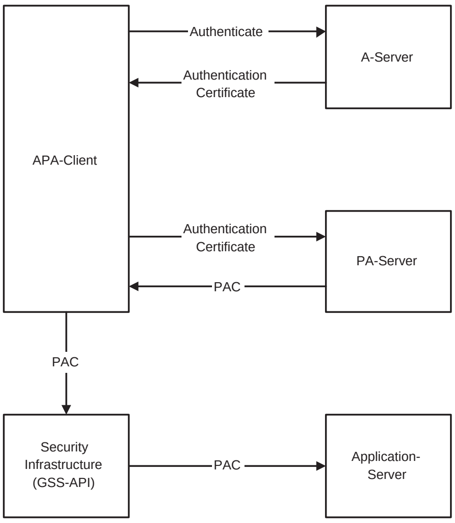

Authenticate
A-Server
Authentication Certificate
APA-Client
Authentication Certificate
PA-Server
PAC
PAC
Security Infrastructure (GSS-API)
PAC
Application-Server

Figure 7.3: Overview of SESAME

The SESAME version 5 PAC can contain the following privilege attributes types as defined in the SESAME V5 Internet Draft[28]:

1. *Access Identity:* An identity either specified in the SESAME Privilege Attribute Server (PAS), or the default authenticated identity;

2. *Primary Group:* The primary group of which the owner of the PAC is a member;

3. *Secondary Group:* The secondary group of which the owner of the PAC is a member; and

4. *Role Attribute:* An attribute corresponding to the role-name specified in the PAC request.

The SESAME Privilege Attribute Server (PAS) is responsible for the administration of roles. Role permissions however, are managed by application servers. The PAC does not contain permissions, however it is instrumental in proving to the application server that the role the user activates has been authorized by the PAS.

SESAME is a distributed authentication service that supports “Identity-based” authorization, where an access identity, generally the authenticated identity, and the corresponding role or group is bound to a PAC and used by the access control mechanism of the target application to make access control decisions.

Application servers that participate in SESAME typically make use of Access Control Lists (ACLs) for application specific authorization, where ACLs contain a role or group and associated permissions. An application server makes access control decisions based on the role or group name provided in the PAC. Both group-based and RBAC methods have the same affect on access control. There is little to distinguish a group from a role in this implementation.

7.3.3.1 SESAME RBAC Compliance

The implementation of RBAC facilitated by the SESAME architecture does not fulfill the following two requirements specified in the core RBAC model of the proposed NIST standard [52]:

1. *Support for multiple roles.* The proposed NIST standard [52] requires that a user is able to exercise the permissions of multiple roles. SESAME V5 does not support multiple roles. For each user there is a list of roles in which a user can act, however each user can only activate a single role at a time; and

2. *Permission-role review.* While this is an advanced review function of the core RBAC model imposed by the proposed NIST standard [52], the SESAME architecture is unable to provide such functionality. It is required that the permissions assigned to a given role and the permissions a role has can be determined. This requirement can been shown using an example used by Sandhu, Coyne, Feinstein and Youman in [83] is UNIX access control. In UNIX, groups and group membership is defined by the files /etc/passwd and /etc/group. The actual permissions are defined by access control lists containing the permission bits associated with a file. To determine the permissions for a given group, the entire file system must be traversed, taking a considerably longer time than determining the group membership. It would be infeasible to determine permission assignments for a given role for all applications in a domain. The control of membership and permissions should be relatively centralized in a few users. SESAME’s RBAC

implementation fails this requirement because role permissions for applications within a domain cannot be determined from the PAS within that domain, whereas role membership can be determined from the PAS very quickly.

In addition to the above non-compliance, there are also a number of issues with SESAME's architecture that hinder its ability to support the other functional packages defined by the proposed NIST standard [52].

1. *SESAME only provides support role activation.* A user must logout of their current role in order to change their role. Any PACs issued in the previous role continue to be valid until the PAC expires;

2. *Privilege revocation is managed through the use of expiry times in PACs.* The SESAME PAS cannot revoke a previous PAC if a user wishes to change their role. This is because the push model used would require that a revocation PAC be issued to the target application via the user. A user could simply choose to not pass on the revocation PAC, defeating the access control policy. In this way, there is no possibility for the implementation of constraints such as separation of duties. Enforcement of such constraints would require the PAS wait till all PACs issued with conflicting roles expire;

3. *The target application server cannot be sure that user is authorized to use resources.* The privilege attributes present in a PAC, ie: group/role membership and access identity, are not necessarily valid at the point in time that they are used. This is due to the use of the "push model". The role membership may have changed during the PAC's lifetime. To combat this problem, SESAME issues PACs with a short expiry time of the order of a few hours; and

4. *SESAME does not provide support for a role hierarchy.* This is because every target application must implement their own access control logic. For SESAME to support a role hierarchy, every target application would need to understand the hierarchy in order to give the subject the correct permissions. This could be facilitated in SESAME by providing the role hierarchy in a PAC, through the use of user containment relations. For example, $role_1$ contains $role_2$ if all users authorized for $role_1$ are also authorized for $role_2$.

7.4 A Review of Emerging Research

Recent security efforts in ubiquitous computing have been targeted at security in pervasive applications for context-aware homes and offices. While our focus is improving security of network applications through the augmentation of access control processes to include context, emerging work in this field has introduced a number of interesting approaches to securing pervasive applications in the smart home/ office environment.

The first attempt of authorization in the smart-home environment was proposed by Al-Muhtadi et al. [2], who approaches authorization by using a scaled-down version of SESAME (described in the previous section), an extension to Kerberos providing minimal RBAC support. As SESAME is based on traditional network security paradigms and a push model using privilege attribute certificates, it is unable to provide support for context-influenced credentials, nor can it support changes to credentials within a session.

The architecture proposed by Covington et al. [22] [21] introduces a pull based model that supports changes to credentials within a session. Access control policy in this architecture is based on an extension to RBAC, the Generalized RBAC model [20], where object and environment roles are introduced in addition to traditional subject roles. Object roles contain a membership of resources and environmental roles contain a membership of environmental conditions. Access control decisions are then made based on a policy combining environment roles, object roles and subject roles. While GRBAC provides an improvement in the flexibility of security policy, it introduces complexity in administering access control policies and difficulty in ensuring they are conflict free.

The architecture proposed by Al-Muhtadi et al. [3] differs from the architecture proposed by Covington et al. [22] [21], in that it uses first order predicate logic to form its access control policy. While this offers increased flexibility to GRBAC, it is also complex to administer due to the requirement of specifying each access control rule and associated actions individually.

Our approach to access control policy differs in that our work extends RBAC to provide a more flexible activation mechanism for roles, as well as providing role-centric context constraints. This allows for simple access control policy, rather than complex policy definitions that attempt to bind context data to credentials.

7.5 Architecture Goals

The proposed architecture was developed with a number of objectives:

1. Support the use of context information in the access control mechanism, as well as the use of triggers for context changes;
2. Provide support for single sign-on[4];
3. Provide support for core RBAC with hierarchies, static separation of duties and dynamic separation of duties as specified in the proposed RBAC NIST standard [52];
4. Provide support for the specification of context constraints in RBAC roles;
5. Provide context transparency to application servers in order to maintain privacy, but allow for future support for aggregation of context data to the application in a secure controlled manner if needed; and
6. Provide migration for existing network applications by using a standard interface.

These goals were achieved by using an existing network authentication protocol that facilitated single sign-on, and by extending the protocol to support the object-oriented RBAC that we specify in § 7.6. This variation of RBAC provides support for context-awareness and context activation triggers. The following sections detail the RBAC specification, the architecture design and justifications for design decisions.

7.6 Specification of RBAC for a Context-aware Multi-vendor Environment

The following subsections discuss our formal specification of an object-oriented representation of RBAC, which is able to provide support for context-awareness.

[4] Single sign-on (SSO) is mechanism that permits a user to access authorized protected resources, without the need to authenticate for each resource. (i.e. enter multiple passwords).

7.6.1 Object-Z Specification for Object-Oriented RBAC

The use of formal specifications not only allows a system's behavior to be characterized more precisely, but also facilitates precise definition of a system's desired properties. While formal specification can be used to prove that a system meets its specification, formal methods do not prove that a system is correct, nor does it guarantee a system is secure.

The choice of Object-Z [27] as the formal specification was heavily influenced by RBAC, which is well suited to an object-oriented environment. It was deemed appropriate to specify the system in an object-oriented way, rather than functionally. The use of object-oriented specification additionally facilitated the transition from the formal specification to traditional design languages such as UML for prototyping. Kim and Carrington in [66] detail a method for the translation of Object-Z into UML class diagrams. In addition, Johnston and Rose in [65] detail guidelines for the conversion of Object-Z to C++.

Object-Z is an extension to the formal specification language Z, where a state schema and operation schemas are specified for each class. The state schema defines variables and their relationships. The operation schemas define relationships between the pre and post states of the state schema. Object-Z supports object-oriented properties such as object instantiation and object inheritance.

The use of object orientation in specifying RBAC increases its clarity, as it allows designers to focus on single parts of a specification at a time. Additionally, verification and refinement are simplified through the use of composition. Refer to [27] for details of Object-Z.

7.6.2 Object-Oriented RBAC

Appendix C.2 details the formal specification of an object oriented RBAC system which can provide support for context-awareness. This model forms the basis for the proposed architecture. In this model, applications are specified, such that an RBAC system has a set of applications, and each application has a set of roles. Figure 7.4 illustrates the class diagram. We now describe modifications to the object-oriented RBAC model to support context awareness.

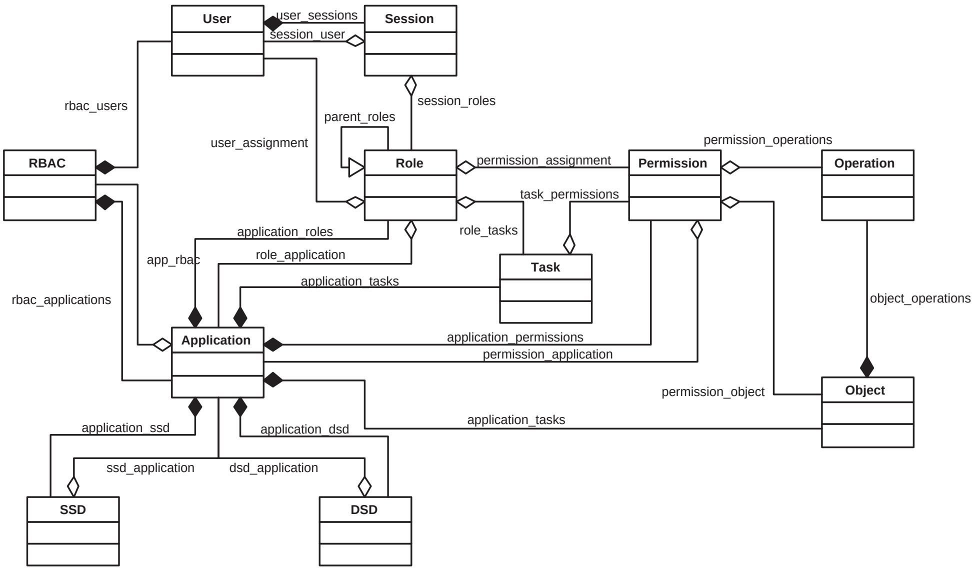


Figure 7.4: Object-Oriented RBAC Class Diagram

7.7 Context-aware Intranet Architecture

The architecture is designed for use in an Intranet environment due to the requirement of application servers having reliable access to the authorization server, and the difficulty of managing access control policy between multiple administrative domains. It is assumed that access to authorization servers over an Internet link would not perform adequately given the required response times and frequency of communications required by the use of centralized permissions and access control logic. The implementation of the architecture operates in the context of a Kerberos (Microsoft Windows 2000) domain.

The proposed architecture has been designed for context-transparency to application servers in that they do not process any context information. For future applications that may have a requirement for certain types of context data, the application server must have activated an appropriate role that allows updates of required type of context data to be requested. This mechanism allows for the centralized storage of user privacy policies which can determine whether such permission is granted to an application server. Figure 7.5 illustrates a component view of the architecture.

The following sections detail the services provided by the architecture.

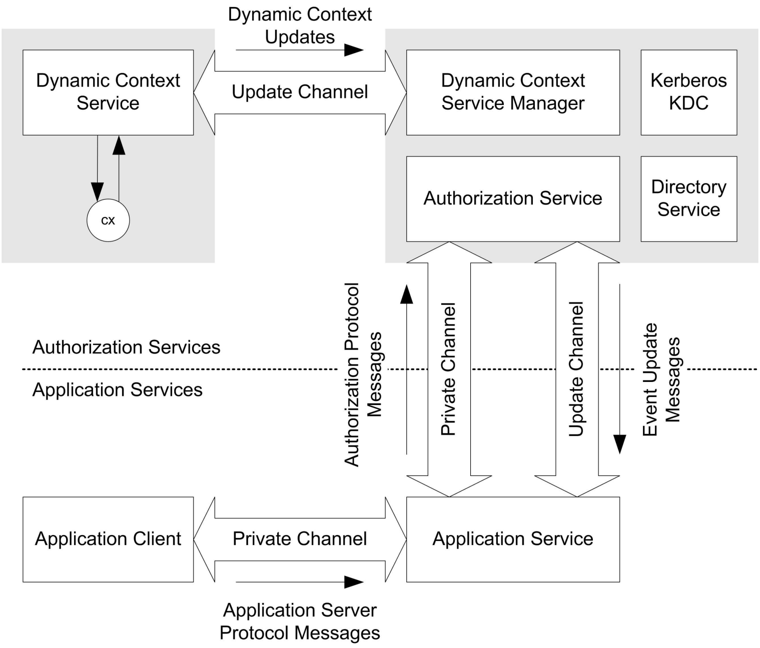


Figure 7.5: Architecture Components

7.7.1 Authentication Service

The authentication service is based on Kerberos [68]. Kerberos authentication is static and session-based, such that when an application ticket is granted, a user is authenticated for the validity duration of the ticket. Depending on the lifetime of the user's ticket granting ticket, a user may be able to renew tickets for application services up till its expiry, which is typically 10 hours.

As user identity information does not dynamically change, Kerberos was deemed an appropriate authentication service, but is not suitable for dynamic access control. It is used solely to authenticate principals to application services, and establish session keys for secure communications. Control of resource access is managed through the proposed on-line authorization protocol, which uses the user's identity from the Kerberos ticket for access control decisions. A ticket can be used to convey pseudonyms instead of a user's real identity, where pseudonyms are resolved by the authorization service, facilitating privacy from application services.

A directory supporting Lightweight Directory Access Protocol (LDAP)[5] provides the data storage for authentication and authorization data, linking Kerberos principals

[5] Refer to RFC2251 Lightweight Directory Access Protocol (v3) `http://www.ietf.org/rfc/rfc2251.txt`.

with roles in the authorization architecture. Kerberos V5 is based on symmetric key cryptography, and is well suited to resource constrained mobile devices. For additional security, devices capable of public key cryptography can perform the initial authentication to the Key Distribution Center (KDC) using public key cryptography as defined in PKINIT [98]. Constrained devices can use a mobile version of PKINIT such as the service proposed by Harbitter [58]. Kerberos is a widely accepted network authentication protocol that is implemented on most Unix variants as well as Microsoft Windows 2000/XP/2003. In addition, support for Kerberos is included in Microsoft Windows CE .NET 4.2. Refer to § 7.8.1 for details of the authentication protocol.

7.7.2 Authorization Service

The role of the authorization service is to provide access control decisions to requests made by application services based on a set of active roles maintained for each application's users by the authorization service.

The Dynamic Context Service Manager (DCSM) is responsible for evaluation and activation of roles in the Authorization service, as illustrated in Figure 7.7. The architecture design segregates the authorization service from the dynamic context service manager, such that the authorization service only makes access control decisions on the current set of active roles for a given user-application context. This independence allows evaluation of events and constraints to operate asynchronously to access control requests.

This significantly increases the performance of the architecture, as access control decisions can be made virtually instantaneously, hence providing a high level of performance to application services. Role activation through mediated principal activation or context triggers is evaluated in an event-driven fashion by the DCSM. The DCSM has a direct RMI-based communications channel that allows the authorization server to request principal activation/deactivation, and allows the DCSM to activate/deactivate roles for a given user-application context. This communications and processing overhead does reduce the performance of principal role activation/deactivation, typically done during a user-application context establishment. It is deemed that moderate delays that are once-off are acceptable.

In this architecture, security context-awareness is modeled through dynamic activation of roles. Standard RBAC roles are extended such that role constraints are evaluated based on activation triggers which may include user request for activation, the activation of other roles, or a context event. Role activation for a given principal depends on the evaluation result of a role's constraints, which are evaluated by the Dynamic Context Service Manager (DCSM). Role activation requests made to the authorization service by the DCSM are only actioned if system-wide constraints, evaluated by the authorization service, are fulfilled.

Activation of roles by a user is mediated by the application service, which performs all role activation/deactivation functions on behalf of the user. The authorization service supports activation and deactivation of roles as requested, but it is ultimately up to the application to provide support for role changes to the user.

7.7.2.1 Implementation of RBAC

The implementation of RBAC in this architecture is an implementation of the Object-Z specification detailed in § 7.6.

The architecture contains two types of roles: a standard role as defined below, and a simplified type of role, the task. A task represents a work function, containing a set of permissions (allowed operations for a given object), and is assigned to a role. For example, an "Administrator" role may include a "AddDomainUser" task. This task contains permissions required to add a domain user. Tasks differ from roles in that they have no members and tasks cannot be assigned to another task, but otherwise have the same attributes as a role. Roles in the architecture are stored in the directory and contain the following attributes:

- *Inherits.* Role inheritance is supported through this attribute, where the role can inherit the members, permissions and constraints from parent roles assuming separation of duties constraints are fulfilled. The attribute contains a list of distinguished names of parent roles in the directory.

- *Members.* This attribute contains the role members, represented as a list of distinguished names of users in the directory.

- *Tasks.* This attribute contains the tasks the role contains, represented as a list of distinguished names of tasks for the given application in the directory.

- *Permissions.* This attribute contains the permissions of the role, represented as a list of distinguished names referencing permissions specified for the given application. A permission contains: (1) Permission Class, the type of the permission represented; (2) Permission Name, the name of the object this permission describes; and (3) Actions, the name of actions granted on the specified object.

- *ActivationTriggers.* This attribute contains the names of supported triggers. The currently supported trigger types include context class, role activation / deactivation, or the principal that requests role activation / deactivation. When an event is fired, the roles containing the corresponding trigger are evaluated. The evaluation considers the constraints before making a decision of whether to activate the role for the principal the event was targeted at.

- *Constraints.* The following attributes represent the supported constraints:

  - *Requires.* This attribute contains the names of the prerequisite roles that must be active in order for this role to be activated, represented as a list of distinguished names.
  - *MutuallyExclusive.* Dynamic separation of duties are implemented as mutual exclusion sets. This attribute contains the name of mutual exclusion sets represented as lists of distinguished names.
  - *ContextRule.* This attribute contains a set of statements that define contextual constraints for role activation.

  ContextRule statements are constructed using the following syntax: `<OBJECT> OPERATION <PARAMETER>`. Boolean operators (and, or, not) can be used between consecutive statements: `(<STATEMENT>) <BOOLEAN_OP> (<STATEMENT>)`. The variables `$PRINCIPAL`, `$CONTEXT(context_class)`, `$LDAP(distinguished_name)` reference the principal, the current principal's context and the directory object referenced by the LDAP distinguished name. An example ContextRule (Figure 7.6) allows the role to be activated only if the principal performed mutual authentication with the application and context location is at the principal's registered home or office location.

```
($CONTEXT(dcaa.dcsm.ConnectionSecurity) hasMutualAuth) and
(($CONTEXT(dcaa.dcsm.Location) within $LDAP(cn=OfficeLocation;cn=...)) or
($CONTEXT(dcaa.dcsm.Location) within $LDAP(cn=HomeLocation;cn=$PRINCIPAL;cn=...)))
```

Figure 7.6: Example ContextRule

7.7.2.2 Authorization Logic

The authorization service maintains the roles for each principal that are active in each application in the domain. For a principal to become active, the application service

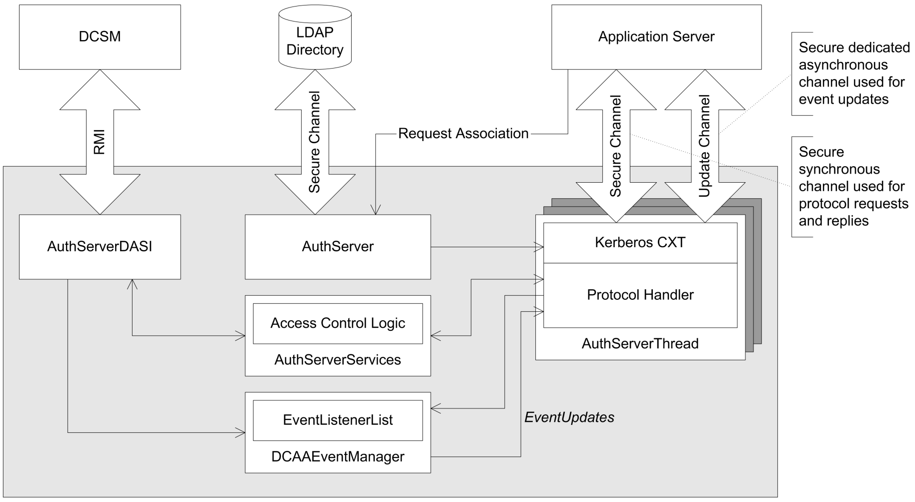


Figure 7.7: Authorization Service

must initialize an access control context with the requested roles on behalf of the user. When a user session ends, the application service ends the context on behalf of the user.

The application service can request an access decision by sending a *checkPermission* request to the authorization service. The *checkPermission* function enumerates all the permissions for the given principal's active roles and tasks. The function then checks that the permission, for which the access decision is requested, is implied by a permission in the set of enumerated permissions. The access control logic is illustrated in Figure 7.8.

```
boolean checkPermission(application, principal, permission) {
   Permissions prms = AuthServer.getPrincipalPermissions(application, principal);
   return prms.implies(permission);
}

Permissions getPrincipalPermissions(application, principal) {
   Permissions prms = new Permissions();
   Role[] roles = AuthServer.getActiveRoles(application, principal)
   for all roles {
      prms.addPermissions(role.getPermissions());
   }
   return prms;
}
```

Figure 7.8: Authorization Service CheckPermission Logic

The implementation has defined a class *DCAAPermission*, a generic permission containing the permission name and actions the permission allows. In addition to these permissions, the implementation supports the use of any permission that extends the *java.security.Permission* class. As such, system, network and runtime permissions can be controlled by the authorization server. These permissions are checked by the access controller, which queries the authorization server, when a privileged piece of code is executed in the access control context of the user using the *doPrivileged()* method. The application service logic is illustrated in Figure 7.9.

```
// Create DCAASubject from Kerberos Principal
DCAASubject subject = new DCAASubject(kerberosPrincipal);

subject.initRemoteAccessControlContext("auth_server@TESTDOMAIN.LOCAL", "testserver.testdomain.local",
   application, roles, new DCAAEventListener());

AccessControlContext acc = subject.getAccessControlContext();

// Request access decision for access control context encapsulated by acc
try {
   acc.checkPermission(permission);
   ...code if granted permission...
} catch (java.security.AccessControlException) {
   ...code if permission denied...
}

// Perform privileged code on behalf of a user’s access control context
somemethod() {
   ...normal code here...
      AccessController.doPrivileged(new PrivilegedAction() {
         public Object run() {
            ...privileged code goes here...
            // checkPermission function as above can be used for checking of arbitrary permissions
         }
      });
   ...normal code here...
}
```

Figure 7.9: Application Service Logic

The *checkPermission* and *doPrivileged* functions throw a *java.security.AccessControlException* if access to the requested resource is denied.

7.7.3 Dynamic Context Service Manager

The Dynamic Context Service Manager (DCSM) is responsible for the activation and deactivation of roles and tasks in the authorization service based on a principal’s context. The DCSM has a context event listener where the triggers defined in roles are registered. The following subsections will provide an overview of its interactions with context services and detail the dynamic context and event update mechanisms.

7.7.4 Dynamic Context Services

Dynamic Context Services (DCS) are trusted services responsible for acquiring context information either directly or via a third party. A DCS notifies the DCSM of context changes based on its policy that specifies how trust and accuracy are quantified given the raw context data, the frequency at which the context is acquired and updated, and the thresholds for notifying the DCSM.

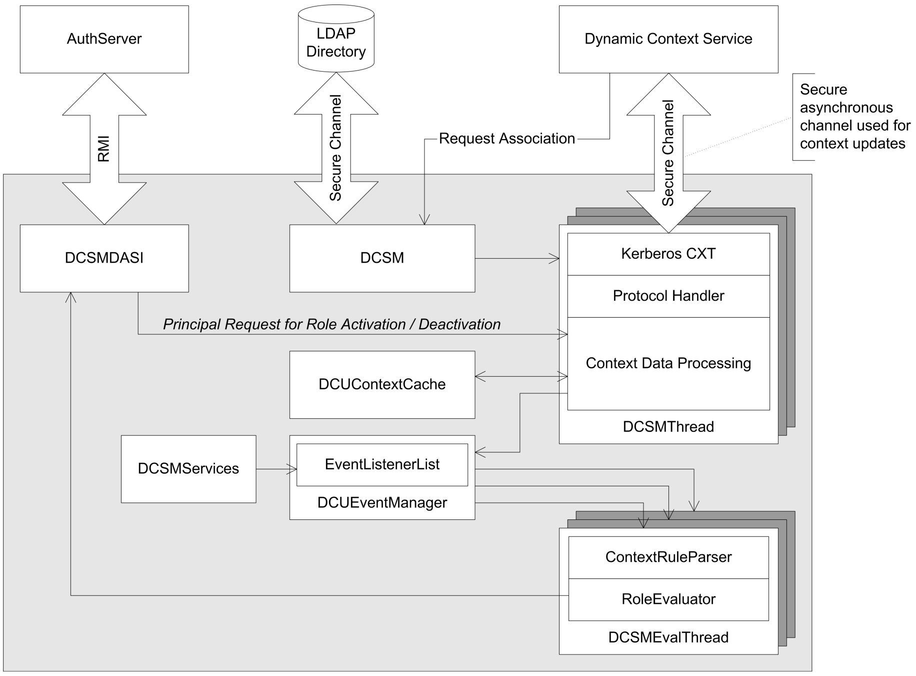


Figure 7.10: Dynamic Context Service Manager

This allows a DCS to collect context information using paradigms other than the event driven model used in the architecture. In addition, context data can be sourced from other context acquisition frameworks,[6] whilst maintaining the trust and security requirements of context acquisition for access control decisions.

For example, we have implemented a GSM location DCS that queries a GSM location service (Gateway Mobile Positioning Center) every 30 seconds for active principals. The DCS maintains a cache of location data. If a context change exceeds the specified threshold, the updated context is communicated to the DCSM in a dynamic context update message containing a common context representation and the principal the update is relevant to. The DCS is responsible for mapping the principal in the context to corresponding principal in the directory. For example, the GSM location service maps the MSISDN[7] of the user to a principal in the directory.

ContextObjects are used to represent context information as well as containing default methods that manage trust, time of last update and accuracy level determined as by the DCS. An example is the more specific instance of ContextObject, LocationContextObject, which contains a common representation of location, and constructors

[6] Such as the Georgia Institute of Technology Context Toolkit `http://www.cc.gatech.edu/fce/contexttoolkit/`

[7] Mobile Subscriber ISDN, the number callers use to reach a mobile subscriber.

that convert a DCS' location data (e.g. WGS-84, lat/long, GSM-Timing Advance arc, etc.) to the common representation. This allows multiple sources of location to have the same representation, such that it can be used by policy and contextRule evaluators. The hierarchy of implemented ContextObjects is illustrated in Figure 7.11.

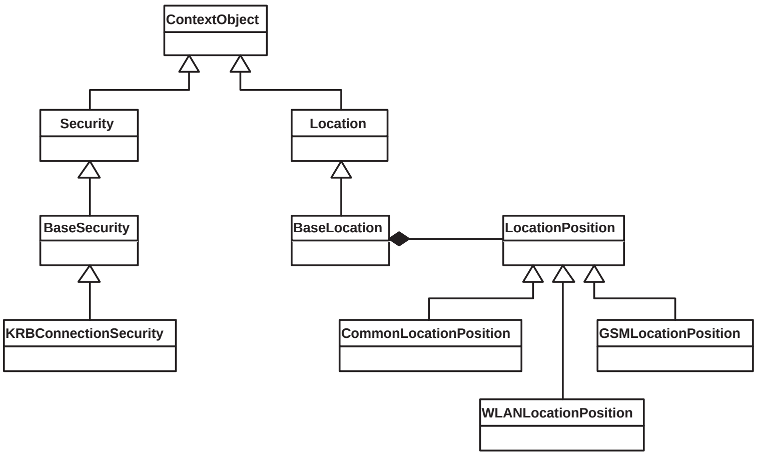


Figure 7.11: Implemented ContextObject Hierarchy

7.7.5 Dynamic Context Update Mechanism

The following processes are executed when a dynamic context update is received from a DCS, or an activation / deactivation request event is received from the authorization service.

When a dynamic context update is received, the last ContextObject for a given principal is retrieved from the context cache. The ContextCombiner method of the new ContextObject (and its subclasses) is then used to create a ContextObject combining the context data from the update and the cache. For example, the combining functionality may use historic data to calculate a more accurate context with the potential to increase trust. Finally, the context cache is updated with the new ContextObject.

An event is fired for the given class of ContextObject, such that all roles or tasks with a trigger of this class are evaluated and activated according to the outcome of the evaluation. In the case of an activation / deactivation request event, the specific role requested is evaluated. Successful activation of a role or task fires the evaluation of any roles or tasks which are triggered based on the activation of that role. Note that tasks will not be evaluated unless a role to which the task is assigned is active for a given principal.

7.7.6 The Event Update Mechanism

The authorization service has an Event Manager that is responsible for informing application services of events that occur. After an application service initiates a secure context with the authorization service, an authorization service-side listener is registered for each application service session. When an event is fired, all event listeners that the event update is relevant to are informed of the event. The event listener for the given application service session sends an event update message over the update channel as shown in Figure 7.5. The application service can then take appropriate action.

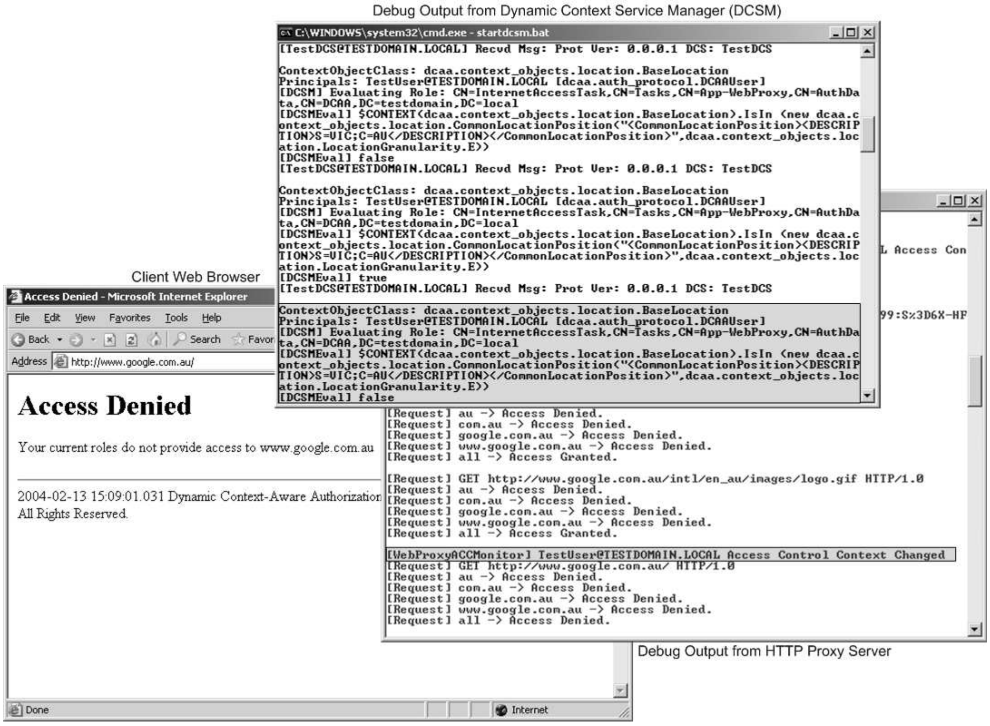
The architecture is designed in such as way as to allow customized events to be implemented. The following events are currently implemented in the architecture:

- *Access Control Context Changed Event.* If a role or task is activated or deactivated for a given principal, the applications in which the principal is active are notified that the access control context for the given principal has changed. The application service can then check that the principals are still authorized to perform an operation by sending a checkPermission request to the authorization service.

- *Dynamic Context Service Manager Failure Event.* In the case there is a failure in the DCSM, the application service is notified and can take evasive action if required.

- *Update Channel Not Responding Event.* In the case of an outage or attack against an update channel, the application service is notified and can take evasive action such as suspending the user's session until the update channel is restored. Failure of the update channel is detected through the use of a heartbeat.

The event update mechanism has been implemented such that future applications that may require context data can be supported through a "Context Data Update Event". An example of an update event can be seen in the prototyped architecture and a webproxy application service illustrated in Figure 7.12. A wireless LAN location DCS (based on the location derivation methods detailed in Chapter 5 provides location proximity information to the DCSM. The proximity is given as a percentage with a threshold defined in a ContextRule of the *InternetAccessTask*. When the user walks out of the room, the DCSM is notified of a context change and the InternetAccessTask is reevaluated. As the user is no longer within the acceptable threshold, the role is deactivated and an "Access Control Context Changed" update is made to the webproxy for the given principal.


Debug Output from Dynamic Context Service Manager (DCSM)
C:\WINDOWS\system32\cmd.exe - startdcsm.bat
Client Web Browser
Access Denied - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favor
Address http://www.google.com.au/
Done
Internet
Debug Output from HTTP Proxy Server

```
[TestDCS@TESTDOMAIN.LOCAL] Recvd Msg: Prot Ver: 0.0.0.1 DCS: TestDCS

ContextObjectClass: dcaa.context_objects.location.BaseLocation
Principals: TestUser@TESTDOMAIN.LOCAL [dcaa.auth_protocol.DCAAUser]
[DCSM] Evaluating Role: CN=InternetAccessTask,CN=Tasks,CN=App-WebProxy,CN=AuthDa
ta,CN=DCAA,DC=testdomain,DC=local
[DCSMEval] $CONTEXT(dcaa.context_objects.location.BaseLocation).IsIn (new dcaa.c
ontext_objects.location.CommonLocationPosition("<CommonLocationPosition><DESCRIP
TION>S=VIC;C=AU</DESCRIPTION></CommonLocationPosition>",dcaa.context_objects.loc
ation.LocationGranularity.E))
[DCSMEval] false
[TestDCS@TESTDOMAIN.LOCAL] Recvd Msg: Prot Ver: 0.0.0.1 DCS: TestDCS

ContextObjectClass: dcaa.context_objects.location.BaseLocation
Principals: TestUser@TESTDOMAIN.LOCAL [dcaa.auth_protocol.DCAAUser]
[DCSM] Evaluating Role: CN=InternetAccessTask,CN=Tasks,CN=App-WebProxy,CN=AuthDa
ta,CN=DCAA,DC=testdomain,DC=local
[DCSMEval] $CONTEXT(dcaa.context_objects.location.BaseLocation).IsIn (new dcaa.c
ontext_objects.location.CommonLocationPosition("<CommonLocationPosition><DESCRIP
TION>S=VIC;C=AU</DESCRIPTION></CommonLocationPosition>",dcaa.context_objects.loc
ation.LocationGranularity.E))
[DCSMEval] true
[TestDCS@TESTDOMAIN.LOCAL] Recvd Msg: Prot Ver: 0.0.0.1 DCS: TestDCS

ContextObjectClass: dcaa.context_objects.location.BaseLocation
Principals: TestUser@TESTDOMAIN.LOCAL [dcaa.auth_protocol.DCAAUser]
[DCSM] Evaluating Role: CN=InternetAccessTask,CN=Tasks,CN=App-WebProxy,CN=AuthDa
ta,CN=DCAA,DC=testdomain,DC=local
[DCSMEval] $CONTEXT(dcaa.context_objects.location.BaseLocation).IsIn (new dcaa.c
ontext_objects.location.CommonLocationPosition("<CommonLocationPosition><DESCRIP
TION>S=VIC;C=AU</DESCRIPTION></CommonLocationPosition>",dcaa.context_objects.loc
ation.LocationGranularity.E))
[DCSMEval] false
```

Access Denied

Your current roles do not provide access to www.google.com.au

2004-02-13 15:09:01.031 Dynamic Context-Aware Authorization
All Rights Reserved.

```
[Request] au -> Access Denied.
[Request] com.au -> Access Denied.
[Request] google.com.au -> Access Denied.
[Request] www.google.com.au -> Access Denied.
[Request] all -> Access Granted.

[Request] GET http://www.google.com.au/intl/en_au/images/logo.gif HTTP/1.0
[Request] au -> Access Denied.
[Request] com.au -> Access Denied.
[Request] google.com.au -> Access Denied.
[Request] www.google.com.au -> Access Denied.
[Request] all -> Access Granted.

[WebProxyACCMonitor] TestUser@TESTDOMAIN.LOCAL Access Control Context Changed
[Request] GET http://www.google.com.au/ HTTP/1.0
[Request] au -> Access Denied.
[Request] com.au -> Access Denied.
[Request] google.com.au -> Access Denied.
[Request] www.google.com.au -> Access Denied.
[Request] all -> Access Denied.
```

Figure 7.12: Triggered Updates

7.8 Architecture Protocols

This section details the protocol messages that are sent in establishment of contexts, performing authorization function, and update messages.

7.8.1 Authentication

Authentication of a client to an application server proceeds as follows. Refer to Appendix B for protocol notation.

The authentication portion of the proposed authentication and authorization protocol proceeds identically to that of Kerberos as shown in Figure 7.13. A client obtains a Ticket Granting Ticket (TGT) by sending a request with an authenticator (1), which the authorization server validates, and if successful, will return a ticket granting ticket

$$
\begin{array}{llll}
C & \rightarrow & AS: & \{TS_C\}_{K_A}, A, TGS, T_s, T_e, N_C \qquad (1) \\
C & \leftarrow & AS: & A, TGT, \{k_{C-TGS}, T_s, T_e, N_c, TGS, C\}_{K_A} \qquad (2) \\
C & \rightarrow & TGS: & \{A, TS'_C\}_{k_{C-TGS}}, TGT, S, T'_s, T'_e, N'_C \qquad (3) \\
C & \leftarrow & TGS: & A, STKT_S, \{k_{C-S}, N'_C, T'_s, T'_e, S, C\}_{k_{C-TGS}} \qquad (4) \\
C & \rightarrow & S: & STKT_S, \{C, TS''_C, SN_{C-S}\}_{k_{C-S}} \qquad (5) \\
C & \leftarrow & S: & \{TS''_C, SN_{C-S}\}_{k_{C-S}} \qquad (6) \\
\vdots & & & \\
C & \rightarrow & S: & \{M^{(i)}, SN_{C-S}\}_{k_{C-S}}, MAC^{(i)}_{k_{C-S}}(M^{(i)}, SN_{C-S}) \qquad (7) \\
C & \leftarrow & S: & \{M^{(i+1)}, SN_{C-S}\}_{k_{C-S}}, MAC^{(i+1)}_{k_{C-S}}(M^{(i+1)}, SN_{C-S}) \qquad (8)
\end{array}
$$

$$
\begin{array}{l}
TGT = TGS, \{k_{C-TGS}, A, T_s, T_e, C\}_{K_{AS-TGS}} \\
STKT_S = S, \{k_{C-S}, A, T'_s, T'_e, C\}_{K_{KDC-S}}
\end{array}
$$

Figure 7.13: Authentication Protocol (Kerberos)

(2) with the key to communicate with the Ticket Granting Service (TGS), k_{C-TGS}, encrypted under the client's long-term key, K_A.

The client uses the TGT to request a ticket for the required application service (S) from the TGS (3). The TGS replies with a ticket for the requested service, containing the key to communicate with S, k_{C-S} (4). Subsequent communications between the client and the server are confidentiality and integrity protected as shown in (7),(8).

The client uses the application service ticket to contact the application for access to desired resources. Control of these resources is mediated through the authorization service. The following subsection will detail the proposed authorization service.

7.8.2 Authorization

The protocol notation detailed below is as defined in Appendix B, with the following additions. U denotes the update port that will be used for the asynchronous update channel, R denotes a role, P denotes a permission which contains a permission class, name and requested actions, I denotes the implication result, such that requested permission, P implies a permission in the user's active set of permissions.

7.8.2.1 Application Service - Authorization Service Context Establishment

An application service must establish a secure Kerberos context with the authorization service before it is able to accept client connections. The application service sends a *SET_UPDATE_PORT* message to the authorization service after the secure context is established (1) as detailed in Figure 7.14. This allows the asynchronous channel to be established, such that update messages can be sent to the application service. S' and AZS' denote the asynchronous communications channel used for update messages.

$$\begin{array}{lllll}
S & \rightarrow & AZS: & \{update\_port, SN_{S-AZS}\}_{k_{S-AZS}}, MAC_{k_{A-AZS}}(update\_port, SN_{S-AZS}) & (1)\\
S' & \leftarrow & AZS': & \{ack, SN_{S-AZS}\}_{k_{S-AZS}}, MAC_{k_{A-AZS}}(ack, SN_{S-AZS}) & (2)
\end{array}$$

Figure 7.14: Set Update Port Message

7.8.2.2 User Access Control Context Establishment

Once a user authenticates and establishes a secure context with an application service, the application service initiates an access control context with the authorization service on behalf of the user. This is done by sending an initiation request to the authorization server containing the user's identity and the set of roles the user wishes to activate (1). The authorization service replies with an acknowledgment that the roles were requested for activation (2). Figure 7.15 details the protocol messages for access control context initiation and checking a user's permissions. When an application service queries the authorization service for an access control decision, the user identity and permission the user requested are sent in a *CHECK_PERMISSION* message (3). The authorization service responds with a decision for the access control request (4).

$$\begin{array}{lllll}
S & \rightarrow & AZS: & \{user, role_{1..n}, SN_{S-AZS}\}_{k_{S-AZS}}, MAC_{k_{A-AZS}}(user, role_{1..n}, SN_{S-AZS}) & (1)\\
S & \leftarrow & AZS: & \{user, ack, SN_{S-AZS}\}_{k_{S-AZS}}, MAC_{k_{A-AZS}}(user, ack, SN_{S-AZS}) & (2)\\
\vdots & & & & \\
S & \rightarrow & AZS: & \{user, permission, SN_{S-AZS}\}_{k_{S-AZS}}, MAC_{k_{A-AZS}}(user, permission, & \\
 & & & SN_{S-AZS}) & (3)\\
S & \leftarrow & AZS: & \{user, permission, ac\_result, SN_{S-AZS}\}_{k_{S-AZS}}, MAC_{k_{A-AZS}}(user, & \\
 & & & permission, ac\_result, SN_{S-AZS}) & (4)
\end{array}$$

Figure 7.15: Access Control Context Establishment and CheckPermission

7.8.2.3 Update Messages

The protocol message as illustrated in message (1), Figure 7.16 shows a single update message sent in the asynchronous channel, where M is the update information for the *ACCESS_CONTROL_CONTEXT_CHANGED*, *DCSM_FAILURE*, and other custom defined events. The *HEATBEAT_UPDATE* message (2) is sent regularly by the

$$\begin{array}{llll} AZS & \rightarrow & S: & \{user, event, SN_{S-AZS}\}_{k_{S-AZS}}, MAC_{k_{A-AZS}}(user, event, SN_{S-AZS}) \quad (1) \\ \vdots & & & \\ AZS & \rightarrow & S: & \{SN_{S-AZS}\}_{k_{S-AZS}}, MAC_{k_{A-AZS}}(SN_{S-AZS}) \quad (2) \end{array}$$

Figure 7.16: Update Messages

authorization service over the asynchronous channel, such that a heartbeat is not received for a predetermined timeout, the application service raises a *CHANNEL_NOT_RESPONDING* event.

7.9 Prototype Implementation of the Architecture

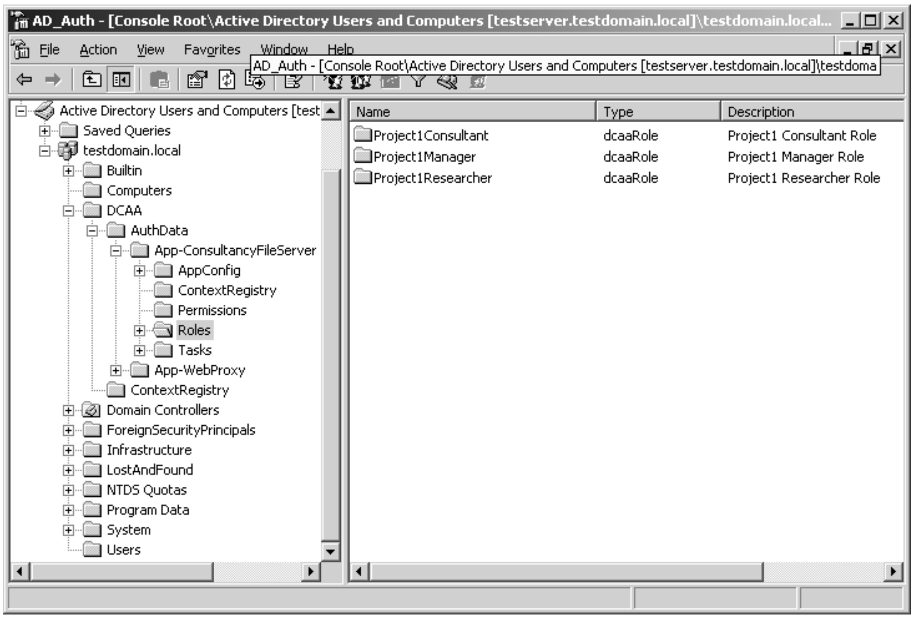
This section will detail the prototype implementation of the architecture, the performance results and an architecture usage scenario.

7.9.1 Test Environment

The architecture is centered around the use of open standards such as Kerberos, LDAP and XML messaging. This allows the architecture to operate over different platforms. Microsoft Windows 2000 Server was used as the platform to host the directory and Kerberos functionality. Figure 7.17 illustrates the Windows management console with the active directory tree containing the applications and their roles, permissions, tasks, etc. used by the prototype. The prototype implementation can also run on Linux with MIT Kerberos and OpenLDAP.

The testing platform, as illustrated in Figure 7.18, was built using a number of Pentium III 833Mhz PCs with 256MB RAM. The authorization data center was configured with Windows 2000 Advanced Server, and hosted the authorization service and DCSM. To date GSM and Wireless LAN location DCSs have been implemented, although the Wireless LAN location tamper resistance is still being investigated. These services were hosted on the context server. The GSM service requires users preregister their MSISDN, stored as an attribute in the User object in the active directory.

The authorization and dynamic context service management services were implemented using Java 1.4.2. The Kerberos implementation in the architecture components uses Java Authentication and Authorization Services (JAAS) for Kerberos authentication and key establishment between architecture entities. The Java security architecture was customized, such that a new Security Manager and Policy class were im-


AD_Auth - [Console Root\Active Directory Users and Computers [testserver.testdomain.local]\testdomain.local...
File Action View Favorites Window Help
AD_Auth - [Console Root\Active Directory Users and Computers [testserver.testdomain.local]\testdoma
Active Directory Users and Computers [test
Saved Queries
testdomain.local
Builtin
Computers
DCAA
AuthData
App-ConsultancyFileServer
AppConfig
ContextRegistry
Permissions
Roles
Tasks
App-WebProxy
ContextRegistry
Domain Controllers
ForeignSecurityPrincipals
Infrastructure
LostAndFound
NTDS Quotas
Program Data
System
Users

Name	Type	Description
Project1Consultant	dcaaRole	Project1 Consultant Role
Project1Manager	dcaaRole	Project1 Manager Role
Project1Researcher	dcaaRole	Project1 Researcher Role

Figure 7.17: Management Console

plemented. The new classes facilitate the initialization of a secure Kerberos context for the standard communications channel and update channel with the authorization server, subsequently allowing XML messages to be sent and received in these channels. This approach allows standard Java Security methods such as *doAsPrivileged()*, *checkPermission()* and standard access control context constructs to be used, resulting in transparent use of the authorization architecture. An additional listener class was implemented, such that event listener methods would be instantiated on events such as “Access Control Context Changed”.

7.9.2 Prototype Performance

A benchmark of the architecture performance was conducted to quantify the efficiency of the architecture. The performance results illustrated in Figures 7.19,7.20,7.21 are average execution times for 1000 consecutive executions of *checkPermission()* for each application service running for an active principal. These figures have two data sets, the data set with the longer times being the remote requests taking into account network latency, and the shorter times being local requests. Figure 7.22 compares the execution of *checkPermission()* for 1,5,and 10 application services used by up to 1000 principals.

The results detailed in Table 7.1 illustrate the performance for dynamic context updates that were continually sent to the DCSM for a principal with an established

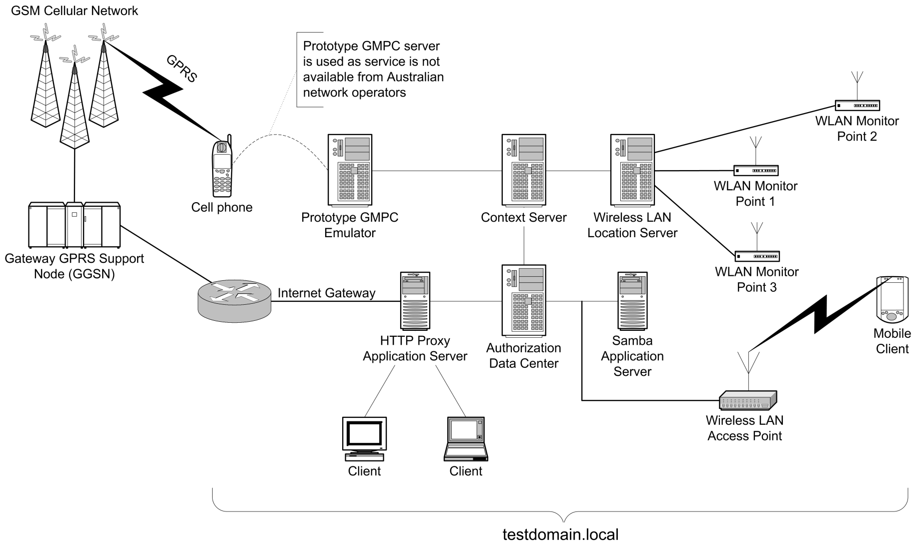

GSM Cellular Network
GPRS
Prototype GMPC server is used as service is not available from Australian network operators
Cell phone
Prototype GMPC Emulator
Context Server
Wireless LAN Location Server
WLAN Monitor Point 2
WLAN Monitor Point 1
WLAN Monitor Point 3
Gateway GPRS Support Node (GGSN)
Internet Gateway
HTTP Proxy Application Server
Authorization Data Center
Samba Application Server
Mobile Client
Wireless LAN Access Point
Client
Client
testdomain.local

Figure 7.18: System Architecture Test Platform

context in each application service. The benchmark was conducted with 1, 5, and 10 simultaneous application services, requesting an access control decision continuously, simulating high load. The dynamic context updates were benchmarked with a load of 10, 100 and 1000 active principals over 1, 5, and 10 application services.

Dynamic Context Updates		
10 Principals (ms)	**100 Principals (ms)**	**1000 Principals (ms)**
328.55	448.55	1079.55
341.61	461.61	1092.61
571.13	691.13	1322.13

Table 7.1: Performance Results

The performance results detailed in Table 7.1 illustrate that the architecture is able to scale well with better than linear growth for context updates as the number of active principals increases. This indicates that the dynamic authorization paradigm presented in this chapter can provide context-based authorization with acceptable performance. The authorization service and DCSM are additionally designed such that they can be distributed to improve performance for large domains.

7.9.3 Architecture Usage Scenario

A prototype fileserver was developed to illustrate possible uses of the architecture. The fileserver is based on Samba 3.0, which supports Kerberos authentication. The

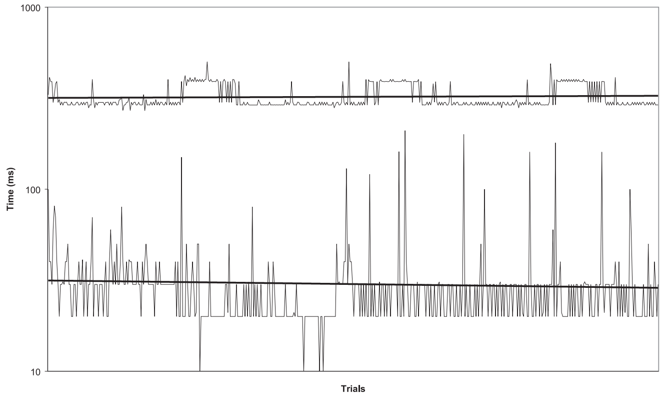


Figure 7.19: *CheckPermission*() Performance for 1 Application

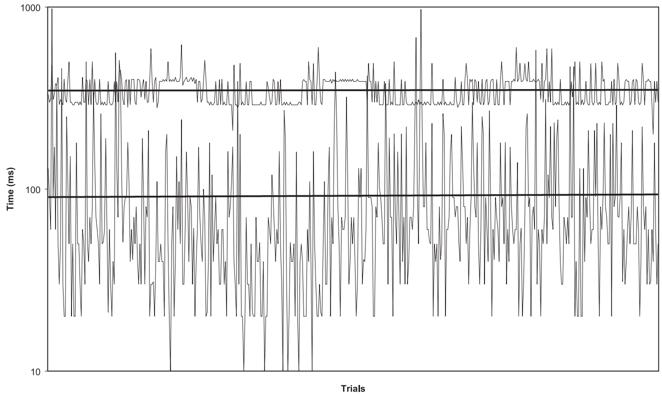


Figure 7.20: *CheckPermission*() Performance for 5 Applications

following scenario illustrates the context-aware architecture in the use of a fileserver that supports the requirements of “commercial in confidence” file access.

There are three DCS’ that are used in this scenario:

1. *GSM Location DCS.* This DCS acquires a trusted location from a GSM cell phone using the methods we have developed in Chapter 4. The DCS authenticates the user and confirms that the user is associated with the cell phone.

2. *iButton*[8] *Location DCS* This DCS provides the location of a user associated with an authenticated iButton based on the known location of a given iButton reader.

[8] iButton®. is a tamper-resistant token that has a unique identifier and may provide support for cryptography. `http://www.ibutton.com`

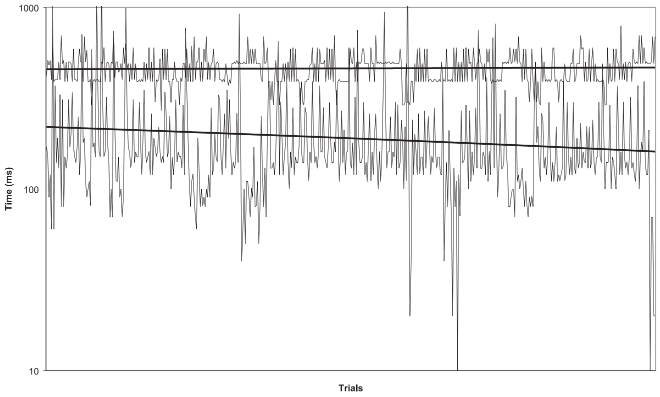


Figure 7.21: *CheckPermission*() Performance for 10 Applications

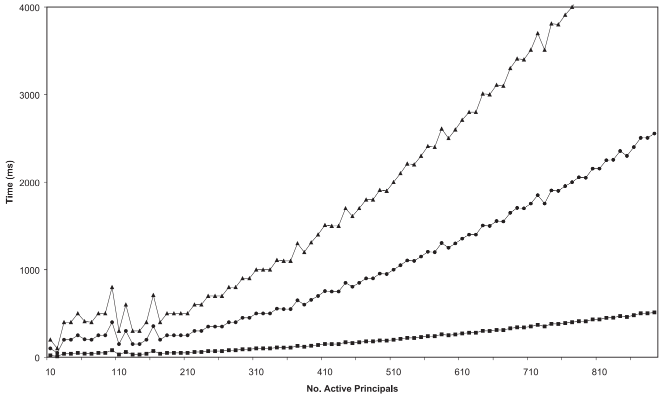


Figure 7.22: *CheckPermission*() Performance comparison of 1 to 1000 principals

3. *Kerberos Connection Security DCS.* This DCS is a component of the fileserver which communicates the properties of the authentication and subsequent communications to the DCSM. The properties supported by the DCS are `hasMutualAuth`, `hasIntegrity`, and `hasPrivacy`.

Commercial in confidence projects have a project manager and a series of consultants who occupy the roles `project1_manager` and `project1_consultant`, which are triggered by principal activation. The fileserver has a shared area for each project and a personal area for each consultant. Consultants are only permitted to work at the office or home where appropriate security arrangements exist. The home location of a consultant is registered in the directory.

The `project1_consultant` role has a number of tasks relating to file and printer access, one of which is `project1_filesystem_access`. This task contains permissions for accessing the project share and the consultant's personal share. The task has triggers of the dcaa.dcsm.Location context class and activation of the `project1_consultant` role. In order to ensure the shares are only accessed in approved locations and with appropriate connection security, the task contains the ContextRule illustrated in Figure 7.23.

```
(($CONTEXT(dcaa.dcsm.ConnectionSecurity) hasMutualAuth) and
($CONTEXT(dcaa.dcsm.ConnectionSecurity) hasIntegrity)) and
(($CONTEXT(dcaa.dcsm.Location) within $LDAP(cn=OfficeLocation;cn=...)) or
($CONTEXT(dcaa.dcsm.Location) within $LDAP(cn=HomeLocation;cn=$PRINCIPAL;cn=...)))
```

Figure 7.23: Example ContextRule

A user wanting to work on "Project1" in the office would tap their iButton at the office door on entry and exit, causing a ContextUpdate with a dcaa.dcsm.Location object to be sent to the DCSM. When the user attempts to access the share, a Kerberos ticket is presented to the fileserver. The fileserver initiates an access control context with the authorization service and requests the role of `project1_consultant`. Assuming the user is a member of the role and no constraints prohibit the activation of the role, the role will be activated. As activation of the role triggers evaluation of the `project1_filesystem_access` task, it too will be activated assuming its constraints are fulfilled.

A remote user would have to prove their location via the GSM location DCS. If a user is to leave a trusted location area, the task will be deactivated following the appropriate DCS notifications. Similarly with the fileserver ConnectionSecurity DCS, it notifies the DCSM of the established Kerberos context security properties.

7.10 Summary

In conclusion, this chapter has introduced a new authorization architecture for Intranet environments that supports context-aware authorization using both local and remote security contexts. We implement extensions to RBAC that facilitate efficient context-aware authorization with simple policies and administration. The implementation of the architecture has been described with the currently implemented dynamic context services as well as the description of a demonstration application that utilizes the architecture.

Chapter 8

Conclusions and Future Research

In this thesis we have examined location services and their use in IP and application-layer access control processes. In Chapter 2, a set of location acquisition models that generalize common location technologies were presented. The security properties of the model components were investigated. These properties were used to establish a set of requirements that trusted location systems should exhibit. A taxonomy of attacks against these common components was presented, based on known attacks against common location technologies. The taxonomy and requirements were used to classify the trust of a number of existing location technologies.

Based on identified vulnerabilities of differential GPS, and an opportunity to provide a solution, an authentication and integrity augmentation to differential GPS was proposed in Chapter 3. The vulnerabilities of differential GPS broadcasts are discussed in more detail, the first known discussion of such vulnerabilities. DGPS vulnerabilities are a significant concern, as they may have implications on safety-critical marine operations. A solution is proposed, augmenting the existing message broadcast protocol with an authentication and integrity scheme that attempts to mitigate the vulnerabilities identified.

In an attempt to engineer trusted location for access control systems, we propose a scheme for providing tamper-resistant location acquisition in GSM that can be used in security services in Chapter 4. The proposal is an entire location scheme, including the geographical representations for trusted location and an association protocol that attempts to mitigate the possibility of a disassociation attack.

The use of location context data at the IP-layer is introduced in Chapter 5, where we presented our proposal for proximity-based packet filtering for wireless LANs. Proximity-based packet filtering can be used to restrict users from accessing a network from an unauthorized location, requiring that users are within a predefined location area.

During the development of the wireless LAN location system, a denial of service exploit was discovered, resulting in an unanticipated research contribution detailed in Chapter 6. The exploit is significant as it can be achieved using commercial off the shelf hardware and software; has low power requirements; and can be executed with minimal chance of detection and localization.

Chapter 7 introduced context-awareness for access control at the application layer. A new authorization architecture was proposed based on the Kerberos authentication service, providing support for context-awareness in access control policy, and dynamic authorization where role activation could be triggered by context events. The context-awareness was modeled in RBAC, where the proposed RBAC model was specified in Object-Z. The architecture was prototyped and results were given detailing the performance of the architecture.

A number of future directions for research have emerged from the research performed in this thesis. The analysis in Chapter 2 indicated that research is needed to address the lack of authenticity and integrity in GPS data-based augmentation systems, and methods to integrate emerging signal integrity techniques in Galileo with these augmentations, such that evidence of signaling trust and augmentation data trust can be provided to a third party in addition to the resulting location data.

In Chapter 6 a number of strategies to mitigate the Wireless LAN denial of service attack were discussed. These strategies need to be further investigated, in particular the need for developing highly available wireless networks that are suitable for safety-critical environments.

The concept of context-aware security introduced in this thesis is a new concept and requires significant future research. In order to support effective context-aware security, host context information must be obtained in a secure manner. Further investigation of trusted computing methods to facilitate trusted host security context acquisition is required. Additionally, future work may include developing support for cross-domain context-awareness in access control policy.

Lastly, further development of context services and migration of existing network applications to support the proposed authorization extensions is required in order to further develop and mature the proposed context-aware access control architecture.

Appendix A

An Introduction to the TESLA Protocol

Time Efficient Stream Loss Tolerant Authentication (TESLA) is a multicast authentication protocol proposed by Canetti et al. [80]. TESLA uses Message Authentication Codes (MAC) to achieve integrity of broadcast messages. The advantage of using MACs is the reduction in computation and communications overhead compared to the use of asymmetric cryptography. It is additionally scalable to large number of receivers. The protocol provides authentication and integrity of the broadcast messages and is currently an IETF standard. The protocol assumes that the sender and the receiver have a weak time synchronization.

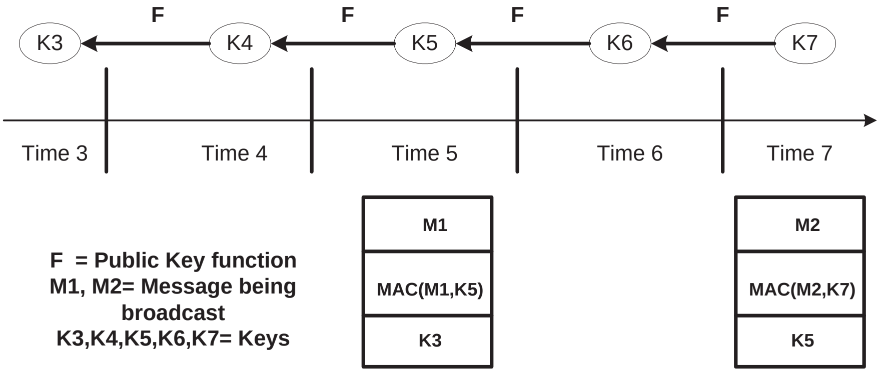


Figure A.1: The TESLA Protocol

The protocol, as illustrated in Figure A.1, is described as follows:

1. *Initialization:*

   Time is synchronized between the sender and receivers and the disclosure delay is agreed between the receiver and sender. The key chain for generation of a one time hash is subsequently committed.

2. *Sender:*

   For each time slot t_i there exists a key K_i, which expires as the time slot expires. Suppose a message M_1 needs to be broadcast to the receiver at time slot 5 and the agreed time delay is 2 timeslots, then the package P_1 contain the Message $M_1, MAC(M_1, K_5)$ and the key K_3. Similarly a message M_2 that needs to be broadcast at time slot 7 is packaged as P_2 containing, $M_2, MAC(M_2, K_7)$, and K_5.

3. *Receiver:*

   The receiver receives the package P_1 and stores $M_1, MAC(M_1, K_5)$. Package P_2 containing M_2 is received after a gap of two time slots. The receiver stores the message $M_2, MAC(M_2, K_7)$. The receiver then uses the key K_5 released in P_2 to compute $MAC'(M_1, K_5)$. The receiver verifies that $MAC'(M_1, K_5)$ is equal to $MAC(M_1, K_5)$ received by the receiver in package P1 at time slot T_5. If these two MAC are equal then the integrity of the message M_1 is assured.

Depending on the nature of the application where the broadcast is taking place, the TESLA protocol can be modified to suit the new environment. μTESLA is a modified version of the TESLA protocol for sensor networks requiring smaller communication overheads compared to traditional TESLA. μTESLA and Secure Network Encryption Protocol (SNEP) were proposed by Perrig et al. in [79] as part of secure protocols for sensor networks. In the following section, we introduce our proposed variation of the TESLA protocol and its application to DGPS.

Appendix B

Protocol Notation

Entities	
A	User
C	Client
AS	Authentication Service
TGS	Ticket Granting Service
$PTGS$	Privilege Ticket Granting Service
$DPAS$	Dynamic Privilege Attribute Service
KDC	Key Distribution Center (encapsulates AS, TGS, PTGS, and DPAS functionality)
S	Application Service
Keys	
K_{i-j}	Long term key shared between i and j
K_i	Key derived from password of user
k_{i-j}	Session key generated for use between i and j
Operations	
$A \rightarrow B / A \leftarrow B$	Protocol message sent in direction of arrow between entity A and B
$\{m\}_{K_{i-j}}$	Denotes encryption of message m using key K_{i-j}
Elements	
$A \mid B$	Entity A or B
TGT	Ticket granting ticket
$PTGT$	Privilege TGT
$STKT_i$	Service ticket for service i
$PTKT$	Privilege ticket
$PAC_{i,j}$	Privilege attribute certificate containing privileges of i for service j
$UPAC_{i,j}$	Update PAC containing privileges of i for service j
E'	Indicates a new instance of element E
TS_i	Timestamp created by i
N_i	Nonce generated by i
SN_{i-j}	Protocol sequence number used by i and j
T_s	Start time of a ticket
T_e	End time of a ticket
TP_r	Transited Path to this realm r

Table B.1: Protocol Notation

Appendix C

Object-Z Specifications

C.1 Brief Overview of Object-Z

Z schemas are used to describe both static and dynamic aspects of a system including its state, invariant relationships, changes that occur to its state, the system operations and relationships between the inputs and outputs of operations. The schemas are specified using predicate logic, such that system functionality can be abstracted and its behavior reasoned about effectively without dependence on platform or program code.

Object-Z is a variation of Z that supports object orientation. Object orientation is characterized by a modular design methodology that is composed of a collection of interacting objects. Objects are abstractions of real-world entities with a state, behavior and identity.

In Object-Z, a state schema with associated operations constitutes the definition of a class. A class is a template for objects as well as a type, which facilitates object referencing through type instances. An Object-Z specification is composed of a number of class definitions that may be related by inheritance.

A class schema has the following definitions[1]:

1. *Visibility list* A class' interface is defined by a visibility list, denoted by $\upharpoonright(...)$. No visibility list results in all attributes and operations being visible;

2. *Type and constant definitions* Type and constant definitions are the same as in Z;

[1] Refer to [90] for an introduction to the Object-Z specification language.

3. *State schema* The state schema is defined in the same way as Z, except that it is nameless;

4. *Initial state schema* The initial state schema is identified by, INIT, composed only of the predicate part assuming declarations of the state schema; and

5. *Operations* The operations may include a list of the state variables that are changed by the operation, $\Delta(...)$. Operations in Object-Z can be combined with other operations using:

   - Composition, where $\parallel$ defines parallel composition, $\fatsemi$ defines sequential composition, and $\parallel_!$ defines associative parallel composition;
   - Conjunction, where $\bigwedge$ defines operation conjunction;
   - Choice, where $[]$ defines angelic choice (similar to $\vee$ in Z); and
   - Scope enrichment, where $\bullet$ defines scope enrichment.

When the class is used as a type, the class represents a set of object identities which uniquely identify objects of that class. Visible attributes of an object can be accessed using the dot notation, $\mathit{obj}.\mathit{attrib}$. Visible operations can be applied to an object using the dot notation, $\mathit{obj}.\mathit{Operation}$.

Inheritance is supported in Object-Z by specification of the parent class names after the visibility list. Parent class signatures must be type compatible with the derived class. An object of a parent class or one of its derivatives is referenced using the notation, $\downarrow\mathit{ParentClass}$, facilitating polymorphism.

C.2 Specification for Object-Oriented RBAC

The following Object-Z schema specifies the proposed RBAC implementation.

The specification is introduced with the specification of an *Operation* type.

$[\mathit{Operation}]$

C.2.1 Object Class Schema

Object

$operations : \mathbb{P}\, Operation_{\copyright};$

INIT

$operations = \varnothing$

CreateOperation

$\Delta(operations)$

$operation? : Operation$

$operation? \notin operations$

$operations' = operations \cup \{operation?\}$

RemoveOperation

$\Delta(operations)$

$operation? : Operation$

$operation? \notin operations$

$operations' = operations \setminus \{operation?\}$

C.2.2 Permission Class Schema

Permission

$object : Object$

$operations : \mathbb{P}\,Operation$

$application : Application$

INIT

$operations = \varnothing$

SetObject

$\Delta(object)$

$object? : Object$

$object? \in application.objects$

$object' = object?$

AddOperation

$\Delta(operations)$

$operation? : Operation$

$operation? \notin operations$

$operation? \in object.operations$

$operations' = operations \cup \{operation?\}$

RemoveOperation

$\Delta(operations)$

$operation? : Operation$

$operation? \in operations$

$operations' = operations \setminus \{operation?\}$

C.2.3 Application Class Schema

Application

$objects : \mathbb{P}\, Object_{©}$

$permissions : \mathbb{P}\, Permission_{©}$

$roles : \mathbb{P}\, Role_{©}$

$ssd\_sets : \mathbb{P}\, SSD_{©}$

$dsd\_sets : \mathbb{P}\, DSD_{©}$

$rbac : RBAC$

$\forall p : Permission \mid p \in permissions \bullet p.application = self$

$\forall r : Role \mid r \in Roles \bullet r.application = self$

$\forall ssd : SSD \mid ssd \in ssd\_sets \bullet ssd.application = self$

$\forall dsd : DSD \mid dsd \in dsd\_sets \bullet dsd.application = self$

INIT

$objects = \varnothing$

$permissions = \varnothing$

$roles = \varnothing$

CreateObject

$\Delta(objects)$

$object? : Object$

$object? \notin objects$

$objects' = objects \cup \{object?\}$

DeleteObject

$\Delta(objects)$

$object? : Object$

$object? \in objects$

$objects' = objects \setminus \{object?\}$

CreatePermission

$\Delta(\mathit{permissions})$

$\mathit{permission}? : \mathit{Permission}$

$\mathit{permission}? \notin \mathit{permissions}$

$\mathit{permissions}' = \mathit{permissions} \cup \{\mathit{permission}?\}$

DeletePermission

$\Delta(\mathit{permissions})$

$\mathit{permission}? : \mathit{Permission}$

$\mathit{permission}? \in \mathit{permissions}$

$\mathit{permissions}' = \mathit{permissions} \setminus \{\mathit{permission}?\}$

AddRole

$\Delta(\mathit{roles})$

$\mathit{role}? : \mathit{Role}$

$\mathit{role}? \notin \mathit{roles}$

$\mathit{roles}' = \mathit{roles} \cup \{\mathit{role}?\}$

RemoveRole

$\Delta(\mathit{roles})$

$\mathit{role}? : \mathit{Role}$

$\mathit{role}? \in \mathit{roles}$

$\mathit{roles}' = \mathit{roles} \setminus \{\mathit{role}?\}$

CreateSSDSet

$\Delta(\mathit{ssd\_sets})$

$\mathit{ssd}? : \mathit{SSD}$

$\mathit{ssd}? \notin \mathit{ssd\_sets}$

$\mathit{ssd\_sets}' = \mathit{ssd\_sets} \cup \{\mathit{ssd}?\}$

DeleteSSDSet

$\Delta(ssd\_sets)$

$ssd? : SSD$

$ssd? \in ssd\_sets$

$ssd\_sets' = ssd\_sets \setminus \{ssd?\}$

CreateDSDSet

$\Delta(dsd\_sets)$

$dsd? : DSD$

$dsd? \notin dsd\_sets$

$dsd\_sets' = dsd\_sets \cup \{dsd?\}$

DeleteDSDSet

$\Delta(dsd\_sets)$

$dsd? : DSD$

$dsd? \in dsd\_sets$

$dsd\_sets' = dsd\_sets \setminus \{dsd?\}$

GetUserRoleSessions

$role? : Role$

$user? : User$

$sessions! : \mathbb{P}\, Session$

$$sessions! = \{s : Session \mid (s \in user?.sessions) \\ \quad \land (role? \in s.active\_roles)\}$$

$$DeleteRole \mathrel{\widehat=} DeleteSessions \ \bigwedge \ RemoveRole$$

$$DeleteSessions \mathrel{\widehat=} \bigwedge u : rbac.users \bullet \\ \quad \left[GetUserRoleSessions\left[sessions/sessions!\right] \middle| user? = u\right] \bullet \\ \quad \bigwedge s : sessions \left[u.DeleteSession \middle| session? = s\right]$$

In this implementation of *DeleteRole*, all sessions containing an active role to be deleted are terminated.

C.2.4 User Class Schema

User

$assigned\_roles : \mathbb{P}\, Role$

$effective\_roles : \mathbb{P}\, Role$

$authorized\_roles : \mathbb{P}\, Role$

$effective\_permissions : \mathbb{P}\, Permission$

$assigned\_roles = \{r : Role \mid self \in r.users\}$

$effective\_roles = \bigcup\{er : \mathbb{P}\, Role \mid \forall\, r : Role \bullet$
$\quad (r \in assigned\_roles) \land (er = (r.inherited\_roles \cup \{r\}))\}$

$authorized\_roles = \bigcup\{r : \mathbb{P}\, Role \mid \forall\, s : Session \bullet$
$\quad (s \in self\ .sessions) \land (r = s.active\_roles)\}$

$effective\_permissions = \bigcup\{p : \mathbb{P}\, Permission \mid \forall\, r : Role \bullet$
$\quad (r \in effective\_roles) \land (p = r.role\_permissions)\}$

$sessions : \mathbb{P}\, Session_{\copyright}$

$\forall\, s : Session \mid s \in sessions \bullet s.user = self$

INIT

$sessions = \varnothing$

CreateSession

$\Delta(sessions)$

$session? : Session$

$session? \notin sessions$

$sessions' = sessions \cup \{session?\}$

DeleteSession

$\Delta(sessions)$

$session? : Session$

$session? \in sessions$

$sessions' = sessions \setminus \{session?\}$

C.2.5 Session Class Schema

$Session$

$effective\_permissions : \mathbb{P}\, Permission;$

$effective\_active\_roles : \mathbb{P}\, Role$

$effective\_active\_roles = \bigcup$

$\quad \{rs : \mathbb{P}\, Role \mid \forall\, r : Role \bullet$

$\quad (r \in self\ .active\_roles) \wedge (rs = r.inherited\_roles \cup \{r\})\}$

$effective\_permissions = \bigcup \{ps : \mathbb{P}\, Permission \mid \forall\, r : Role \bullet$

$\quad (r \in effective\_active\_roles) \wedge (ps = r.role\_permissions)\}$

$active\_roles : \mathbb{P}\, Role;$

$user : User;$

$INIT$

$active\_roles = \varnothing$

$AddActiveRole$

$\Delta(active\_roles)$

$role? : Role$

$application? : Application$

$role? \notin active\_roles$

$role? \in application?.roles$

$user \in role?.users$

$\forall\, d : DSD \mid (d \in application?.dsd\_sets) \wedge (role? \in d.roles) \bullet$

$\quad d.cardinality \geq (\#\{r : Role \mid r \in ((effective\_active\_roles \cup$

$\quad \{role?\}) \cap d.roles)\})$

$active\_roles' = active\_roles \cup \{role?\}$

DropActiveRole

$\Delta(active\_roles)$

$role? : Role$

$role? \in active\_roles$

$active\_roles' = active\_roles \setminus \{role?\}$

CheckAccess

$operation? : Operation$

$object? : Object$

$operation? \in object?.operations$

$\forall\, r : Role \mid r \in roles \bullet$

$\quad operation? \in r.application.objects$

$\exists\, p : Permission \mid p \in effective\_permissions \bullet$

$\quad (object? = p.object) \land (operation? \in p.operations)$

C.2.6 Role Class Schema

Role

$assigned\_users : \mathbb{P}\, User$

$inherited\_roles : \mathbb{P}\, Role$

$descended\_roles : \mathbb{P}\, Role$

$inherited\_permissions : \mathbb{P}\, Permission$

$assigned\_users = self\,.users$

$inherited\_roles = \bigcup \{ir : \mathbb{P}\, Role \mid \forall\, r : Role \bullet$

$\quad (r \in self\,.parent\_roles) \land (ir =$

$\quad (r.inherited\_roles \cup \{r\}))\}$

$descended\_roles = \bigcup \{dr : \mathbb{P}\, Role \mid \forall\, r : Role \bullet$

$\quad (self \in r.parent\_roles) \land (dr =$

$\quad (\{r\} \cup r.descended\_roles))\}$

$inherited\_permissions = \bigcup \{ir : \mathbb{P}\, Role \mid \forall\, r : Role \bullet$

$\quad (r \in (self\,.inherited\_roles \cup \{self\})) \land$

$\quad (ir = r.permissions)\}$

$role\_permissions = inherited\_permissions \cup self\,.permissions$

$authorized\_users = \bigcup \{au : \mathbb{P}\, User \mid \forall\, r : Role \bullet$

$\quad (r \in self\,.descended\_roles) \land (au = r.users)\}$

$users : \mathbb{P}\, User;$

$permissions : \mathbb{P}\, Permission;$

$parent\_roles : \mathbb{P}\, Role;$

$application : Application$

INIT

$users = \varnothing$

$permissions = \varnothing$

$parent\_roles = \varnothing$

AssignUser

$$\begin{array}{|l}
\Delta(users) \\
user? : User \\
\hline
user? \notin users \\
user? \in application.rbac.users \\
\forall s : SSD \mid self \in s.roles \bullet \\
\qquad s.cardinality \geq (\#\{r : Role \mid r \in ((user?.effective\_roles \cup \{self\}) \cap s.roles)\}) \\
users' = users \cup \{user?\} \\
\hline
\end{array}$$

DeassignUser

$$\begin{array}{|l}
\Delta(users) \\
user? : User \\
\hline
user? \in users \\
user? \in application.rbac.users \\
\forall s : Session \mid s \in user?.sessions \bullet \\
\qquad s.active\_roles' = s.active\_roles \setminus \{self\} \\
users' = users \setminus \{user?\} \\
\hline
\end{array}$$

GrantPermission

$$\begin{array}{|l}
\Delta(permissions) \\
permission? : Permission \\
\hline
permission? \notin permissions \\
permission? \in application.permissions \\
permissions' = permissions \cup \{permission?\} \\
\hline
\end{array}$$

RevokePermission

$$\begin{array}{|l}
\Delta(permissions) \\
permission? : Permission \\
\hline
permission? \in permissions \\
permission? \in application.permissions \\
permissions' = permissions \setminus \{permission?\} \\
\hline
\end{array}$$

AddParentRole

$\Delta(parent\_roles)$

$role? : Role$

$role? \notin parent\_roles$

$role? \in application.roles$

$role? \notin self\ .inherited\_roles$

$\forall s : SSD \mid self \in s.roles \bullet$

$\quad s.cardinality \geq (\#\{r : Role \mid r \in ((self\ .inherited\_roles \cup$

$\quad \{self\} \cup \{role?\}) \cap s.roles)\})$

$parent\_roles' = parent\_roles \cup \{role?\}$

RemoveParentRole

$\Delta(parent\_roles)$

$role? : Role$

$role? \in parent\_roles$

$role? \in application.roles$

$parent\_roles' = parent\_roles \setminus \{role?\}$

C.2.7 Static Separation of Duties Class Schema

SSD

$roles : \mathbb{P}\, Role$

$cardinality : \mathbb{N}$

$application : Application$

$(cardinality = 0) \lor (cardinality \geq 2)$

$cardinality \leq \# roles$

$INIT$

$roles = \varnothing$

$cardinality = 0$

$AddSSDRoleMember$

$\Delta(roles)$

$role? : Role$

$role? \notin roles$

$role? \in application.roles$

$$
\begin{array}{l}
\forall\, s : SSD \mid s \in application.ssd\_sets \bullet \\
\quad (s.cardinality \geq (\#\{r : Role \mid r \in (role?.descended\_roles \cap \\
\quad (s.roles \cup \{role?\}))\})) \land \\
\quad (\forall\, u : User \mid u \in application?.rbac.users \bullet \\
\qquad s.cardinality \geq (\#\{r : Role \mid r \in (u.effective\_roles \cap \\
\qquad (s.roles \cup \{role?\}))\}))
\end{array}
$$

$roles' = roles \cup \{role?\}$

$RemoveSSDRoleMember$

$\Delta(roles)$

$role? : Role$

$role? \in roles$

$roles' = roles \setminus \{role?\}$

SetSSDCardinality

$\Delta(\mathit{cardinality})$

$\mathit{cardinality}? : \mathbb{N}$

$\mathit{cardinality}' = \mathit{cardinality}?$

C.2.8 Dynamic Separation of Duties Class Schema

DSD

$roles : \mathbb{P}\, Role$

$cardinality : \mathbb{N}$

$application : Application$

$(cardinality = 0) \lor (cardinality \geq 2) \land$
$\quad (cardinality \leq \#roles)$

INIT

$roles = \varnothing$

$cardinality = 0$

AddDSDRoleMember

$\Delta(roles)$

$role? : Role$

$role? \notin roles$

$role? \in application.roles$

$\forall\, d : DSD \mid d \in application.dsd\_sets \bullet$
$\quad (d.cardinality \geq (\#\{r : Role \mid r \in (role?.descended\_roles \cap$
$\qquad (d.roles \cup \{role?\}))\})) \land$
$\quad (\forall\, u : User \mid u \in application.rbac.users \bullet$
$\qquad (\forall\, s : Session \mid s \in u.sessions \bullet$
$\qquad\quad d.cardinality \geq (\#\{r : Role \mid r \in (s.effective\_active\_roles \cap$
$\qquad\quad (d.roles \cup \{role?\}))\}))$

$roles' = roles \cup \{role?\}$

RemoveDSDRoleMember

$\Delta(roles)$

$role? : Role$

$role? \in roles$

$roles' = roles \setminus \{role?\}$

SetDSDCardinality

$\Delta(cardinality)$

$cardinality? : \mathbb{N}$

$cardinality' = cardinality?$

C.2.9 RBAC System Class Schema

RBAC

$users : \mathbb{P}\, User_{\copyright}$

$applications : \mathbb{P}\, Application_{\copyright}$

$\forall\, a : Application \mid a \in applications \bullet a.rbac = self$

INIT

$users = \varnothing$

$applications = \varnothing$

AddUser

$\Delta(users)$

$user? : User$

$user? \notin users$

$users' = users \cup \{user?\}$

RemoveUser

$\Delta(users)$

$user? : User$

$user? \in users$

$users' = users \setminus \{user?\}$

CreateApplication

$\Delta(applications)$

$application? : Application$

$application? \notin applications$

$applications' = applications \cup \{application?\}$

$$\begin{array}{|l}
\multicolumn{1}{l}{\mathit{DeleteApplication}} \\
\hline
\Delta(\mathit{applications}) \\
\mathit{application}? : \mathit{Application} \\
\hline
\mathit{application}? \in \mathit{applications} \\
\mathit{applications}' = \mathit{applications} \setminus \{\mathit{application}?\} \\
\hline
\end{array}$$

$$\begin{array}{|l}
\multicolumn{1}{l}{\mathit{GetUserSessions}} \\
\hline
\mathit{user}? : \mathit{User} \\
\mathit{sessions}! : \mathbb{P}\,\mathit{Session} \\
\hline
\mathit{user}? \in \mathit{users} \\
\mathit{sessions}! = \mathit{user}?.\mathit{sessions} \\
\hline
\end{array}$$

$$\begin{array}{|l}
\multicolumn{1}{l}{\mathit{GetApplicationRoles}} \\
\hline
\mathit{application}? : \mathit{Application} \\
\mathit{roles}! : \mathbb{P}\,\mathit{Role} \\
\hline
\mathit{application}? \in \mathit{applications} \\
\mathit{roles}! = \mathit{application}?.\mathit{roles} \\
\hline
\end{array}$$

$$\mathit{DeleteUser} \mathrel{\widehat=} \mathit{DeassignUser} \ \bigwedge \ \mathit{RemoveUser}$$

$$\begin{array}{l}
\mathit{DeassignUser} \mathrel{\widehat=} \bigwedge a : \mathit{applications} \bullet \\
\qquad \left[\, \mathit{GetApplicationRoles}\left[\mathit{roles}/\mathit{roles}!\right] \,\middle|\, \mathit{application}? = a \,\right] \bullet \\
\qquad \bigwedge r : \mathit{roles} \bullet r.\mathit{DeassignUser}
\end{array}$$

Appendix D

Architecture Protocol Messages

D.1 Supported Messages

No	Message
101	MSG_INIT_ACCESS_CONTROL_CONTEXT_REQ
102	MSG_INIT_ACCESS_CONTROL_CONTEXT_REP
103	MSG_GET_PERMISSIONS_REQ
104	MSG_GET_PERMISSIONS_REP
105	MSG_CHECK_PERMISSION_REQ
106	MSG_CHECK_PERMISSION_REP
107	MSG_ACCESS_CONTROL_CONTEXT_CHANGED_UPD
108	MSG_ACCESS_CONTROL_CONTEXT_CHANGED_ACK
110	MSG_HEARTBEAT_UPD
111	MSG_HEARTBEAT_ACK
112	MSG_GET_AUTH_ROLES_REQ
113	MSG_GET_AUTH_ROLES_REP
114	MSG_DISPOSE_ACCESS_CONTROL_CONTEXT_REQ
115	MSG_DISPOSE_ACCESS_CONTROL_CONTEXT_REP
116	MSG_SET_UPDATE_PORT_UPD
199	MSG_ERR

Table D.1: Protocol Message Numbers

No	Error
201	MSG_INIT_ACCESS_CONTROL_CONTEXT_REQ
202	ERR_DISPOSE_ACCESS_CONTROL_CONTEXT_FAILED
202	ERR_AUTH_SYNC_FAILED
203	ERR_USER_UNKNOWN
204	ERR_ROLE_UNKNOWN
205	ERR_ROLE_DENIED
206	ERR_APP_UNKNOWN
207	ERR_PROTOCOL_VERSION
209	ERR_MSG_TYPE
210	ERR_BAD_MSG
299	ERR_GENERIC

Table D.2: Error Numbers

D.2 Implemented Message Examples

D.2.1 MSG_INIT_ACCESS_CONTROL_CONTEXT_REQ

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
    <message_type>101</message_type>
  </header>
  <body>
    <application>CN=App-WebProxy,CN=AuthData,CN=DCAA,
       DC=testdomain,DC=local</application>
    <principal>TestUser@TESTDOMAIN.LOCAL</principal>
    <roles>
      <role_cn>CN=MobileUser,CN=Roles,CN=App-WebProxy,
         CN=AuthData,CN=DCAA,DC=testdomain,DC=local
      </role_cn>
    </roles>
  </body>
</dcaa_protocol_msg>
```

D.2.2 MSG_INIT_ACCESS_CONTROL_CONTEXT_REP

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
    <message_type>102</message_type>
  </header>
  <body />
</dcaa_protocol_msg>
```

D.2.3 MSG_CHECK_PERMISSION_REQ

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
    <message_type>105</message_type>
  </header>
  <body>
    <principal>TestUser@TESTDOMAIN.LOCAL</principal>
    <permission>
      <class>dcaa.app_server.DCAAPermission</class>
      <name>AccessInternetDomain</name>
      <actions>qut.edu.au</actions>
    </permission>
  </body>
</dcaa_protocol_msg>
```

D.2.4 MSG_CHECK_PERMISSION_REP

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
  <message_type>106</message_type>
  </header>
  <body>
    <implies>true</implies>
  </body>
</dcaa_protocol_msg>
```

D.2.5 MSG_DISPOSE_ACCESS_CONTROL_CONTEXT_REQ

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
    <message_type>114</message_type>
  </header>
  <body>
    <application>CN=App-WebProxy,CN=AuthData,CN=DCAA,
       DC=testdomain,DC=local</application>
    <principal>TestUser@TESTDOMAIN.LOCAL</principal>
  </body>
</dcaa_protocol_msg>
```

D.2.6 MSG_DISPOSE_ACCESS_CONTROL_CONTEXT_REP

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
    <message_type>115</message_type>
  </header>
  <body />
</dcaa_protocol_msg>
```

D.2.7 MSG_SET_UPDATE_PORT_UPD

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
    <message_type>116</message_type>
  </header>
  <body>
   <update_port>3244</update_port>
  </body>
</dcaa_protocol_msg>
```

D.2.8 MSG_ACCESS_CONTROL_CONTEXT_CHANGED_UPD

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
    <message_type>107</message_type>
  </header>
  <body />
</dcaa_protocol_msg>
```

D.2.9 MSG_HEARTBEAT_UPD

```
<?xml version="1.0" encoding="UTF-8"?>
<dcaa_protocol_msg>
  <header>
    <protocol_version>2</protocol_version>
    <message_type>110</message_type>
  </header>
  <body />
</dcaa_protocol_msg>
```

D.2.10 Dynamic Context Service Update Message

```
<?xml version="1.0" encoding="UTF-8"?>
<DCUpdate>
  <DCUHeader>
    <DCSVersion>0.0.0.1</DCSVersion>
    <DCSName>dcaa.context_objects.location.BaseLocation</DCSName>
  </DCUHeader>
  <UpdatePrincipals>
    <DCAAPrincipal>
       rO0ABXNyABtkY2FhLmF1dGhfcHJvdG9jb2wuRENBQVVzZXJ1137cbB
       dX6QIAAUwACXByaW5jaXBhbHQAMExqYXZheC9zZWN1cml0eS9hdXRo
       L2tlcmJlcm9zL0tlcmJlcm9zUHJpbmNpcGFsO3hyACBkY2FhLmF1dG
       hfcHJvdG9jb2wuRENBQVByaW5jaXBhbBaBo3hY3nK1AgACSQAQaW50
       UHJpbmNpcGFsVHlwZUwADHZBY3RpdmVSb2xlc3QAEkxqYXZhL3V0aW
       wvVmVjdG9yO3hwAAAAAXNyABBqYXZhLnV0aWwuVmVjdG9y2Zd9W4A7
       rwEDAANJABFjYXBhY2l0eUluY3JlbWVudEkADGVsZW1lbnRDb3VudF
       sAC2VsZW1lbnREYXRhdAATW0xqYXZhL2xhbmcvT2JqZWN0O3hwAAAA
       AAAAAAB1cgATW0xqYXZhLmxhbmcuT2JqZWN0O5DOWJ8QcylsAgAAeH
       AAAAAKcHBwcHBwcHBwcHhzcgAuamF2YXguc2VjdXJpdHkuYXV0aC5r
       ZXJiZXJvcy5LZXJiZXJvc1ByaW5jaXBhbJmnfV0PHjMpAwAAeHB1cg
       ACW0Ks8xf4BghU4AIAAHhwAAAAFTAToAMCAQGhDDAKGwhUZXN0VXNl
       cnVxAH4ADAAAABIbEFRFU1RET01BSU4uTE9DQUx4
    </DCAAPrincipal>
  </UpdatePrincipals>
  <ContextObject>
    rO0ABXNyABVkY2FhLmRjc20uVGVzdENvbnRleHQuAXZdKN0rWQIAAUk
    ADGludFRocmVzaG9sZHhyABdkY2FhLmRjc20uQ29udGV4dE9iamVjdO
    EdOz3JHd57AgAFSQAPY29udGV4dEFjY3VyYWN5SQAMY29udGV4dFRyd
    XN0TAAKZGNzQXR0cmlic3QAEkxqYXZhL3V0aWwvVmVjdG9yO0wAEmRj
    c1VwZGF0ZVRpbWVzdGFtcHQAFExqYXZhL3NxbC9UaW1lc3RhbXA7TAA
    Nc3RyUGFyYW1ldGVyc3QAEkxqYXZhL2xhbmcvU3RyaW5nO3hwAAAAZA
    AAAGRwc3IAEmphdmEuc3FsLlRpbWVzdGFtcCYY1cgBU79lAgABSQAFb
    mFub3N4cgAOamF2YS51dGlsLkRhdGVoaoEBS1l0GQMAAHhwdwgAAAD5
    45VjOHgAp9jAcAAAAFQ=
  </ContextObject>
</DCUpdate>
```

Appendix E

Application of Jordan Curve Theorem to Location Applications

The Jordan Curve Theorem as defined by Weisstein in [100], is as follows:

> If J is a simple closed curve in $\mathbb{R}^2$, then $\mathbb{R}^2 - J$ has two components (an "inside" and "outside"), with J the boundary of each.

Using this theorem it is possible to determine whether a point is inside a polygon, based on the number of intersections detected on a ray from a given point to a maximum value as illustrated in Figure E.1. An even number of intersections indicates that the point is outside the polygon, where as an odd number of intersections indicates the point is within the polygon. We provide a simple Java implementation as detailed in Figure E.2.

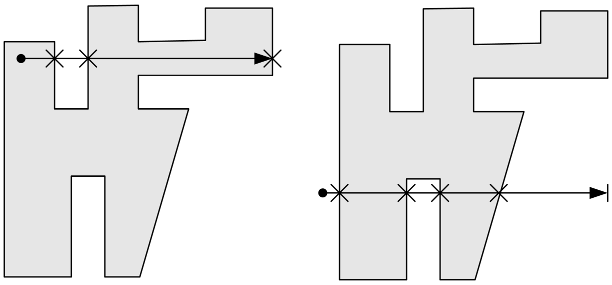


Figure E.1: Testing Whether a Point is Within a Polygon using Jordan Curve Theorum

```
/** Use Jordan curve theorum to calculate whether points are within
  * the polygon
  */
private boolean PointInPolygon(POLYGON p, UTM_POINT u) {

   /** Calculate the maximum and minimum points of the polygon on the X axis
     */
   int min = Integer.MAX_VALUE;
   int max = Integer.MIN_VALUE;
   int intersections = 0;
   for (int i=0;i<p.getNumberOfLLPoints();i++) {
      int val = Math.round(p.getLLPoint(i).getUTMPoint().getEasting());
      if (val < min) {
         min = val;
      }
      if (val > max) {
         max = val;
      }
   }


   /** Create a ray from UTM_POINT u to max, checking for the
     * number of intersections
     */
   for (int x = new Float(u.getEasting()).intValue();x<(max+1);x++){
      for (int i=0;i<p.getNumberOfLLPoints();i++) {
         int val = Math.round(p.getLLPoint(i).getUTMPoint().getEasting());
         if (val == x) {
            intersections++;
         }
      }
   }

   /** If there were an even number of intersections, the point is not
     * inside polygon, otherwise the point is inside the polygon
     */
   if ((intersections % 2) == 0) {
      return false;
   } else {
      return true;
   }
}
```

Figure E.2: Java Code to Check Whether a Point is Within a Polygon

Appendix F

An Overview of GSM

F.1 Introduction

This chapter provides an overview of GSM and components of the GSM architecture that have been discussed in this thesis. More in-depth information can be sourced from the GSM specifications[1].

F.2 Components of GSM

The entities of the GSM system are detailed below with a brief description of their functionality and relationship to other entities as defined in the GSM Architecture Specification [44].

The GSM system uses a number of registers to store data required to manage mobile subscribers. These registers are detailed as follows:

Home Location Register (HLR). The home location register is a database containing information to facilitate the management of mobile subscribers. The HLR stores information such as:

1. Mobile subscription information;
    (a) International Mobile Subscriber Identity (IMSI);
    (b) Mobile Station International Subscriber Dialing Number (MSISDN);

[1] Refer to `http://www.etsi.org/` for the GSM specifications.

2. MS Roaming Number (MSRN);
3. Visitor Location Register (VLR) address;
4. Mobile Switching Center (MSC) address;
5. Local MS identity; and
6. Other information:
   (a) Teleservices / bearer services subscription information;
   (b) Service restrictions; and
   (c) Supplementary services.

The HLR communicates with the Authentication Center (AuC) via the H-interface to retrieve the authentication and ciphering data for a mobile subscriber during an authentication request. Refer to [39] for more information on the organization of subscriber data.

Visitor Location Register (VLR). An MS roaming in a given location area must perform a registration when it moves into a new location area. The MSC controlling this area transfers the identity of the new location area to the VLR. In the case the MS has not been registered, the VLR and HLR exchange information. The VLR contains the following data required for call-setup or paging:

1. International Mobile Subscriber Identity (IMSI);
2. Temporary Mobile Subscriber Identity (TMSI);
3. Local Mobile Station Identity (LMSI);
4. Mobile Station ISDN (MSISDN);
5. Mobile Station Roaming Number (MSRN); and
6. Location area in which the MS has been registered.

VLRs communicate with each other on the G-interface (See Figure F.1[2]) for operations such as the retrieval of the IMSI and authentication parameters from an old VLR during a location registration procedure. Refer to [39] for more information on the organization of subscriber data.

Equipment Identity Register (EIR). This entity contains a database storing the International Mobile Equipment Identities (IMEIs) of subscribers which may be classified as:

[2] Image sourced from [44].

1. White listed;
2. Grey listed; and
3. Black listed.

F.2.1 Authentication Center (AuC)

The Authentication Center stores an identity key associated for each subscriber registered with a given HLR. This key is used for authentication and ciphering of data over the network. Refer to [48] for more information on authentication and ciphering procedures.

F.2.2 Mobile-services Switching Center (MSC)

The mobile-services switching center, is principally an exchange that performs additional mobile-related procedures such as location registration[33] and handover[35]. The MSC performs switching and signaling for all subscribers in the location area managed by the MSC. Many MSCs make up a Public Land Mobile Network (PLMN), provides interfaces between fixed networks and the PLMN.

There are a number of interfaces connecting the MSC to various entities. Communication between the BSC and the MSC is facilitated by the A-interface, used for BSS management, call handling and mobility management. The interface used for communication between the MSC and VLR is the B-interface. This interface is used in situations where the MSC must query the VLR or perform a location update. The C-interface is used for communication between the MSC and HLR. This interface is typically used in situations where the MSC must query the HLR for information such as call routing information. The F-interface facilitates communication between the MSC and the EIR for status verification of the mobile station IEMI. These interfaces are illustrated in Figure F.1[3].

A Gateway MSC (GMSC) performs routing from the current MSC to the MSC where the MS is located. This functionality is used when the network delivering a call to the Public Land Mobile Network (PLMN) is unable to query the HLR. The call is subsequently routed to an MSC which performs the HLR query and then routes the call to the appropriate MSC. For more information refer to [34].

[3] Image sourced from [44].

F.2.2.1 SMS Delivery

SMS Gateway MSC (SGMSC). The SMS gateway MSC facilitates the delivery of SMS messages from Short Message Service Center (SMSC) to mobile stations.

SMS Interworking MSC. The SMS Interworking MSC facilitates the delivery of SMS messages from a mobile station to an SMSC.

F.2.3 Base Station System (BSS)

The base station system consists of numerous Base Transceiver Stations (BTS) connected to a Base Station Controller (BSC) via the Abis-interface. This interface is used to carry data for controlling the radio equipment and radio frequency allocation of the BTS connected to it. The BSC is connected to the MSC via the A-interface as shown in Figure F.1[4]. The A-interface is used to carry data for BSS management, call handling and mobility management. Each BSS contains only one BSC and can have many BTSs, each of which is responsible for serving a single cell. This functionality is detailed in [36].

F.2.4 Mobile Station (MS)

The mobile station is the physical terminal used by a subscriber containing a Subscriber Identity Module (SIM). Communication between the MS and BTS is performed over the radio interface. This communication process is discussed in section F.2.5.

F.2.5 GSM Radio Subsystem

GSM uses two bands of 25MHz, 890-915Mhz for the reverse link (MS to BTS transmissions) and 935-960MHz for the forward link (BTS to MS transmissions)[82]. GSM uses Frequency Division Duplexing (FDD) and utilizes a combination of Time Division Multiple Access (TDMA - See Figure F.2[5]) and Frequency Hopping Multiple Access (FHMA) to facilitate multiple subscribers on a single carrier frequency. The forward and reverse links are divided into Associated Radio Frequency Channel Numbers (ARFCN) of 200kHz where the forward and reverse channel pairs are separated by 45MHz. These channels are further divided into 8 timeslots using TDMA, in which

[4] Image sourced from [44].

[5] Image sourced from [82].

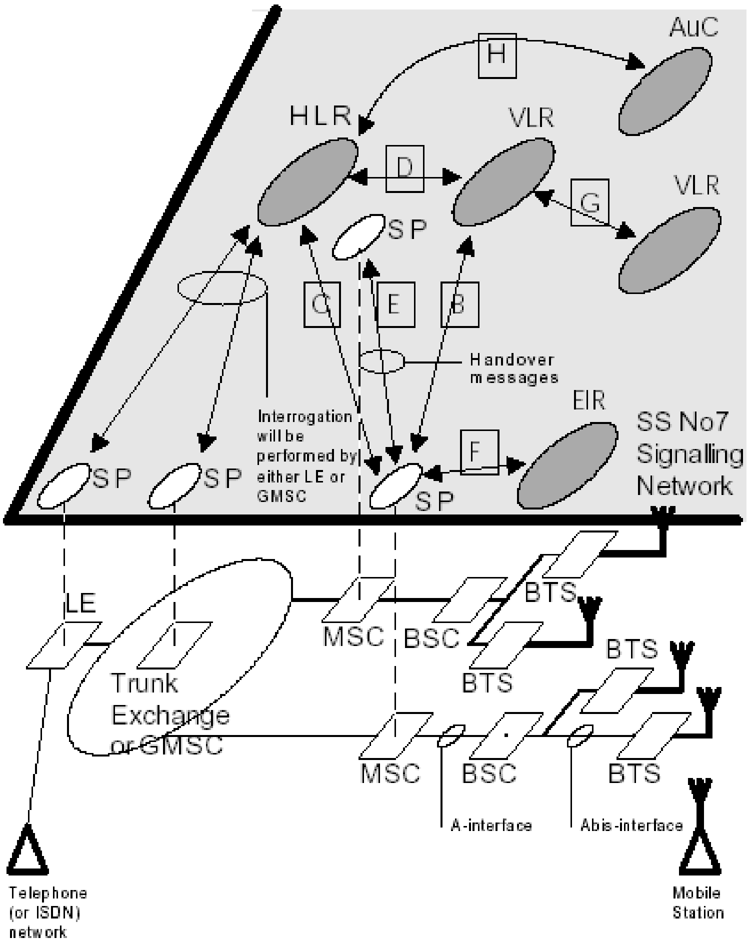


Figure F.1: Configuration of a PLMN and its Interfaces

up to 8 subscribers can simultaneously use the same ARFCN by occupying a single timeslot in every frame.

A single TDMA frame has a duration of 4.615ms where each timeslot has a duration of 576.9μs and where a bit period has a duration of 3.692μs. Each timeslot has an allocation of 156.25 bits where 8.25 bits are guard time and 6 bits are start and stop time for preventing timeslot overlap. In effect, each subscriber only has the carrier frequency for 576.9μs and must transmit within that time period. This requires that the mobile station is time-synchronized with the BTS.

F.2.6 Handover Procedure

The strengths of signals from surrounding BTSs are important in deciding which cell to handover to. In order for the BSC to initiate a handover, it must be aware of the

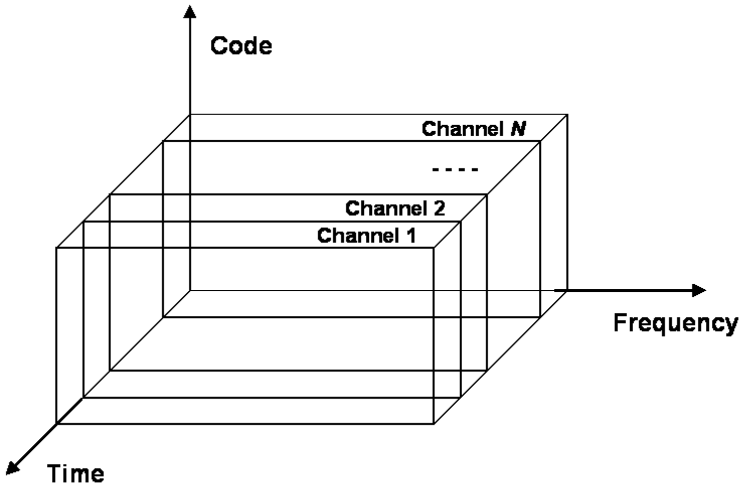


Figure F.2: Time Division Multiple Access

signal strengths and signal reception quality of neighboring BTSs. Hence, the MS must continuously measure the signal strengths from surrounding BTSs. To identify the cells that are being measured, the MS synchronizes and demodulates surrounding BCCH carriers to identify their BSIC. This is done as the carrier frequency alone is not sufficient to determine the cell due to frequency reuse.

The radio measurements and the BSICs identifying the BTS from which they were obtained are sent in a MEAUSREMENT REPORT[37] message over the Slow Associated Control Channel (SACCH) to the BTS, which forwards the MEASUREMENT REPORT to the BSC. The conversion of measured signal levels to RXLEVs in the MEASUREMENT REPORT and the conversion of error rate to RXQUAL is detailed below:

Mapping of signal level to RXLEV[40]

RXLEV 0	= less than -110 dBm + SCALE
RXLEV 1	= -110 dBm + SCALE to -109 dBm + SCALE
RXLEV 2	= -109 dBm + SCALE to -108 dBm + SCALE
..	
..	
RXLEV 62	= -49 dBm + SCALE to -48 dBm + SCALE
RXLEV 63	= greater than -48 dBm + SCALE

Where SCALE is an offset parameter with a default of 0 dB.

Mapping of signal quality (Bit Error Rate) to RXQUAL[40]

RXQUAL 0	= less than 0.2%
RXQUAL 1	= 0.2% to 0.4%
RXQUAL 2	= 0.4% to 0.8%
RXQUAL 3	= 0.8% to 1.6%
RXQUAL 4	= 1.6% to 3.2%
RXQUAL 5	= 3.2% to 6.4%
RXQUAL 6	= 6.4% to 12.8%
RXQUAL 7	= greater than 12.8%

For details of the information elements within the measurement report, refer to section F.2.7.1 and for more details on the measurement reporting process, refer to [40]. The transmittal of measurements to the BSC is defined in [38].

Based on the measurement reporting from the MS, the BSC may choose to initiate a handover if the serving cell's signal strength is less than a neighboring cell or the signal quality is less than a neighboring cell. There are two types of handover:

1. **Inter-cell Handover** Occurs when the measurements reported from the MS indicate a low RXLEV and/or RXQUAL for the current serving cell and a neighboring cell with a better RXLEV is available.

2. **Intra-cell Handover** Occurs when the measurements reported from the MS indicate a low RXQUAL but a high RXLEV for the current serving cell indicating interference. The intra-cell handover attempts to provide the MS with a channel less affected by interference than the current cell.

The BSC initiates a handover by issuing a HANDOVER COMMAND[37] message to the MS, requiring that it moves to the new channel. The MS proceeds by transmitting the HANDOVER ACCESS[37] command on the new channel until it receives the PHYSICAL INFORMATION[37] message, containing the timing advance for new channel. The layer 2 connection is then established, and on successful establishment, the MS sends a HANDOVER COMPLETE[37] message after which the network releases the old channels.

F.2.7 GSM Measurements Facilitating Location Determination

Numerous techniques exist to derive location from network information used by GSM for signal synchronization and cell handover. The following information can be obtained from GSM Layer 3 messages[37] to assist in deriving the location of a mobile station:

1. *Timing Advance.* Used to correct MS transmission timing for signal propagation delay. This measurement indicates the round-trip propagation time of a signal transmitted by the BTS to the MS and back to the BTS.

2. *Observed Time Difference.* Used for pseudo-synchronized handover, such that the MS is pre-synchronized to the BTS it is handed over to.

3. *Signal Level Measurement Results.* Observed by MS and sent to the BSC via the BTS for the use of making handover decisions. These measurements are comprised of the neighbor and serving cell signal reception levels.

4. *Current Serving Cell.* This measurement provides the country, location area, cell ID and network for the current serving cell. The radio communications regulatory body such as the ACA[5] in Australia provide publicly available information such as the easting and northing of BTS antennas, antenna azimuth and antenna beamwidth.

The information described above is contained an some of the following GSM layer 3 information elements:

F.2.7.1 Measurement Results Information Element

This measurement element forms part of the MEASUREMENT REPORT Message [37].

BA-USED. Bit indicates whether the BA-IND value, the BCCH allocation sequence number indication from the Neighbor Cells description information element, is used in the coding of BCCH-FREQ-NCELL.

DTX-USED. Bit indicates whether the MS used DTX (discontinuous transmission) during previous measurement period. The measurement period has a duration of 480ms between each measurement report which is sent over the Slow Associated Control Channel (SACCH) frame.[40]

<table>
<tr><td>8</td><td>7</td><td>6</td><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td><td></td></tr>
<tr><td></td><td colspan="7">Measurement Results IEI</td><td>octet 1</td></tr>
<tr><td>BA-USED</td><td>DTX USED</td><td colspan="6">RXLEV-FULL-SERVING-CELL</td><td>octet 2</td></tr>
<tr><td>0 spare</td><td>MEAS-VALID</td><td colspan="6">RXLEV-SUB-SERVING-CELL</td><td>octet 3</td></tr>
<tr><td>0 spare</td><td colspan="3">RXQUAL-FULL SERVING-CELL</td><td colspan="3">RXQUAL-SUB SERVING-CELL</td><td>NO-NCELL M (high part)</td><td>octet 4</td></tr>
<tr><td colspan="2">NO-NCELL-M (low part)</td><td colspan="6">RXLEV-NCELL 1</td><td>octet 5</td></tr>
<tr><td colspan="5">BCCH-FREQ-NCELL 1</td><td colspan="3">BSIC-NCELL 1 (high part)</td><td>octet 6</td></tr>
<tr><td colspan="3">BSIC-NCELL 1 (low part)</td><td colspan="5">RXLEV-NCELL 2 (high part)</td><td>octet 7</td></tr>
<tr><td>RXLEV NCELL 2 (low part)</td><td colspan="5">BCCH-FREQ-NCELL 2</td><td colspan="2">BSIC-NCELL 2 (high part)</td><td>octet 8</td></tr>
<tr><td colspan="4">BSIC-NCELL 2 (low part)</td><td colspan="4">RXLEV-NCELL 3 (high part)</td><td>octet 9</td></tr>
<tr><td colspan="2">RXLEV-NCELL 3 (low part)</td><td colspan="5">BCCH-FREQ-NCELL 3</td><td>BSIC-NCELL 3 (high part)</td><td>octet 10</td></tr>
<tr><td colspan="5">BSIC-NCELL 3 (low part)</td><td colspan="3">RXLEV-NCELL 4 (high part)</td><td>octet 11</td></tr>
<tr><td colspan="3">RXLEV-NCELL 4 (low part)</td><td colspan="5">BCCH-FREQ-NCELL 4</td><td>octet 12</td></tr>
<tr><td colspan="6">BSIC-NCELL 4</td><td colspan="2">RXLEV-NCELL 5 (high part)</td><td>octet 13</td></tr>
<tr><td colspan="4">RXLEV-NCELL 5 (low part)</td><td colspan="4">BCCH-FREQ-NCELL 5 (high part)</td><td>octet 14</td></tr>
<tr><td>BCCH-FREQ-NCELL 5 (low part)</td><td colspan="6">BSIC-NCELL 5</td><td>RXLEV-NCELL 6 (high part)</td><td>octet 15</td></tr>
<tr><td colspan="5">RXLEV-NCELL 6 (low part)</td><td colspan="3">BCCH-FREQ-NCELL 6 (high part)</td><td>octet 16</td></tr>
<tr><td colspan="2">BCCH-FREQ-NCELL 6 (low part)</td><td colspan="6">BSIC-NCELL 6</td><td>octet 17</td></tr>
</table>

Table F.1: Measurement Results Information Element

RXLEV-FULL-SERVING Received signal strength of serving cell measured on all slots.

RXLEV-SUB-SERVING Received signal strength of serving cell measured on a subset of slots.

RXQUAL-FULL-SERVING-CELL Received signal quality of serving cell measured on all slots.

RXQUAL-SUB-SERVING-CELL Received signal quality of serving cell measured on a subset of slots.

MEAS-VALID Bit indicates whether the measurements for the dedicated channel are valid.

NO-NCELL-M The number of neighboring cell measurements.

RXLEV-NCELL Received signal strength for given neighbor.

BCCH-FREQ-NCELL BCCH carrier of for the given neighbor.

BSIC-NCELL Base Station Identity Code for the given neighbor.

The RXLEV-FULL-SERVING and RXLEV-NCELL measurements can be used for location calculation in both idle and active mode. By using a propagation-pathloss model, it is possible to derive a scalar distance from each BTS. These scalars can then be used to trilaterate the position of the mobile station.

F.2.7.2 Timing Advance Information Element

This information element forms part of the PHYSICAL INFORMATION Message.[37]

8	7	6	5	4	3	2	1	
	Timing Advance IEI							octet 1
0 spare	0 spare	Timing advance value						octet 2

Table F.2: Timing Advance Information Element

Timing advance value The coding of the timing advance value is the binary representation of the timing advance in bit periods, where 1 bit period = 48/13 μs.

The timing advance value can be used to derive a scalar distance from the serving BTS. The timing advance is only obtainable in active mode.

F.2.7.3 Pseudo-Synchronization Information Elements

The Phase 2 GSM Layer 3 specification [37] defines a number of messages, the HANDOVER COMMAND message which contains the Synchronization Indication and Real Time Difference information elements, and the HANDOVER COMPLETE message which contains the Mobile Observed Time Difference information element. These information elements are principally used for the purpose of synchronized handovers.

Synchronization Indicator Information Element This information element forms part of the HANDOVER COMMAND message issued by the BSC.

8	7	6	5	4	3	2	1	
	Synch. Indic. IEI			NCI	ROT	SI		octet 1

Table F.3: Synchronization Indicator Information Element

ROT Report Observed Time Difference

- **0** - Mobile Time Difference information element shall not be included in the HANDOVER COMPLETE message. This is dependent on whether the network supports some form of synchronized handover.
- **1** - Mobile Time Difference information element shall be included in the HANDOVER COMPLETE message

SI Synchronization indication

- **0 0** - Non-synchronized
- **0 1** - Synchronized
- **1 0** - Pre-synchronized
- **1 1** - Pseudo-synchronized

NCI Normal cell indication

- **0** - Out of range timing advance is ignored
- **1** - Out of range timing advance shall trigger a handover failure procedure

This information element is important in location determination, in ensuring the Time Difference information element is sent to the BSC. This allows the network operator to correct the mobile time difference by $\frac{1}{2}$ the Timing Advance and then return it to the ME as the in the HANDOVER COMPLETE message.

Time Difference Information Element The time difference provides information about the synchronization difference between the time bases of two base stations. This information element forms part of the HANDOVER COMMAND message issued by the BSC.

8	7	6	5	4	3	2	1	
	Time Difference IEI							octet 1
Length of Time Difference contents								octet 2
time difference value								octet 3

Table F.4: Time Difference Information Element

The time difference information element is included in the HANDOVER COMMAND message when the Synchronization Indication information element has a ROT value of 1. The time difference value is encoded in a binary format where the time difference is represented by half bit periods modulo 256 where $\frac{1}{2}$ bit period = 24/13 μs.

F.2.7.4 Mobile Observed Time Difference Information Element

This information element forms part of the HANDOVER COMPLETE message.

8	7	6	5	4	3	2	1	
	Mobile Time Difference IEI							octet 1
Length of Mobile Time Difference contents								octet 2
Mobile Time Difference contents (high)								octet 3
Mobile Time Difference contents (contd)								octet 4
Mobile Time Difference contents (low)					0 spare	0 spare	0 spare	octet 5

Table F.5: Mobile Observed Time Difference Information Element

The mobile observed time difference information element is included in the HANDOVER COMPLETE message when the Synchronization Indication information element has a ROT value of 1. The time difference value is encoded in a binary format where the time difference is represented by half bit periods modulo 256 where $\frac{1}{2}$ bit period = 24/13 μs.

F.2.7.5 Cell Identity Information Element

The cell identity information element is included in the SYSTEM INFORMATION TYPE 6 message. This information element identifies the serving cell within a given location area code (LAC).

8	7	6	5	4	3	2	1	
	Cell Identity IEI							octet 1
CI value								octet 2
CI value (continued)								octet 3

Table F.6: Cell Identity Information Element

CI Cell Identity

This information in combination with the Location Area Identification Element can identify the serving cell. The radio communications regulatory body such as the

ACA[5] in Australia provide publicly available information such as the easting and northing of BTS antennas, antenna azimuth and antenna beamwidth. This information can be used in combination with signal measurement-based and or timing-based to derive the mobile equipment's location.

F.2.7.6 Location Area Identification Information Element

The Location Area Identification message identifies an unambiguous location area covered by GSM.

8	7	6	5	4	3	2	1	
	Location Area Identification IEI							octet 1
MCC digit 2				MCC digit 1				octet 2
1 1 1 1				MCC digit 3				octet 3
MCC digit 2				MCC digit 1				octet 4
LAC								octet 5
LAC (continued)								octet 6

Table F.7: Location Area Information Element

MCC Mobile Country Code uniquely identifies the country within which a mobile network operates.

MNC Mobile Network Code uniquely identifies the mobile network within a country.

LAC Location Area Code uniquely identifies an area within a given network and country.

Appendix G

An Overview of GSM Mobile Positioning Protocol

G.1 Introduction

The following subsections detail the mobile positioning protocol used between the LCS client and the Gateway Mobile Positioning Center (GMPC) and the supported location shapes as defined in the GMS LCS98 specifications.

G.2 Mobile Positioning Protocol

In this section, the Ericsson Mobile Positioning Protocol (MPP) [29] will be detailed, as it is a working implementation of the LCS98 specifications. MPP is a protocol implemented over HTTP or HTTPS for making location requests from an LCS client to a GMPC server. The protocol supports two types of location requests:

- *Location Immediate Request (LIR):* A request that is sent in a HTTP POST message to the GMPC server to request the location of one or more Mobile Stations (MS). This request includes parameters such as the quality of service required, the type of location whether current or last known, and geographical information including the coordinate system, format and datum. This request is followed by a Location Immediate Answer (LIA), containing the position for the requested MS', or an error indicating that the location request failed. This protocol is illustrated in Figure G.1.

- *Location Push Answer (LPA):* This type of location request may originate from an MS and would be pushed to the LCS client using HTTP POST. This message would contain location information similar to the LIA. Figure G.2 illustrates the Push protocol.

$$\begin{array}{llll} C & \rightarrow & GMPC: & C, C_K, MS_{1\ldots n}, QoS, GeoInfo, LocType \qquad (1) \\ C & \leftarrow & GMPC: & T_{GMPC}, Pos_{1\ldots n} \qquad (2) \end{array}$$

GeoInfo = {*CoordSystem, Datum, Format*}
QoS = {T_{resp}, *HorizontalAccuracy*}
Pos = {*MS*, T_{pos}, *ConfidenceLevel, Shape*}
Where *C* is the LCS client, C_K is the LCS client key or password, *MS* is the ID of the MS, and *T* is a timestamp.

Table G.1: Mobile Positioning Protocol - LIR/LIA

$$GMPC \;\rightarrow\; C: \quad PushUser, PushUser_K, T_{GMPC}, Pos \qquad (1)$$

Pos = {*MS*, T_{pos}, *ConfidenceLevel, Shape*}

Table G.2: Mobile Positioning Protocol - LPA

These protocol messages are encoded in XML[1]. Examples of the LIR, LIA messages are given in Figures G.1, G.2 and G.3 respectively[2].

The LIA message in Figure G.2 illustrates the LIA of a point with an uncertainty circle. Figure G.7 illustrates the LIA of a polygon. These two examples are given, as they are pertinent to the proposed GSM location system detailed in Chapter 4.

G.3 Location Area Representations

This section details the location shapes supported in the GSM LCS98 specifications, and by the MPP protocol. These shapes with their corresponding encoding and protocol messages are defined in [46]. The definition of the shapes below are taken directly from the Universal Geographical Area Description (GAD) of the GSM specifications[46].

[1] Refer to [29] for the DTDs of the protocol messages.

[2] Example XML messages are sourced from [29].

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes"?>
<!DOCTYPE REQ SYSTEM "file://mpp50_req.dtd">
<REQ ver="5.0">
   <CLIENT>
      <ID>TheUser</ID>
      <PWD>The5PW</PWD>
   </CLIENT>
   <LIR>
      <MSIDS>
         <MSID>461011334411</MSID>
         <MSID>461011334414</MSID>
         <MSID_RANGE>
            <START_MSID>461011334500<START_MSID>
            <STOP_MSID>461011334599</STOP_MSID>
         </MSID_RANGE>
      </MSIDS>
      <QoS>
         <RESP_TIME>0</RESP_TIME>
         <HORIZON_ACC>0</HORIZON_ACC>
      </QoS>
      <GEO_INFO>
         <COORD_SYS>LL</COORD_SYS>
         <DATUM>WGS-84</DATUM>
         <FORMAT>IDMS0</FORMAT>
         </GEO_INFO>
   </LIR>
</REQ>
```

Figure G.1: Example Location Immediate Request (LIR)

G.3.1 Ellipsoid Point

The description of an ellipsoid point is that of a point on the surface of the ellipsoid, and consists of a latitude and a longitude. In practice, such a description can be used to refer to a point on Earth's surface, or close to Earth's surface, with the same longitude and latitude. No provision is made in this version of the standard to give the height of a point.

Figure G.4 illustrates a point on the surface of the ellipsoid and its co-ordinates. The latitude is the angle between the equatorial plane and the perpendicular to the plane tangent to the ellipsoid surface at the point. Positive latitudes correspond to the North hemisphere. The longitude is the angle between the half-plane determined by the Greenwich meridian and the half-plane defined by the point and the polar axis, measured Eastward.

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes"?>
<!DOCTYPE ANS SYSTEM "file://mpp50_ans.dtd">
<ANS ver="5.00">
   <LIA>
      <GMT_OFF>+0100</GMT_OFF>
      <POS msid="1234512345">
         <PD>
            <TIME>20020626171825</TIME>
            <POINT_UNCERT_CIR>
               <LL_POINT>
                  <LAT>N561157</LAT>
                  <LONG>E0151716</LONG>
               </LL_POINT>
               <UNCERT>200</UNCERT>
            </POINT_UNCERT_CIR>
         </PD>
      </POS>
   </LIA>
</ANS>
```

Figure G.2: Example Location Immediate Answer (LIA) for point with uncertainty circle

G.3.2 Ellipsoid Point with Uncertainty Circle

The "ellipsoid point with uncertainty circle" is characterized by the coordinates of an ellipsoid point (the origin) and a distance r. It describes formally the set of points on the ellipsoid which are at a distance from the origin less than or equal to r, the distance being the geodesic distance over the ellipsoid, i.e., the minimum length of a path staying on the ellipsoid and joining the two points, as shown in Figure G.5.

As for the ellipsoid point, this can be used to indicate points on the Earth surface, or near the Earth surface, of same latitude and longitude. The typical use of this shape is to indicate a point when its position is known only with a limited accuracy.

G.3.3 Ellipsoid Point with Uncertainty Ellipse

The "ellipsoid point with uncertainty ellipse" is characterized by the co-ordinates of an ellipsoid point (the origin), distances r_1 and r_2 and an angle of orientation A. It describes formally the set of points on the ellipsoid which fall within or on the boundary of an ellipse with semi-major axis of length r_1 oriented at angle $A(0^\circ$ to $180^\circ)$ measure clockwise from north and semi-minor axis of length r_2, the distances being the geodesic distance over the ellipsoid, i.e., the minimum length of a path staying on the ellipsoid and joining the two points, as shown in Figure G.6.

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes"?>
<!DOCTYPE ANS SYSTEM "file://mpp50_ans.dtd">
<ANS ver="5.00">
   <LIA>
      <GMT_OFF>+0100</GMT_OFF>
      <POS msid="1234512345">
         <PD>
            <TIME>20020626171825</TIME>
            <POLYGON>
               <LL_POINT>
                  <LAT>N561157</LAT>
                  <LONG>E0151716</LONG>
               </LL_POINT>
               <LL_POINT>
                  <LAT>N561212</LAT>
                  <LONG>E0151746</LONG>
                  </LL_POINT>
               <LL_POINT>
                  <LAT>N561201</LAT>
                  <LONG>E0151801</LONG>
                  </LL_POINT>
               <LL_POINT>
                  <LAT>N561144</LAT>
                  <LONG>E0151752</LONG>
               </LL_POINT>
               <LL_POINT>
                  <LAT>N561151</LAT>
                  <LONG>E0151725</LONG>
               </LL_POINT>
            </POLYGON>
         </PD>
      </POS>
   </LIA>
</ANS>
```

Figure G.3: Example Location Immediate Answer (LIA) for polygon shape

As for the ellipsoid point, this can be used to indicate points on the Earth's surface, or near the Earth's surface, of same latitude and longitude. The confidence level with which the position of a target entity is included within this set of points is also included with this shape. The typical use of this shape is to indicate a point when its position is known only with a limited accuracy, but the geometrical contributions to uncertainty can be quantified.

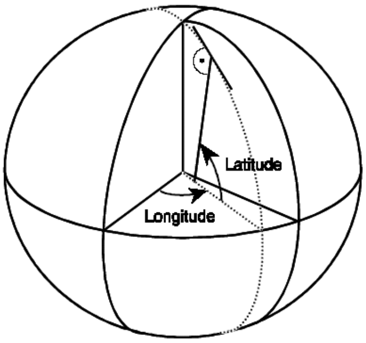


Figure G.4: Ellipsoid Point

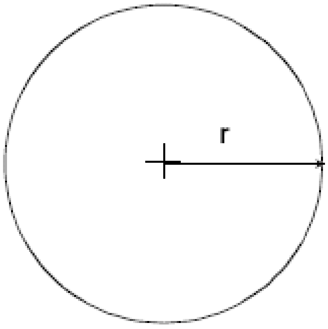


Figure G.5: Ellipsoid Point with Uncertainty Circle

G.3.4 Polygon

A polygon is an arbitrary shape described by an ordered series of points (in the example pictured in the drawing, *A* to *E*). The minimum number of points allowed is 3, and the maximum number of points allowed is 15. The points shall be connected in the order that they are given. A connecting line is defined as the line over the ellipsoid joining the two points and of minimum distance (geodesic). The last point is connected to the first. The list of points shall respect a number of conditions[3]:

- a connecting line shall not cross another connecting line; and
- two successive points must not be diametrically opposed on the ellipsoid.

[3] NOTE: This definition does not permit connecting lines greater than roughly 20,000 km. If such a need arises, the polygon can be described by adding an intermediate point. Computation of geodesic lines is not simple. Approximations leading to a maximum distance between the computed line and the geodesic line of less than 3 meters are acceptable.

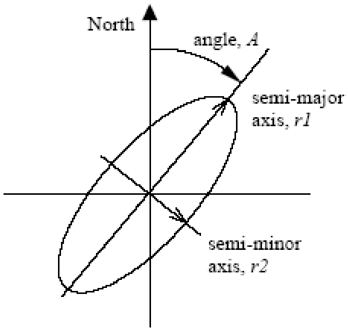


Figure G.6: Ellipsoid Point with Uncertainty Ellipse

The described area is situated to the right of the lines with the downward direction being toward the Earth's center and the forward direction being from a point to the next.

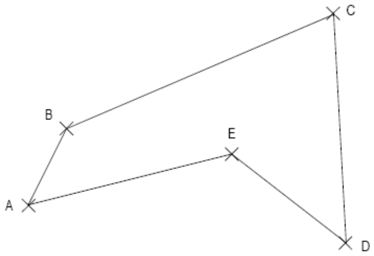


Figure G.7: Polygon

G.3.5 Ellipsoid Point with Altitude

The description of an ellipsoid point with altitude is that of a point at a specified distance above or below a point on the earths surface. This is defined by an ellipsoid point with the given longitude and latitude and the altitude above or below the ellipsoid point. Figure G.8 illustrates the altitude aspect of this description.

G.3.6 Ellipsoid Point with Altitude and Uncertainty Ellipsoid

The "ellipsoid point with altitude and uncertainty ellipsoid" is characterized by the co-ordinates of an ellipsoid point with altitude, distances r_1 (the "semi-major uncertainty"), r_2 (the "semi-minor uncertainty") and r_3 (the "vertical uncertainty") and an angle of orientation A (the "angle of the major axis"). It describes formally the set of

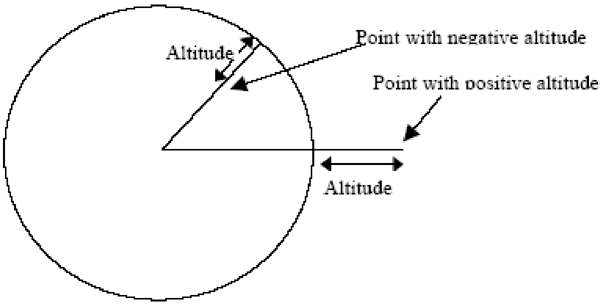


Figure G.8: Ellipsoid Point with Altitude

points which fall within or on the surface of a general (three dimensional) ellipsoid centered on an ellipsoid point with altitude whose real semi-major, semi-mean and semi-minor axis are some permutation of r_1, r_2, r_3 with $r_1 \geq r_2$. The r_3 axis is aligned vertically, while the r_1 axis, which is the semi-major axis of the ellipse in a horizontal plane that bisects the ellipsoid, is oriented at an angle A (0° to $180°$) measured clockwise from north, as illustrated in Figure G.9.

The typical use of this shape is to indicate a point when its horizontal position and altitude are known only with a limited accuracy, but the geometrical contributions to uncertainty can be quantified. The confidence level with which the position of a target entity is included within the shape is also included.

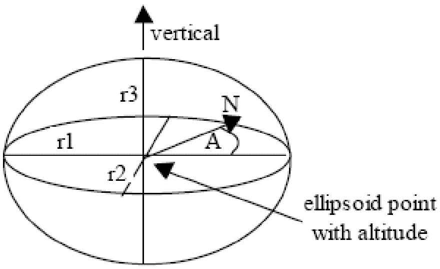


Figure G.9: Ellipsoid Point with Altitude and Uncertainty Ellipsoid

G.3.7 Ellipsoid Arc

An ellipsoid arc is a shape characterized by the co-ordinates of an ellipsoid point o (the origin), inner radius r_1, uncertainty radius r_2, both radii being geodesic distances over the surface of the ellipsoid, the offset angle (θ) between the first defining radius of the ellipsoid arc and North, and the included angle (β) being the angle between the first and second defining radii. The offset angle is within the range of $0°$ to $359°$ while the included angle is within the range from $1°$ to $360°$. This is to be able to describe a full circle, $0°$ to $360°$.

This shape-definition can also be used to describe a sector (inner radius equal to zero), a circle (included angle equal to $360°$) and other circular shaped areas. The confidence level with which the position of a target entity is included within the shape is also included.

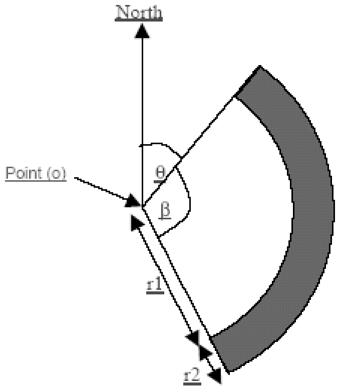


Figure G.10: Ellipsoid Arc

Appendix H

Prototype Implementation of MPC

A prototype network operator location system was developed for the primary purpose of testing MPC-based prototype applications.

H.1 Mobile Location Center Gateway Emulation

It has been established that location information is most securely obtained from the network operator as discussed in section 4.9.1. Due to the unavailability of network-based location systems in Australia at the time research was conducted, an emulated network operator had to be built.

This prototype server implements the Ericsson MPP, version 3.0. The location information is obtained from a direct cable-connection to a GSM mobile phone, rather than from the network operator. This solution was designed to facilitate the development of a prototype location server for security services such as authentication and authorization, providing them with real-time location data.

The following sections detail the design of the MPC server and associated components.

H.2 Prototype Development Discussion

This prototype obtains its location from the MS, effectively voiding all security benefits of obtaining location from the network operator. It does, however, serve as a reliable source of real time location data for testing.

The Nokia 5110 phone was chosen for use in the network operator prototype due to the ease with which timing advance and cell information could be obtained from the handset. The Nokia 5110 phone protocol has been analyzed by contributors to the Gnokii project[1] and a high-level ANSI C API for Linux has been written to perform functions using the Nokia serial communications protocol.

The Ericsson MPSSDK version 3.0 was distributed with a number of Java MPC client, as well as the MPC emulator and associated documentation. While the client examples were distributed with source code, the MPC emulator source was not distributed, requiring the redevelopment of an MPC emulator.

H.3 Detailed Design

Figure H.1 details the design of the prototype MPC emulator. This subsection details the components of the emulator and the data used to calculate the location of a cell phone.

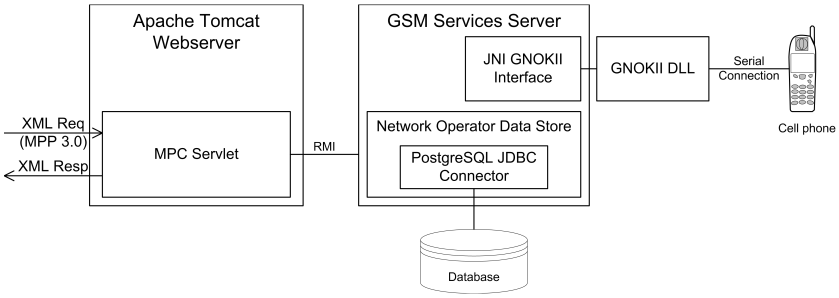


Figure H.1: MPC Emulator Prototype

H.3.1 Network Operator Database Component

The network operator database component design was based on the data required to calculate the location of a mobile phone using the Timing Advance (TA) and cell information provided by Vodafone Australia.

1. Network data files from Vodafone Australia consisting of the following information:
   - *Location Area Code (LAC).* An identification code for a location area unique to a given network operator.

[1] Refer to `http://www.gnokii.org` for the Gnokii project.

- *Cell ID (CID).* An identification code for a BTS unique to a given location area.
- *Broadcast Channel (BCCH).* The broadcast channel of a given BTS.
- *Base Station Controller (BSC).* The Base Station Controller of a given BTS.
- *BTS Site Name.* The name of the site containing where a given BTS is deployed.
- *BTS Site Location.* The location of a given BTS.
  - Easting
  - Northing
  - Zone
- *Neighbouring Cells' CID.*
- *Cell Broadcast.* A cell broadcast message from the active cell containing the BTS identifier.

2. Information available from Australian Communications Authority (ACA)[5]

    - *Coordinates of each BTS.* Represented in Eastings and Northings for a given zone.
    - *Antenna Height.* The height of the antenna.
    - *Antenna Azimuth.* The angular distance along the horizon to the location of the object. By convention, azimuth is measured from north towards the east along the horizon.
    - *Antenna Tilt.* The tilt of the antenna. +90 through 0 to -90 degrees.
    - *Antenna Size.* The physical measurement of a parabolic antenna.
    - *Antenna Beam Width.* The angle between the -3db beamwidth points off the main lobe of the antenna.
    - *Antenna Gain.* The main lobe gain of the antenna in (db) referenced to an isotropic radiator.
    - *Antenna Front to Back.* The ratio of forward gain to reverse gain of a directional antenna.
    - *Antenna Type.* The type of antenna used.
    - *Effective Radius.* The effective radius of a mobile site.

- *Carrier Frequency.* This information can be used by propagation models to enhance location accuracy.
- *Transmitter Power.* The maximum level of RF power permitted to be produced by the transmitter. This information can be used by propagation models to enhance location accuracy.

3. International Telecommunication Union (ITU) [63]

- *Mobile Network Code (MNC).* An identification code for a network operator unique to a given country.
- *Mobile Country Code (MCC).* An identification code for the country in which GSM networks are installed.

H.3.2 Database Design

The conceptual model of the database is detailed in figure H.2.

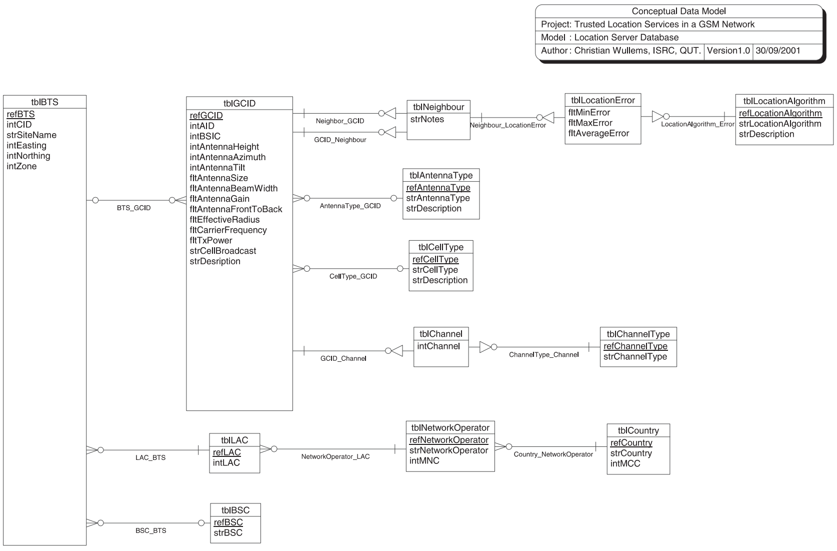


Figure H.2: Conceptual Data Model of Database on Secure Location Server

Based on this conceptual model, a physical database model was created for the PostgreSQL[2] platform.

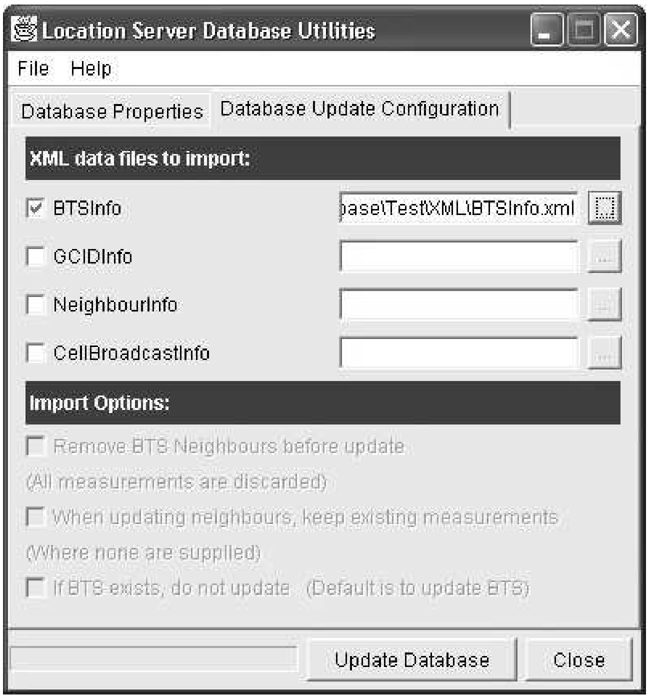
[2] See `http://www.postgresql.org`

H.3.3 Database Utility Applications

The database requires a mechanism for network data to be initially installed and future updates to be applied. Raw data files were provided by the network operator, however, the format often changed between each update requiring a common format for updating the database. The following set of database utilities was developed to facilitate this functionality:

Network Operator Data to XML Converter. Facilitates the conversion of raw data files provided by the network operator into our XML exchange format. A conversion program was developed such that the constantly changing network data obtained from the network operator can be converted into a common XML format, recognized by the database import utility. This is advantageous in that it only requires modification of the appropriate conversion class, rather than modification of the more complex database import utility.

XML Database Import Utility. This utility provides a mechanism to import the XML files into the database. This tool is illustrated in Figure H.3.



Location Server Database Utilities
File
Help
Database Properties
Database Update Configuration
Update Database
Close

XML data files to import:

☑ BTSInfo ase\Test\XML\BTSInfo.xml

□ GCIDInfo

□ NeighbourInfo

□ CellBroadcastInfo

Import Options:

□ Remove BTS Neighbours before update
(All measurements are discarded)

□ When updating neighbours, keep existing measurements
(Where none are supplied)

□ If BTS exists, do not update (Default is to update BTS)

Figure H.3: XML Database Import Utility

H.3.4 GSM Interface Component

The GSM Interface provides the connection to a GSM mobile phone for the acquisition of measurements. A set of Java interfaces defined the common functionality required for the acquisition of measurements from a GSM handset. The implementation classes are dynamically instantiated at runtime by the MPC servlet (Refer to § H.3.5).

Classes implementing the interface were developed for the Nokia 5110 mobile phone, handling issues such as serial communication and physical data acquisition. An open-source project, gnokii[3], provides an interface to common Nokia mobile phones facilitating the acquisition of network monitor data and basic SMS and phonebook management. The gnokii project was originally developed in ANSI C for Linux, but has subsequently been ported to other platforms including Windows. This set of C interfaces needed to be accessed in Java, requiring that a library was built containing the gnokii code. The gnokii functions are made available to Java classes by using Java Native Interface (JNI) declarations in the library.

H.3.5 MPC Server Component

The MPC prototype was developed as location services were not implemented on any Australia network operator. The prototype version obtains the Timing Advance and Cell Global Identifier (CGI) from a serial interface to the GSM mobile phone. It is implemented as a Java servlet on Apache Tomcat[4]. Apache Tomcat supports secure sockets layer (SSL) for secure connections to the MPC server.

The XML-based Mobile Positioning Protocol (MPP) version 3.0 is implemented by the servlet. The servlet instantiates the GSM interface as detailed in section § H.3.4.

The MPC server calculates location by correlating measurements with the database to produce the following attributes constituting a location arc for a given mobile station:

Point of origin. This geographic point at which the cell is located, determined by a database lookup of the following information given the CGI.

- Easting of the BTS within a given zone
- Northing of the BTS within a given zone
- Zone

[3] See `http://www.gnokii.org/`

[4] See `http://jakarta.apache.org/tomcat/`

Start Angle. The start angle of a cell sector in which the mobile station is located. This is determined by performing a database lookup for the antenna azimuth and beamwidth for a given GCID and performing the following calculation: $angle_{start} = antenna\_azimuth - \left(\frac{1}{2}antenna\_beamwidth\right)$

Stop Angle Similarly to the start angle, a database lookup is performed and the following is calculated: $angle_{start} = antenna\_azimuth + \left(\frac{1}{2}antenna\_beamwidth\right)$

Inner Radius. The inner radius is the radius of the arc less the maximum measurement error. This is determined by performing the following calculation:
$inner\_radius = TA \cdot TA\_distance - \left(\frac{1}{2}TA\_distance\right)$ where
$TA\_distance = 550m$ per bit of TA. Refer to § 4.4 for more details on timing advance calculation.

Outer Radius. Similarly to outer radius, the following calculation is performed:
$outer\_radius = TA \cdot TA\_distance + \left(\frac{1}{2}TA\_distance\right)$ where
$TA\_distance = 550m$

Appendix I

This figure is not available online.
Please consult the hardcopy thesis
available from the QUT Library

This figure is not available online.
Please consult the hardcopy thesis
available from the QUT Library

Acronym List

A-GPS	Assisted-GPS
AUC	Authentication Center
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
C/A Code	Coarse Acquisition Code
CBC	Cell Broadcast Center
CBCH	Cell Broadcast Channel
D-GPS	Differential GPS
DRM	Digital Rights Management
E-OTD	Enhanced-Observed Time Difference
EGNOS	European Geostationary Navigation Overlay Service
GBAS	Ground Based Augmentation System
GCI	Cell Global Identifier
GMLC	Gateway Mobile Location Center
GPS	Global Positioning System
GSM	Global System for Mobile telecommunications
IMSI	International Mobile Subscriber Identity

KDP	Key Data Processor
KLIF	KDP Loading and Installation Facility
LMU	Location Measurement Unit
MLC-PCF	MLC Positioning Calculation Function
MOT	Mobile Observed Time
MPP	Mobile Positioning Protocol
MS	Mobile Station
MSAS	Multifunctional Transport Satellite Augmentation Systems
NMEA	National Marine Electronics Association
NTP	Network Time Protocol
P Code	Precision Code
RACH	Random Access Channel
RTCA	Radio Technical Commission for Aeronautics
RTD	Real Time Difference
SA	Selective Availability
SACCH	Slow Associated Control Channel
SBAS	Space Based Augmentation System
SCH	Synchronization Channel
SIM	Subscriber Identity Module
SIR	Signal to Interference Ratio
SMLC	Serving Mobile Location Center
SMSCB	Short Message Service Cell Broadcast
TA	Timing Advance

TDMA	Time Division Multiple Access
TOA	Time Of Arrival
UCON	Usage Control
WAAS	Wide Area Augmentation Service
WAP	Wireless Application Protocol

Bibliography

[1] T. K. Adams. GPS Vulnerabilities. In *Military Review: Information-Age Warfare*, volume LXXXI, pages 10–16. March-April 2001.

[2] Jalal Al-Muhtadi, Manish Anand, M. Dennis Mickunas, and Roy H. Campbell. Secure Smart Homes using Jini and UIUC SESAME. Uiucdcs-r-99-2142, University of Illinois at Urbana Champaign, December 1999.

[3] Jalal Al-Muhtadi, Anand Ranganathan, Roy Campbell, and M. Dennis Mickunas. Cerberus: A Context-Aware Security Scheme for Smart Spaces, March 2003.

[4] Myla Archer. Proving Correctness of the Basic TESLA Multicast Stream Authentication Protocol with TAME. In *Workshop on Issues in the Theory of Security (WITS)*, January 2002.

[5] Australian Communications Authority. Record of Radiocommunications Licences. Database, October 2001.

[6] Australian Communications Authority. Radiocommunications (Spread Spectrum Devices) Class Licence 2002. Australian Communication Austhority, November 2002.

[7] Paramvir Bahl and Venkata N Padmanabhan. Radar: An In-building RF Based User Location and Tracking System. In *IEEE INFOCOM*, pages 775–784, March 2000.

[8] E. Barkan, E. Biham, and N. Keller. Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication. In *The 23rd Annual International Cryptology Conference (Crypto 2003)*, August 2003.

[9] John Bellardo and Stefan Savage. 802.11 denial-of-service attacks: Real vulnerabilities and practical solutions. In *Proceedings of the 11th USENIX Security Symposium*, pages 15 – 28. USENIX, August 2003.

[10] Steven M. Bellovin and Michael Merritt. Limitations of the Kerberos Authentication System. In *USENIX*, 1991.

[11] Elisa Bertino, Piero Andrea Bonatti, and Elena Ferrari. TRBAC: A Temporal Role-based Access Control Model. In *Fifth ACM workshop on Role-based access control*, 2000.

[12] Alex Biryukov, Adi Shamir, and David Wagner. Real Time Cryptanalysis of A5/1 on a PC. In *Fast Software Encryption Workshop*, April 2000.

[13] N. Borisov, I. Goldberg, and D. Wagner. Intercepting mobile communications: The insecurity of 802.11. In *Proceedings of the Seventh Annual International Conference on Mobile Computing and Networking (MOBICOM-01)*, pages 180–188, New York, July 16–21 2001. ACM Press.

[14] John Brezak. The Windows 2000 RC4-HMAC Kerberos encryption type. Internet Draft draft-brezak-win2k-krb-rc4-hmac-01, October 1999.

[15] John Brezak. Utilizing the Windows 2000 Authorization Data in Kerberos Tickets for Access Control to Resources. Internet Draft draft-brezak-win2k-krb-authz-00, February 2002.

[16] Marc Briceno and Ian Goldberg. GSM Cloning. `http://www.isaac.cs.berkley.edu/isaac/gsm-faq.html`, April 1998.

[17] Steve Callaghan and Hugo Fruehauf. SAASM and Direct P(Y) Signal Acquisition. *The Journal of Defense Software Engineering*, 16(6):12–16, June 2003.

[18] James Carroll, Karen Van Dyke, and John Kraemer Rodgers Charles. Vulnerability Assesment of the Transportation Infrastructure Relying on the Global Positioning System. Technical report, National Transportation Systems Center, August 2001.

[19] J. L. Caton. We Can Reduce Satellite Vulnerability. In *Proceedings of the U.S. Naval Institute*, pages 81–83, November 1995.

[20] M. Covington, M. Moyer, and M. Ahamad. Generalized Role-Based Access Control for Securing Euture Applications, 2000.

[21] Michael J. Covington, Prahlad Fogla, Zhiyuan Zhan, and Mustaque Ahamad. A Context-aware Security Architecture for Emerging Applications. In *Proceeding of the Annual Computer Security Applications Conference (ACSAC)*, December 2002.

[22] Michael J. Covington, Wende Long, Srividhya Srinivasan, Anind Dey, Mustaque Ahamad, and Gregory Abowd. Security Context-Aware Applications Using Environment Roles. In *Proceedings of the 6th ACM Symposium on Access Control Models and Technologies (SACMAT '01)*, May 2001.

[23] G. K. Crosby, W. S. Ely, K. W. McPherson, J. M. Stewart, D. K. Kraus, and T. P. Cashin. A Ground-based Regional Augmentation System (GRAS) - The Australian Proposal. In *ION GPS2000*, Salt Lake City UT, 2000.

[24] Dorothy E. Denning and Peter F. MacDoran. Location-based Authentication: Grounding Cyberspace for Better Security. In *Computer Fraud and Security*, pages 167–174. Elsevier Science Ltd., February 1996.

[25] Department of Defense. *Report of the Defense Science Board Task Force on Tactical Air Warfare*. Department of Defense, United States of America, Washington D.C., 1993.

[26] Goran M. Djuknic and Robert E. Richton. Geolocation and Assisted GPS. In *IEEE Computer*, volume 34, pages 123–125. February 2001.

[27] R. Duke, G. Rose, and G. Smith. Object-Z: A Specification Language Advocated for the Description of Standards. *Computer Standards & Interfaces*, 17(5–6):511–533, 1995.

[28] Stephen Farrell Eric Baize and Tom Parker. The SESAME V5 GSS-API Mechanism, November 1996.

[29] Ericsson. Mobile Positioning Protocol Version 5.0; GMPC 6.0, June 2003.

[30] European Computer Manufacturer's Association. Authentication and Privilge Attribute Security Application with Related Key Distribution Function, March 1996.

[31] European Space Agency. Galileo - Mission High Level Definition. Technical report, European Space Agency, 2001.

[32] European Space Agency. Galileo: The European Programme for Global Navigation Service. Technical report, European Space Agency, 2003.

[33] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Location registration procedures (GSM 03.12 version 7.0.0 Release 1998)*. European Telecommunications Standards Institute, August 1999.

[34] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Signalling requirements relating to routeing of calls to mobile subscribers (GSM 03.04 version 6.0.0 Release 1997)*. European Telecommunications Standards Institute, April 1999.

[35] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Unstructured Supplementary Service Data (USSD) - Stage 2 (GSM 03.90 version 7.0.0 release 1998)*. European Telecommunications Standards Institute, August 1999.

[36] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Base Station System - Mobile-services Switching Centre (BSS - MSC) interface; Interface principles (GSM 08.02 version 8.0.0 Release 1999)*. European Telecommunications Standards Institute, June 2000.

[37] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Mobile radio interface; Layer 3 specification (GSM 04.08 version 7.10.0 Release 1998)*. European Telecommunications Standards Institute, December 2000.

[38] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Mobile radio interface layer 3 specification; Radio Resource Control Protocol (GSM 04.18 version 8.4.1 Release 1999)*. European Telecommunications Standards Institute, October 2000.

[39] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Organization of subscriber data (GSM*

03.08 version 7.3.0 Release 1998). European Telecommunications Standards Institute, June 2000.

[40] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Radio subsystem link control; (GSM 05.08 version 8.7.1 Release 1999)*. European Telecommunications Standards Institute, November 2000.

[41] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Radio subsystem synchronization (GSM 05.10 version 8.4.0 Release 1999)*. European Telecommunications Standards Institute, August 2000.

[42] European Telecommunications Standards Institute. *Digital Cellular Telecommunications System (Phase 2+); Location Services (LCS); Broascast Network Assistance for Enhanced Observed Time Difference (E-OTD) and Global Positioning System (GPS) Positioning Methods*. European Telecommunications Standards Institute, 2000.

[43] European Telecommunications Standards Institute. *Digital Cellular Telecommunications System (Phase 2+); Location Services (LCS); (Functional Description) - Stage 2*. European Telecommunications Standards Institute, 2000.

[44] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+); Network architecture (GSM 03.02 version 7.1.0 Release 1998)*. European Telecommunications Standards Institute, February 2000.

[45] European Telecommunications Standards Institute. *Digital Cellular Telecommunications System (Phase 2+); Technical realization of Cell Broadcast Service (CBS)*. European Telecommunication Standards Institute, 2000.

[46] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+); Universal Geographical Area Description (GAD)*, gsm 03.2 version 7.1.0 release 1998 edition, November 2000.

[47] European Telecommunications Standards Institute. *Universal Mobile Telecommunications System (UMTS); 3G Security; Security Principles and Objectives (3G TS 33.120) version 3.0.0 Release 1999*. European Telecommunications Standards Institute, 2000.

[48] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+) (GSM); Security related network functions (GSM 03.20 version 8.1.0 Release 1999)*. European Telecommunications Standards Institute, July 2001.

[49] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+); Security Aspects*, gsm 02.09 version 8.0.1 release 1999 edition, June 2001.

[50] European Telecommunications Standards Institute. *Digital cellular telecommunications system (Phase 2+); Security related network functions*, gsm 03.20 version 8.1.0 release 1999 edition, July 2001.

[51] D. B. Faria and D. R. Cheriton. DoS and authentication in wireless public access networks. In *Proceedings of the ACM Workshop on Wireless Security (WiSe-02)*, pages 47–56, New York, September 28 2002. ACM Press.

[52] David F. Ferraiolo, Ravi Sandhu, Serban Gavrila, D. Richard Kuhn, and Ramaswamy Chandramouli. Proposed NIST Standard for Role-Based Access Control. *ACM Transactions on Information and System Security*, 4:224–274, August 2001.

[53] S. Fluhrer, I. Mantin, and A. Shamir. Weaknesses in the key scheduling algorithm of RC4. In *8th Annual Workshop on Selected Areas in Cryptography*, volume 2259 of *Lecture Notes in Computer Science*, pages 1 – 24, Toronto, Canado, August 2001. Springer-Verlag, Berlin Germany.

[54] Radio Technical Commission for Maritime Services. *RTCM Recommended Standards for Differential Navstar GPS Service, Version 2.2, RTCM Special Committee No. 104*, version 2.2 edition, January 1998.

[55] Eran Gabber and Avishai Wool. How to Prove Where You Are: Tracking the Location of Customer Equipment. In *ACM Conference on Computer and Communications Security*, pages 142–149, 1998.

[56] Galilei Consortium. The Galilei Project: GALILEO Design Consolidation. European Comission, August 2003.

[57] Gary Gaskell. Integrating Smartcards into Kerberos . Master's thesis, Data Communications, Faculty of Information Technology, Queensland University of Technology, February 2000.

[58] Allan Harbitter and Daniel A. Menasce. The Performance of Public Key-enabled Kerberos Authentication in Mobile Computing Applications. In *ACM CCS'01*, November 2001.

[59] Guenter W. Hein and et. al. Status of Galileo Frequency and Signal Design. Technical report, Galileo Signal Task force of the European Commission, Brussels, September 2002.

[60] Hoffmann-Wellenhof, B. H. Lichtenegger, and J. Colins. *GPS: Theory and Practice*. Springer-Verlag, New York, 3rd edition, 1994.

[61] IEEE-SA Standards Board. IEEE Std 802.11-1999 Information Technology-Telecommunications And Information exchange Between Systems-Local And Metropolitan Area Networks-specific Requirements-part 11: Wireless Lan Medium Access Control (MAC) And Physical Layer (PHY) Specifications. Technical report, IEEE, 1999.

[62] International Organization for Standardization. *ISO/IEC 9798-3 Information technology - Security techniques - Entity authentication - Part 3: Mechanisms using digital signature techniques*, 1998.

[63] International Telecommunication Union. International Numbering Resources. `http://www.itu.int/ITU-T/inr/index.html`, October 2001. Accessed 13 October, 2001.

[64] Naomaru Itoi and Peter Honeyman. Smartcard Integration with Kerberos V5. Technical report, University of Michigan, December 1998.

[65] Wendy Johnston and Gordon Rose. Guidelines for the Manual Conversion of Object-Z to C++. Technical report 93-14, Software Verification Research Centre, School of Information Technology, The University of Queensland, Brisbane 4072. Australia, September 1993.

[66] Soon-Kyeong Kim and David Carrington. A Formal Mapping between UML Models and Object-Z Specifications. Technical Report 00-03, Software Verifi-

cation Research Centre, School of Information Technology, The University of Queensland, Brisbane 4072, Australia, February 2000.

[67] A. Kishan, M. Michael, S. Rihan, and R. Biswas. Halibut: An Infrastructure for Wireless LAN Location-Based Services. Technical report, Stanford University, June 2001.

[68] J. Kohl and C. Neuman. *The Kerberos Network Authentication Service (V5)*. Networking Working Group Request for Comments, September 1993.

[69] Richard C. Linger, Howard F. Lipson, John McHugh, Nancy R. Mead, and Carol A. Sledge. Life-Cycle Models for Survivable Systems. Technical Report CMU/SEI-2002-TR-026, Carnegie Mellon Software Engineering Institute, October 2002.

[70] Peter F. MacDoran. *Method and Apparatus for Authenticating the Location of Remote Users of Networked Computing Systems*, May 1998. United States Patent 5757916.

[71] A. Menezes, P. van Oorschot, and S. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

[72] Alan O'Connor Michael Gallaher and Brian Krop. The Economic Inpact of Role-Based Access Control. Technical report, National Institute of Standards and Technology, March 2002.

[73] S. P. Miller, B. C. Neuman, J. I. Schiller, and J. H. Saltzer. Project Athena Technical Plan: Kerberos Authentication and Authorization System. Technical report, Massachuttes Institute of Technology, 1988.

[74] National Marine Electronics Association. *NMEA 0183 Standard for Interfacing Marine Electronic Devices*, 2.20 edition, January 1997.

[75] Aleksandar Neskovic, Natasa Neskovic, and George Paunovic. Modern Approaches in Modeling of Mobile Radio Systems Propagation Environment. *IEEE Communications Surveys*, Third Quarter:2–12, 2000.

[76] B. Clifford Neuman and Stuart G. Subblebine. A Note on the Use of Timestamps as Nounces. In *Operating Systems Review*, chapter 2, pages 10–14. April 1993.

[77] Frank O'Dwyer. Feasibility of attacking Windows 2000 Kerberos Passwords, March 2002.

[78] Jaehong Park and Ravi Sandhu. Towards Usage Control Models: Beyond Traditional Access Control. In *SACMAT'02*, June 2002.

[79] A. Perrig, R. Szewczyk, V. Wen, D. Cullar, and J. Tygar. SPINS: Security protocols for sensor networks. In *Proceedings of MOBICOM 2001*, 2001.

[80] Adrian Perrig, Ran Canetti, J. D. Tygar, and Dawn Song. The TESLA Broadcast Authentication Protocol. *Cryptobytes*, 5(2):2–13, Summer/Fall 2002.

[81] Josyula R. Rao, Pankaj Rohatgi, and Helmut Scherzer. Partitioning Attacks: Or How to Rapidly Clone Some GSM Cards. In *IEEE Syposium on Security and Privacy*, May 2002.

[82] Theodore S. Rappaport. *Wireless Communications Principles and Practice*. Prentice-Hall, Inc., 1996.

[83] Hal L. Feinstein Ravi S. Sandhu, Edward J. Coyne and Charles E Youman. Role-Based Access Control Models. *IEEE Computer*, 29(2):38–47, February 1996.

[84] Rusty Russell and Harald Welte. Linux netfilter Hacking HOWTO, August 2003.

[85] Siddhartha Saha, Kamalika Chaudhuri, Dheeraj Sanghi, and Pravin Bhagwat. Location Determination of a Mobile Device Using IEEE 802.11b Access Point Signals, 2001.

[86] Ravi Sandhu and Jaehong Park. Usage Control: A Vision for Next Generation Access Control. In *MMM-ACNS*, 2003.

[87] Logan Scott. Anti-Spoofing and Authenticated Signal Architectures for Civil Navigation Systems. In *Institute of Navigation GNSS 2003*, pages 1543–1552, Portland, OR, USA, September 2003.

[88] Asim Smailagic, Daniel P. Siewiorek, Joshua Anhalt, David Kogan, and Yang Wang. Location Sensing and Privacy in a Context-aware Computing Environment. volume 9, pages 10–17, 2002.

[89] Asim Smailagic, Jason Small, and Daniel P. Siewiorek. Determining User Location for Context-aware Computing Through the Use of a Wireless LAN Infrastructure, December 2000.

[90] Graeme Smith. An Introduction to Object-Z. Technical report, Software Verification Research Centre, University of Queensland, Australia, 2000.

[91] Mika Ståhlberg. Radio Jamming Attacks Against Two Popular Mobile Networks, November 2000.

[92] Paul Stergiou and David Kalokitis. Keeping the Lights On: GPS and Power Grid Intermesh. *GPS World*, November 2003.

[93] A. Stubblefield, J. Ioannidis, and A. D. Rubin. Using the Fluhrer, Mantin, and Shamir attack to break WEP. In *Proceedings of the Symposium on Network and Distributed Systems Security (NDSS 2002)*, San Diego, CA, February 2002. Internet Society.

[94] Mark A. Sturza. Navigation System Integrity Monitoring Using Redundant Measurements. In *Journal of The Institute of Navigation*, volume 35, pages 69–87. October 1988.

[95] Texas Instruments. The Effects of Adjacent Channel Rejection and Adjacent Channel Interference on 802.11 WLAN Performance. White Paper SPLY005, November 2003.

[96] The Open Group. Distributed Computing Environment 1.2.2. `http://www.opengroup.org`, 1997.

[97] James Bao-Yen Tsui. *Fundamentals of Global Positioning System Receivers: A Software Approach*. John Wiley and Sons, Inc., 2000.

[98] B. Tung, C. Neuman, M. Hur, A. Medvinsky, S. Medvinsky, J. Wray, and J. Trostle. *Public Key Cryptography for Initial Authentication in Kerberos*. Kerberos WG Working Group of the IETF, 16 edition, September 2002.

[99] U.S. Department of Transportation United States Coast Guard. *Broadcast Standard for the USCG DGPS Navigation Service*, comdtinst m16577.1 edition, April 1993.

[100] Eric W. Weisstein. Jordan Curve Theorem. http://mathworld.wolfram.com/JordanCurveTheorem.html, March 2004.

[101] F. Whitwam. Integration of Wireless Network Technology with Signaling in the Rail Transit Industry. pages 1 – 7, 2003. Alcatel Telecommunications Review.

[102] Wireless Application Protocol Forum. *WAP TLS Profile and Tunneling; Version 11-Apr-2001; Wireless Application Protocol*. Wireless Application Protocol Forum, April 2001.

[103] Wireless Application Protocol Forum. *Wireless Identity Module; Part: Security; Version 12-July-2001; Wireless Application Protocol*. Wireless Application Protocol Forum, July 2001.

[104] Wireless Application Protocol Forum. *WMLScript Crypto Library; Version 20-Jun-2001; Wireless Application Protocol*. Wireless Application Protocol Forum, June 2001.

[105] WorldPay. WorldPay deploy GPS technology to deliver pinpoint security for Business to Business transactions and launch WorldPay Genesis. http://www.worldpay.co.kr/kr/news/2001/news_genisis.shtml, 2001. Retrieved September, 2003.

[106] Thomas Wu. A Real-World Analysis of Kerberos Password Security. In *Network and Distributed System Security*, February 1999.