A HYBRID FUZZY/GENETIC ALGORITHM FOR INTRUSION DETECTION IN RFID SYSTEMS

by

Gemechu Shonora Geta

Submitted in
partial fulfillment of the requirements for the degree of
Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
November 2011

DALHOUSIE UNIVERSITY

FACULTY OF COMPUTER SCIENCE

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "A HYBRID FUZZY/GENETIC ALGORITHM FOR INTRUSION DETECTION IN RFID SYSTEMS" by Gemechu Shonora Geta in partial fulfillment of the requirements for the degree of Master of Computer Science.

Dated: November 16, 2011

Co-Supervisor: _____

Co-Supervisor: _____

Reader: _____

DALHOUSIE UNIVERSITY

DATE: November 16, 2011

AUTHOR: Gemechu Shonora Geta

TITLE: A HYBRID FUZZY/GENETIC ALGORITHM FOR INTRUSION DETECTION IN RFID SYSTEMS

DEPARTMENT OR SCHOOL: Faculty of Computer Science

DEGREE: MCSc CONVOCATION: May YEAR: 2012

_____

Signature of Author

Dedication

this thesis is dedicated to my younger brothers and sisters

Table of Contents

List of Figures

List of Tables

List of Abbreviations Used

CMOS	Complementary metal–oxide–semiconductor
CRC	Cyclic Redundancy Check
D.C	Don't Care
DoS	Denial of Service
DR	Detection Rate
EPC	Electronic Product Code
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GA	Genetic Algorithm
GBML	Genetics-Based Machine Learning
HF	High Frequency
IC	Integrated Circuit
ID	Identification
LF	Low Frequency
LFP	Location Frequency Profile
MD5	Message Digest 5
PIN	Personal Identification Number
PRNG	Pseudo Random Number Generator

QR-Code	Quick Response Code
RF	Radio Frequency
RFID	Radio Frequency Identification
SHA-1	Secure Hash Algorithm-1
TFP	Time Frequency Profile
TN	True Negative
TP	True Positive
UHF	Ultra High Frequency

Abstract

Various established and emerging applications of RFID technology have been and are being implemented by companies in different parts of the world. However, RFID technology is susceptible to a variety of security and privacy concerns, as it is prone to attacks such as eavesdropping, denial of service, tag cloning and user tracking. This is mainly because RFID tags, specifically low-cost tags, have low computational capability to support complex cryptographic algorithms. Tag cloning is a key problem to be considered since it leads to severe economic losses. One of the possible approaches to address tag cloning is using an intrusion detection system. Intrusion detection systems in RFID networks, on top of the existing light-weight cryptographic algorithms, provide an additional layer of protection where other security mechanisms may fail. This thesis presents an intrusion detection mechanism that detects anomalies caused by one or more cloned RFID tags in the system. We make use of a Hybrid Fuzzy Genetics-Based Machine Learning algorithm to design an intrusion detection model from RFID system-generated event logs. For the purpose of training and evaluation of our proposed approach, part of the RFID system-generated dataset provided by the University of Tasmania's School of Computing and Information Systems was used, in addition to simulated datasets. The results of our experiments show that the model can achieve high detection rates and low false positive rates when identifying anomalies caused by one or more cloned tags. In addition, the model yields linguistically interpretable rules that can be used to support decision making during the detection of anomaly caused by the cloned tags.

Acknowledgements

First, I would like to thank my supervisor Dr. Srinivas Sampalli and my co-supervisor Dr. Denis Riordan, Dr. Sampalli, for his guidance, encouragement and support during the entire length of my program; and Dr. Riordan for his advice, support and important feedback during the program. Additionally, I would like to express my sincere gratitude to Dr. Malcolm Heywood for reading my thesis and providing valuable feedback, and Dr. Stephen Brooks for chairing my thesis's committee.

Next I would like to thank my family. I would like to thank my father and mother for always providing me their consistent love and support which has enabled me to get to where I am today. I am also grateful to all the other members of my family who stuck by me during the challenging years.

Next I would like to thank all the members of WISE RFID lab for their important comments and suggestions to my thesis. Special thanks to Raghav Sampangi, Erik Wibowo, Ashraf Mohammed Iqbal and Komalesh Narayan for proof reading the initial draft of this thesis; and Diana Paterson for proofreading the abstract, and the acknowledgements. I am also thankful to Dalhousie University's Writing Center for their important comments on my writing. I also want to extend my thanks to all the staff and faculty members of the Faculty of Computer Science for their valuable assistance and encouragement.

I would also like to extend my special thanks to Dr. Oudessa Kerro for his advice and encouragement, especially during the initial phases of my degree.

Finally, I would like to thank all who made the completion of this thesis possible! And above all, I would like to give glory to the lord God who surrounded me with supportive and valuable people!

Chapter 1: Introduction

In this chapter, we briefly discuss Radio Frequency Identification (RFID) technology and its current trends, the research problem addressed by the thesis, thesis motivation, and an overview of other approaches. Background and reviews of other approaches are discussed in detail in the chapters 2, 3 and 4. We will start by giving an overview of an RFID system, and then we will discuss fuzzy logic, genetic algorithms and anomaly detection problems.

1.1 RFID Overview

Radio Frequency Identification (RFID) is a technology used to identify objects by means of radio waves. Despite its current rapid growth and development, RFID is not a new technology. It has been around for more than five decades but its potential applications are not fully exploited. The concept of using radio waves to identify objects dates back to the beginning of the twentieth century, 1906 [Landt 2005]. RFID technology in its modern sense was introduced in 1948 [Stockman 1948].

A typical RFID system consists of an RFID tag (transponder), reader (interrogator) and RFID application server (host computer). An RFID tag is a programmable device which carries an ID and other data associated with the tagged object. The tag is composed of a coupling element (antenna) for communication and a small CMOS IC Chip for performing logical operations and storage. An RFID tag reader is a component which interrogates tags for their content and transfers information to an RFID application server at the back end. The reader is composed of storage, signal processing and controller units, RF transceiver and a serial interface to the backed application server. Middleware is the software component that converts raw RFID read data into more understandable high level information. RFID middleware is mostly incorporated into the reader.

RFID systems operate in four common frequency ranges: Low Frequency (LF), High Frequency (HF), Ultra-High Frequency (UHF), and Microwave Frequency range. LF and HF frequency range RFID systems have read distance of 10-20 centimeter for passive tags and the other two have read ranges of up to 3 meters for passive tags.

As the size of RFID tags get smaller, their price cheaper and their read rate improves, the demand to implement RFID system in various agencies as well as small and big businesses has rapidly increased [Karmakar 2010; Landt 2005]. Among the most common applications where RFID is currently being actively utilized are: electronic toll payment systems, car lock keys and engine starters, interactive advertisements, inventory control systems, asset management, animal tracking, electronic passport, automatic library book check-out, in-door object tracking, and patient safety or care in hospital environment. The use of the technology with smart phones for mobile payment, in-door object tracking as well as other location aware services are also gaining more public attention [Mark 2011; Dave 2011; Zhang et al. 2011; Richard 2010].

The use of RFID systems in various agencies or business organizations pays back in terms of low operational cost, less processing delay and higher visibility of an object [Motorola Inc. 2007]. RFID is also believed to be one of the key enabling technologies behind the "Internet of Things". The "Internet of Things" is a larger context of the current internet where not only computers and other peripherals are connected but other real world objects are also connected so as to receive, transmit and share information and other resources [Gerenshenfeld 2004].

1.2 Fuzzy Logic, Genetic Algorithms and Anomaly Detection

Fuzzy logic is an extension of the classical logic, which was first introduced in 1973 by Zadeh [1975] following the introduction of fuzzy set theory in 1965 [Zadeh 1965]. A Proposition in fuzzy logic is allowed to bear two or more truth values unlike classical logic where a proposition can only be either true or false. Fuzzy logic handles real world imprecision and uncertainty by allowing partial truth values. Similar to human reasoning, fuzzy logic precisely reasons imprecise or uncertain (vague) conditions by using linguistic terms.

Genetic Algorithms are evolutionary search strategies, which add learning and adaptation capability to fuzzy logic's approximate reasoning power. Genetic algorithms follow the Darwinian principle of natural selection, where only the fittest individuals evolve into the next generation while the unfit cease to exist [Holland 1992]. This principle is commonly known as the survival of the fittest. Genetic algorithms have been used for generating fuzzy rules from data [Ishibuchi et al. 1992; Wang and Mendel 1992], optimization of fuzzy rulebases [Ishibuchi 2007], as well as for automatic tuning of membership functions.

Anomaly based intrusion detection systems provide security when other security protocols implemented in the system may fail. Various anomaly intrusion detection systems have been proposed and implemented. Such systems play a major role in monitoring the traffic in the system and report or alert when unrecognized patterns are observed within the system. Applying fuzzy genetics-based machine learning method to the intrusion detection problem has been shown to be successful in providing high performance in LANs (KDD-cup dataset) [Abadeh et al. 2007].

1.3 The Research Problem

Object identity cloning is a serious issue due to its global social and economic impact. According to Aberdeen research group [Aberdeen Group 1996-2011], worldwide identity theft losses were $73.8 billion in 2002, reaching $2 trillion by the end of the year 2005 [Jim 2003], a number expected to increase in the years to come.

According to a report by the World Health Organization (WHO), the number of counterfeit products entering global markets has been increasing every year [WHO 2008]. The number of counterfeiting incidents in 2007 has increased by ten-fold than it was in the year 2000. Counterfeiting of medical products imposes serious health risks or even death to patients and causes high economic losses to the industries producing genuine products.

RFID is a promising technology used to combat object identity cloning in applications such as pharmaceutical supply chains because it increases track and trace visibility [Staake et al. 2005] [King 2007]. RFID is also being increasingly used in various applications such as automatic payment, where high security is required.

For an RFID system to be successfully employed in a ubiquitous environment, the cost of the tag is one of the major factors to consider [Wu et al. 2006]. RFID tags have to be low cost devices to be viable in many applications. Some estimates suggest that with a large scale production, the cost of EPC tag can drop below five cents. In 2001, researchers at AUTO-ID lab proposed the design for five cent RFID tag, which they pointed it to be difficult but achievable without the need to scale the production of the tag [Sarma 2002]. There are also some successes in producing printed RFID tags, which is estimated to reduce the current cost of silicon-based tags by more than ten-fold [Mary 2010].

However, due to the trade-off between the cost and computational capability, securing Low-Cost RFID systems has been a remarkable challenge in the research community. This is mainly because reduction in tag storage and processing capacity limits the ability of the tag to provide enough resources for execution of computationally intensive but more secure cryptographic algorithms.

1.4 Motivation to the Thesis

It has been indicated that implementing intrusion detection systems in RFID (at the middleware and backend level) is a promising approach to serve as an additional layer of security on top of the existing security protocols. For example, Mirowski et al. [2007] implemented statistical intrusion detection mechanism introduced by Denning [1987], while Lehtonen et al.[2007] is based on the probabilistic technique. Most of these approaches achieve high detection rate, however, they come with high false positive rate.

We started by assessing the possibility of implementing an intrusion detection mechanism, based on fuzzy logic, that achieves high detection and low false positive rates on RFID system.

1.5 Previous Approaches

Much of the RFID research over the past few years has been aimed at addressing security and privacy challenges of low-cost tags. For example, Dimitriou [2005] proposed a lightweight authentication protocol in order to protect privacy of users and defend tags against cloning. In this protocol, both readers and tags authenticate each other based on shared secret key, which is refreshed every time the tag is read.

Generally, there are three approaches taken to solve security and privacy issues in RFID system. These are:

Low-cost tag hardware design – involves designing a small size, low-cost and computationally efficient tag hardware manufacturing

Lightweight cryptographic algorithm design – involves designing secure algorithms that can run over a low computational capacity tags

1. Intrusion detection system design – implements mechanisms to detection attacks when they occur in the system

The first approach is the one related to manufacturing of small but efficient hardware technology that can support computationally intensive cryptographic algorithms on tags. This is the area related to re-engineering of the efficient and low-cost tag memory and battery systems.

The second research area is about designing efficient cryptographic algorithms that require less resource for execution. This revolves around designing light weight cryptographic algorithms, which are broadly discussed in the next section (Chapter 4) of this thesis.

The third approach deals with whether it is possible to come up with efficient additional algorithms on the top of the existing low-cost authentication protocols. This provides additional detection layer, without the need to depend on the tag resources, by avoiding the implementation of complex security protocols on the tag rather on the other components (reader, middleware or the backend system).

1.6 Approach of this Thesis

This thesis contributes to the third method described in section 1.5 above. Approach by Mirowski et al. [2007] has achieved relatively high detection rate but it also has high false positive rate. We propose an intrusion detection mechanism that makes use of fuzzy based approximate reasoning and the learning capability of genetic algorithm [Ishibuchi et al. 1992].

In an anomaly detection system, the profile for normal activity of the system is first built from the normal operation. During monitoring, the current observation in the system is compared with the normal profile so as to detect changes in the utilization pattern or the known behavior of the system. The normal and anomalous behaviors of the system have some characteristics that they share, since the anomalous behavior shows the characteristics of normal behavior with some additional patterns. This makes it difficult to draw clear boundaries between the normal and anomalous behaviors within a system, marking the presence of a fuzzy boundary between the two behaviors.

Our hypothesis is that applying fuzzy genetics-based intrusion detection system on an event logs generated by the RFID system will achieve high detection rate and low false positive rate. Figure 1.1 shows the process flow diagram, a hybrid fuzzy genetics-based intrusion detection system applied to RFID system. The details of the proposed procedure are given in Chapter 5.

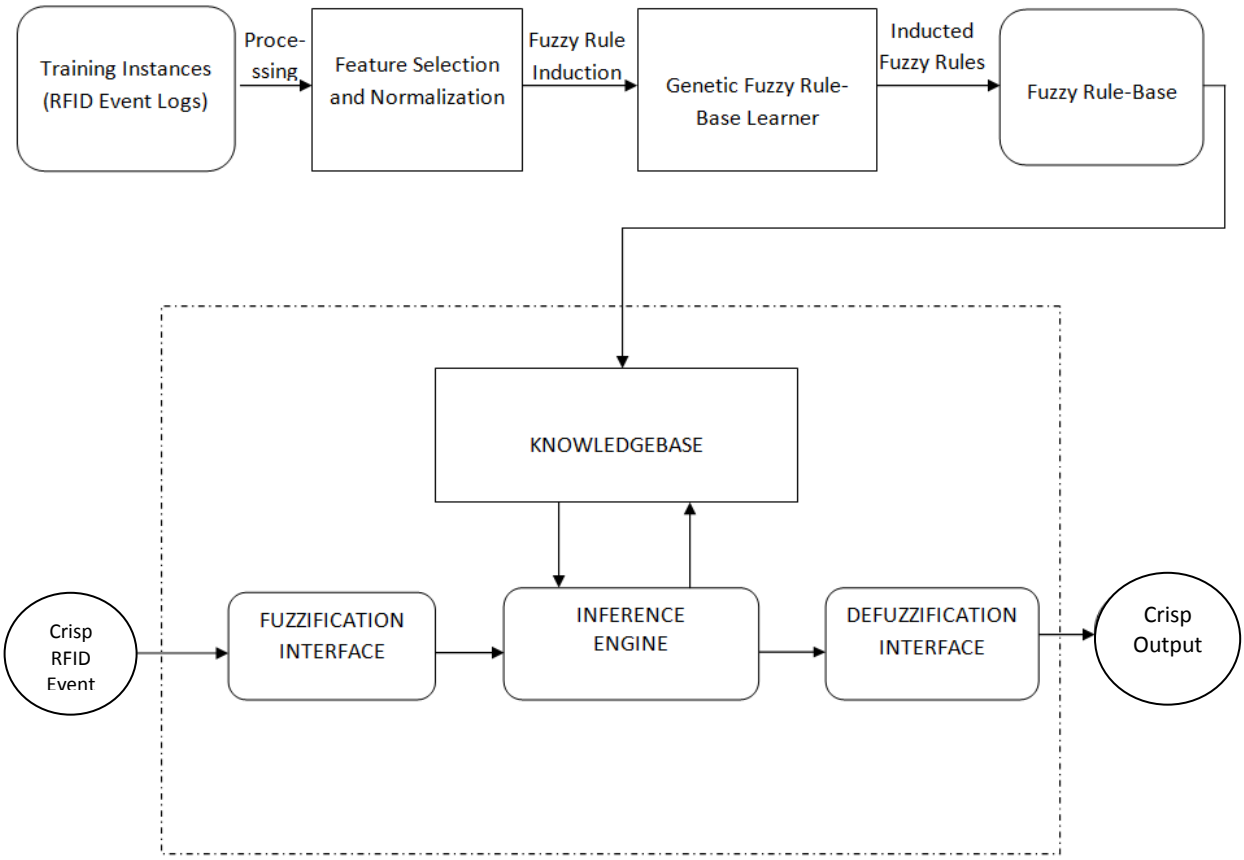

Figure 1.1: Hybrid Fuzzy Genetics-Based RFID Anomaly Intrusion Detection System Diagram

1.7 Summary

This chapter has given a brief overview on RFID systems, its past and the current trends. RFID is a technology which has existed for several years but is still growing with several new applications invented every day. RFID systems, however, are vulnerable to various security and privacy threats. RFID tag cloning is one of the attacks that can be launched on an RFID system and it imposes serious economic and social concerns. There have been a number of approaches taken to combat RFID tag cloning. These approaches come under one of the three research areas which include designing efficient low-cost RFID tag hardware with sufficient computational capability, designing lightweight cryptographic algorithms, and designing intrusion detection systems, security protocols which do not require tag resources.

We have also briefly introduced approaches taken in this thesis to address cloning of RFID system. The next chapter will broadly discuss the background and definition of the technical terms used in this thesis.

1.8 Organization of the Thesis

The rest of the thesis is organized as follows:

Chapter 2 provides background discussion on RFID technology, frequency ranges, architecture of RFID system, key RFID system components and explanations of key RFID terms and concepts used in this thesis. The chapter also provides background material on Fuzzy Systems and Genetic Algorithm.

Chapter 3 discuses about security and privacy issue in RFID system. The common form of attacks RFID system such as cloning, eavesdropping, denial of service, relay, and replay are explained. Discussion on the challenges and issues related to designing security and privacy protocols for RFID system are also presented in this chapter.

Chapter 4 gives literature review covering previous solutions and research directions taken to overcome the challenges in design secure protocols for low-cost RFID systems. This chapter assesses solutions given in designing light-weight/minimalist approaches as well as the intrusion detection approaches presented in literature.

Chapter 5 explains the model used in this thesis to address the problem of anomaly intrusion detection in RFID systems. All the core concepts of the implemented algorithm is presented in this chapter.

Chapter 6 gives description of the dataset used for the evaluation of approached used in this thesis and preprocessing performed on the original dataset.

Chapter 7 provides the results obtained from computational experiment as presented. The presented results are based up on the real world dataset and simulated dataset based up on the observations on the real world dataset.

Chapter 8 concludes the thesis by giving future work that can be done following this thesis.

Chapter 2: Background

In this chapter we will provide background on RFID systems, Fuzzy Logic and Genetic Algorithms. We begin with explanations of RFID system's operating frequency ranges, RFID system architecture, RFID system components, an RFID anti-collision protocols and RFID system applications. Then, explanation on Fuzzy Systems, Fuzzy Set Theory and Fuzzy Logic is provided. Finally, concepts of genetic algorithms and the core techniques involved in applying genetic algorithm are discussed.

2.1 Radio Frequency Identification

Radio Frequency Identification (RFID) is a technology used for automatic identification physical objects (humans, animals, grocery items, etc). It was successfully used in World War II for identification of enemy aircrafts [Landt 2005]. The following sections explain RFID frequency ranges, RFID system components, and the prominent as well as emerging applications of RFID technology.

2.1.2 RFID Frequency Ranges

RFID operates in various frequency ranges. There are generally four major classes of frequency ranges in which RFID operates at Low Frequency (LF), High Frequency (HF), Ultra-high Frequency (UHF) and Microwave Frequency. The following subsections provide brief explanation about each of these classes [Landt 2005].

Low Frequency (LF) RFID Systems

Frequency ranges from 30 KHz to 300 KHz are known as Low Frequency ranges. LF RFID systems operate at 125 KHz and 134.2 KHz. The tags which come under this class are mostly passive tags. LF RFID systems have slow transfer rate and shorter communication distance. The advantage of this system is that it is less affected by the environment. LF RFID devices operate well in harsh environment which contains dust, mud, liquid or metal.

High Frequency (HF) RFID Systems

High Frequency ranges are frequency between 3MHz and 30MHz. The typical HF RFID system operates with frequency of 13.56MHz. Both passive and active RFID tags are used in this HF RFID. Such systems also have slow tag to reader data transfer rate but it performs fairly well in harsh environment such as in the presence of liquid and metal.

Ultra-High Frequency (UHF) RFID Systems

Ultra-high frequency ranges are frequency ranges from 300 MHz to 30 MHz. Typical Passive RFID devices which operate within UHF range have frequency of 915 MHz in North America and 868 MHz. Active RFID tags under this category have operation frequency range of 315MHz in United State and 433MHz in Europe.

UHF RFID systems have poor performance in harsh environments such as in liquids or metals. They use both active and passive tags and it has fast data transfer rate. However, this frequency ranges are not accepted worldwide.

Microwave Frequency RFID System

Microwave RFID systems come under microwave frequency range which is greater than 1 GHz. The typical microwave RFID systems operate at a frequency range of 2.45 GHz and 5.8 GHz. Microwave Frequency RFID systems have the fastest data transfer rate, they perform quite poorly in harsh environments where we have liquids and other metals and they are the smallest size tags. Table 2.1 summarizes classes of RFID systems based on their frequency ranges.

Table 2.1: Classes of RFID System based on Frequency ranges

	Low Frequency (LF)	**High Frequency (HF)**	**Ultra-High Frequency**	**Microwave Frequency**
Frequency Range	125 KHz and 134.2 KHz	13.56MHz	915MHz, 868MHz, 315Mhz, 433MHz	2.45GHz and 5.8GHz
Speed	Slowest	Slow	Slow	Fastest
Environment	Operates well in the presence of liquids and metals	Operates fairly good in the presence of liquids and metals	Operates poor in the presence of liquids and metals	Operates very poor in the presence of liquids and metals
Tags	Mostly passive	Both passive and active	Both passive and active	Passive, semi-passive, active

2.1.2 RFID System Components

RFID Technology typically consists of three basic components: RFID tag (transponder), reader (interrogator) and RFID application server (host computer) [Lehpamer 2007], see Figure 2.1 below. An RFID tag is programmable device which carries an ID and other data associated to the RFID tagged object. The tag is composed of a coupling element (antenna) for communication and a small CMOS IC Chip for performing logical operations and storage. An RFID tag reader is the component which interrogates tags for their content and transfer information to an RFID application server at the backend. The reader is composed of storage, signal processing and controller unit, RF transceiver and a serial interface to the backed application server. Middleware is the software component that converts raw RFID read data into more understandable high level information. RFID middleware is mostly incorporated into the reader.

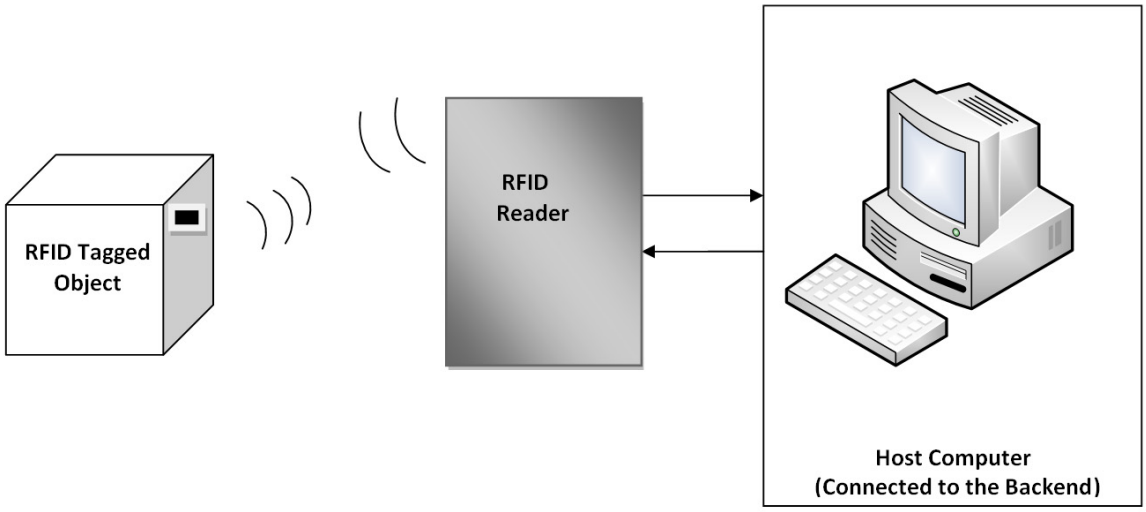

Figure 2.1: The Fundamental Components of RFID System

Communication between the tag and the reader normally starts when the reader interrogates the tag or whenever the tag and the reader come close enough to be in one another's range. When the tag enters reader's range, the reader interrogates the tag and the tag replies with its ID and data stored on its memory. Some applications, however, do not require data stored on the tag. In such cases, the tag replies or reflects only with its ID. The reader receives the response, performs some processing and passes it to the backend host. The backend system uses the tag's unique ID and its associated data to update its database or infer various reasoning depending on the type of enterprise application. The communication between the reader and the backend host is either through serial interface, wired or wireless network.

Tags can either be passive, active or semi-passive depending on how they acquire power for operation. Active tags have integrated on-board power source such as battery, passive tags do not have integrated power source but they are powered by the external power source(usually reader) and the semi-passive tags have integrated power source for powering their chip but depend on the power from external power source (reader) for transmitting information[Landt 2005]. We describe the detail on the next sub-section.

Data from the reader are processed and converted into more meaningful streams of data by using the middleware. Applications and services of the enterprise resides at the backend enterprise servers.

Depending on the functions they offer, RFID tags are divided into different classes [EPCglobal Inc. 2010]. Class-1 tags are solely used for identification. They are passive tags with limited functionality. Class 2 tags are passive tags with some additional functionality than Class-1 tags. They allow tasks such as multiple write and they come with an extended memory. Class-3 tags have an on-board battery to power their microchip. Class 4 tags include active tags. As discussed earlier, active tags can function independent of the power from the reader and they are also able to initiate communication by themselves without interrogation from the reader in the range both with readers and tags. The last class, known as Class 5, has much more functionality than the previous classes of tags. Table 2.2 summarizes the RFID class hierarchy.

Table 2.2: RFID Class Hierarchy modified from [Damith et al. 2005]

Class V
All interrogators
Class IV
Autonomous RFID tags
Class III
EPC Tags with Additional Functionality
Class 0/Class I
Simple EPC read-only label

RFID Tag

RFID tag is the label attached to an object to provide it a unique identity. RFID tag also stores and transmits data associated to the object to the reader by means of radio waves. RFID Tags, also known as transponders, consists of a silicon microchip and an antenna. The silicon microchips are used for storage of information such as tag ID and to perform limited level logical operations. The antennas are used for obtaining power or receiving and responding to the radio signals from the reader. Some RFID tags may also come without memory chips and in such cases RF material's configuration property is used to identify the tag. There are three types of RFID tags: passive, active and semi-passive (also known as semi-active). The following sections provide explanations to these tag types.

Passive Tags

Passive RFID tags do not have any form of power source integrated to its circuit board. They depend on the power emitted by the reader so as to energize it circuits and transmit data to the reader. This type of tag has longer life span and is more resistant to the harsh environment. They can operate well even in the presence of liquid and metal. They are smaller in size and have low cost.

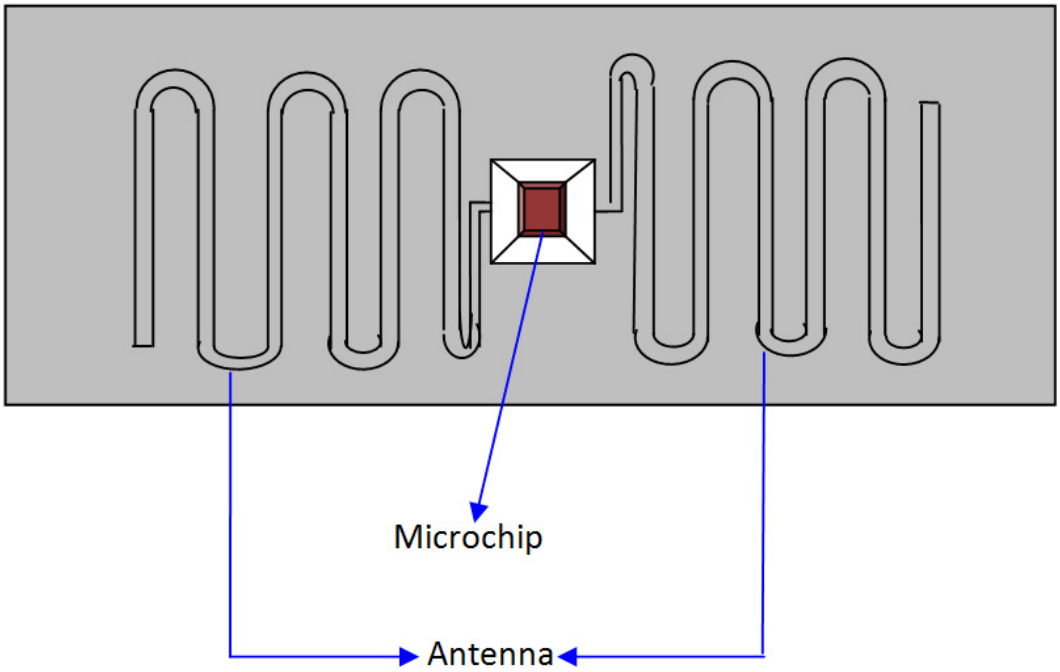

Figure 2.2: Components of Passive RFID tag [Lahiri 2006]

The common passive RFID tag consists of Antenna and Microchip. However, there are some tags, which consist of only an antenna as their identity is determined by the nature of the reflection from the material they are made of.

Microchip: It is a component of RFID tag which consists of memory for storing data, modulator to receive reader signal, the logic unit to implement communication between the tag and the reader, and clock extractor to extract clock signal from the reader's antenna.

Antenna: It is part of a tag which is used to receive and transmit signals. In addition to receiving and transmitting signals, it is also used to draw energy from the reader to energize the tag.

Active Tags

Active tags have an on-board power source, such as battery integrated to them. It doesn't depend on the power from reader, rather it uses power from its own on-board source to energize itself and transmit data.

Active tags do not wait for the reader to start communication as they can initiate communication themselves when there is a reader in their range. In addition to normal tag operation, active tags can also serve as sensor nodes by integrating an additional microchip that makes use of the onboard power source (e.g. temperature sensing). Some active tags save their power by entering sleep mode in the absence of the reader in their range and expect interrogation by the reader to start communication while others keep broadcasting signals even in the absence of the reader resulting on shorter battery life time.

Semi-Passive Tags

Semi-passive tags have an on-board power source like active tags. However, they use it only for energizing their circuits. They depend on power from the reader to transmit data. Semi-passive tags are also called battery-assisted tags or semi-active tags.

RFID Readers

RFID readers, also called interrogators, are the component of RFID systems that are used to read/write data from/to an RFID tag. Readers transmit information read from the tag to the backend system for further processing. Action is taken depending on the type of application.

A typical RFID reader consists of a transceiver, also called reader antenna, to receive and transmit data from and to the tag, a microprocessor to perform certain data processing operations, memory for storage and a power adaptor. Readers perform tasks such as decoding of RF signal, error checking, and filtering of the raw data from tag into high level information. Readers can be stationary, hand-held or integrated with other devices such as mobile phones.

Readers can have one or more antennae so as to communicate and transfer energy to the tag or other readers in the range. While communication between the tag and reader takes place wirelessly by using RF signal, communication channel between the reader and backend systems can either be through serial port, wireless or wired.

RFID Middleware

RFID middleware is the software component of an RFID system. Middleware plays the role between an RFID reader and the backend system. It provides functionalities such as managing of massive data from RFID systems (multiple readers), and sharing of data both within and outside the enterprise.

Backend Enterprise

The backend enterprise encompasses all the enterprise applications and IT infrastructure of the enterprise. It is the data repository system and business processing engine for the enterprise. Backend enterprise systems are usually already built system independent of the RFID systems integration. In Table 2.3 we have summarized comparison between the three types of tags, based on criteria discussed in this section.

Table 2.3: Comparison between passive, semi-passive and active tags

Tag Type	Power Source	Battery Life	Storage, Processing and Read Range	Tag Size	Tag Cost	Reader Cost	EPC Compatibility
Passive	External	Long	Low	Small	Least	High	Compatible
Active	Integrated	Limited	High	Larger	Highest	Least	Not Compatible
Semi-Passive	Both	Limited	Moderate	Largest	Moderate	High	Compatible

2.1.3 RFID Anti-Collision Protocols

Readers differentiate (singulate) a tag from a population of tags by using anti-collision algorithm that manages multiple tags' responses. There have been several anti-collision algorithms proposed for tag singulation. One common algorithm is the deterministic binary tree-walking algorithm [Juels et al. 2003]. In this algorithm, the reader identifies the ID number of individual nearby tags bit-by-bit in a manner that resembles depth first search. The reader first queries all the nearby tags for the next bit of their ID number and if collision is detected in the tags' response, the reader sends a bit indicating the tags that are allowed to be active and those required to be temporarily deactivated. Each bit represents a branch in a binary tree and the leaves represent tags ID numbers.

Another anti-collision algorithm is a probabilistic Slotted Aloha anti-collision scheme [Lee et al. 2005], which is also used as an anti-collision mechanism in local area networks. In this algorithm, tags send responses to the queries of the reader at a random time avoiding collision with other tags' responses. If a collision detected, then the tags will be made to wait for another random amount of time (mostly longer) before they try transmitting again. The disadvantage of this protocol is that it may result in degraded performance of an RFID system in situation where there are a large number of tags which implies more collision probability.

2.1.4 Applications of RFID Technology

RFID systems have been successfully implemented in various applications and the potential application of RFID technology is enormous. Some of the established applications of RFID systems include but not limited to the following:

- Asset monitoring and control
- Inventory management
- Electronic toll payment system
- Access control such as building access and car key lock
- Tracking and tracing of item such as in supplies chain management
- Anti-theft such as library book checkout

- Anti-tampering

In addition to the above stated established applications of RFID, there are also several applications such as anti-counterfeiting and mobile payment systems are also being implemented by a number of companies [Staake et al. 2005; Ondrus 2007].

One of the widely discussed applications of the RFID technology is to use it as a replacement to the current barcode technology. The following subsection provides discussion on the advantages and disadvantages while considering RFID in place of barcode/QR-code systems.

2.1.5 RFID Vs Barcode/QR-code Systems

Replacement of the existing barcode/QR-code technology with an RFID tags is one of the widely spoken application of RFID by stores and several other organizations. Even if RFID has several advantages over the traditional barcode/QR-code systems, there are also certain cases where RFID technology hasn't surpassed the barcode technology to be a perfect replacement. In this section, we will provide a brief comparison between barcodes and RFID systems to clearly point some the key reason why RFID failed to surpass barcode in retail and other industries.

Unlike traditional barcode systems, communication between tag and reader in RFID does not require line of sight and can take place over a longer distance [Lahiri 2006]. RFID readers can also read multiple objects at a time while barcode scanners can scan only one object at a time. In addition, information on RFID tags can be rewritten and RFID is more resistant to harsh environment than traditional barcode systems. These and many other advantages make RFID technology more attractive than barcodes for wider range of applications. On the contrary, barcodes can be produced at relatively no cost that RFID tags. Barcode systems are matured technology with worldwide acceptance, while there are issues, such as security and privacy, raised with RFID technology. The following table provides summary of the comparison between the two technologies.

Table 2.4: Comparison between Barcode/QR-code Systems and RFID [Lahiri 2006]

Barcodes, QR Code Systems	RFID Systems
Requires line of sight	Doesn't require line of sight
Can read only a single object at a time	Can read multiple objects at a time
Limited to communication over a shorter distance	Can communicate over a longer distance
Is read only	Can be rewritten
Can be produce with relatively no cost	Higher tag cost
No intelligence can be integrated	Some intelligence can be integrated(e.g temperature sensing with active tags)
Can easily be destroyed by environmental factors, e.g fading as a result of moist, sun, dust etc	more resistant to harsh environment
Doesn't require special purpose RF device, any image scanner with supporting API works	Requires RFID reader hardware to be added to the existing device hardware infrastructures e.g. Smart-phones
Less read accuracy	Better read accuracy
No item level identification	Item level identification
No social issues because it is widely accepted	Much privacy issues
No legal limitations	Some legal limitations
Can be tagged to any physical object without consideration of its RF property	Affected by RF Lucent and RF Opaque materials

In certain cases, issues such as cost, security and privacy become prevalent than the benefits of replacing traditional barcode or QR-code systems with an RFID system [Lahiri 2006]. For an RFID system to replace barcodes, a significant reduction in RFID tag cost is the major requirement. In addition providing firm solution to the privacy issues and increasing the technology's world wide acceptance are also a requirement.

2.2 Fuzzy Systems

Fuzzy expert systems have been applied to solve problems in various application domains including pattern recognition, control systems etc. Fuzzy systems are built based on the principle of fuzzy set theory where a set may contain elements with full or partial membership.

Fuzzy expert system consists of four major components. These are: Fuzzification interface, Fuzzy Inference engine, Fuzzy Rule Base and Deffuzification interface. See Fig 2.1 below.

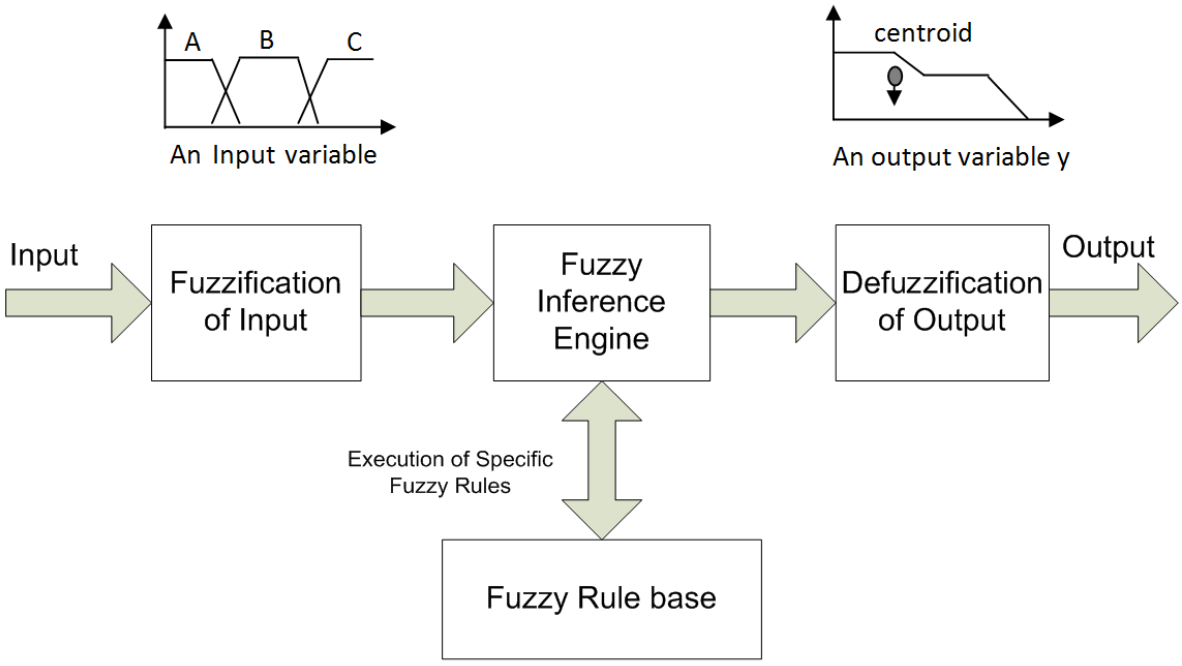

Figure 2.3: The basic components of Fuzzy Inference System [Fulcher 2008, P. 24]

The Fuzzification Interface of the fuzzy inference engine converts the given value of input variable into its fuzzy linguistic equivalent; the fuzzy rule base stores the fuzzy rules associated

with the system, fuzzy Inference Engine uses rules from the rule base to deduce an output to the given antecedent fuzzy inputs, the linguistic output rules (conclusions) from the fuzzy inference engine are converted in to their equivalent crisp (non-fuzzy) output by the defuzzification interface. Various diffuzzification techniques, such as centroid, are usually used to convert the fuzzy output values into its crisp equivalent.

2.2.1 Fuzzy Set Theory

The theory of Fuzzy Sets was first proposed in 1965 by Zadeh [1965]. The fundamental concept in the classical set theory is that an element can have only either of the two membership values to a given set. If an object belongs to a set then it is referred to as a member and non-member otherwise. Classical set theory assumes a sharp, crisp, an unambiguous distinction between a member and non-member of a set. The boundary between a member and non-member is assumed to be clear and precise. This notion of classical set theory holds for both deterministic and stochastic cases where the value used to determine full membership or no membership an object to a given set.

However, the notion of classical set theory cannot describe many real world problems involving elements with only partial membership to a set and therefore, cannot be used to solve such problems. For example, consider the set of tall people. In a classical set theory, if we set a boundary at 1.8 meter, then all people whose height is over 1.8 meter are considered members of the set and all others whose height is below 1.8 are non members. But in a real world, we use linguistic terms to express the degree to which a person is considered too tall and not tall. For instance, if someone's height is 1.79, then we say that person is nearly tall or if someone's height measures 1.91 meter, then we say that person is very tall and so on.

Fuzzy set theory allows partial membership of an object to a given set. In other words, elements of the fuzzy set can be either full member, no member or partial member. Hence, fuzzy sets are generalizations of the notion of classical set theory by allowing partial membership to a set called fuzzy set. Fuzzy set theory is therefore the strict superset of the classical set theory.

> As Zadeh [1965] puts it, "Fuzzy set is a class of objects with a continuum of grade of membership. Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership between zero and one."

The membership of an object to a given fuzzy set is the matter of membership grade. The membership grade of one indicates entire membership while zero indicates non-membership and any values between zero and one represents partial membership. An object can have a membership grade of zero, one or any value between zero and one. Fuzzy set therefore is a superset of the classical set, handling most real world scenarios where there are no clear boundaries between membership and non-membership an object to a given set.

2.2.2 Fuzzy Logic

Fuzzy logic is an extension of the classical logic and was first introduced in 1973 by Zadeh [1975] following the introduction of fuzzy set theory in 1965 [Zadeh 1965]. Classical logic deals with propositions which are either true or false but not in between or both true or false. Only one truth value (either true or false) is allowed in a classical logic. In other words, every statement in classical logic is either true or false but not both or in between.

However, in some real world scenario, there are cases where propositions can be both partially true and partially false. Fuzzy logic handles such real world imprecision and uncertainty by allowing partial truth values. Because multiple values are allowed in fuzzy logic (unlike the two valued classical logic), it is also called multi-valued logic. Fuzzy logic provides the ability to precisely reason imprecise or vague conditions like humans do by using linguistic terms. In a fuzzy logic, fuzzy sets define linguistic notion and the truth value of the given linguistic expression is defined by a membership function. The membership degree for an object in a fuzzy set is defined as a function where the universe of discourse is the domain and the interval [0, 1] is the range.

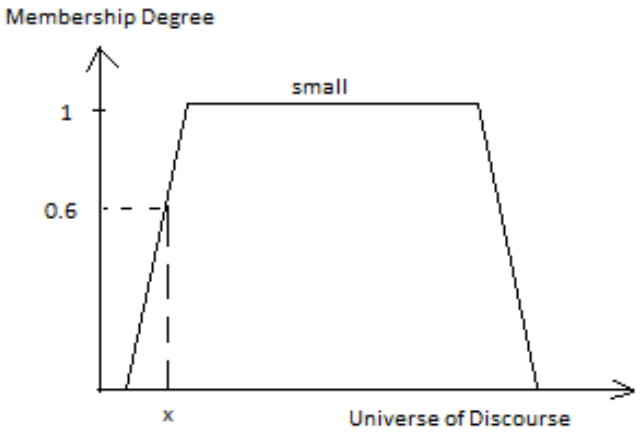

Figure 2.4: Trapezoidal Membership Function for a given universe of discourse

In the Figure 2.4, an object x has a membership degree of 0.6 to the fuzzy set small, which shows x is a member and not member of the fuzzy set small as the same time. An object x is a partial member of another set (which is not indicated here) with the membership degree of 0.4.

As in classical logic, logical operations can be applied to atomic fuzzy expressions to yield resultant fuzzy expression. For each logical operator in a classical logic there is equivalent fuzzy logic operator corresponding to it as summarized in Table 2.5.

Table 2.5: Classical versus fuzzy logical operators [Zadeh 1975]

Classical Logic Operators	**Equivalent Fuzzy Logic Operators**
⌐p	1-p
p v q	Max (p, q)
P ∧ q	Min (p, q)
v (p => q)	Min (1, 1 - p + q)

Fuzzy Rules takes the form: **IF** condition **THEN** consequent [weight]

Where,

"Condition" stands for an atomic fuzzy expression or the combination of atomic expressions interconnected by fuzzy logical operators (such as those given in table 2.5); and

"Consequent" is the resultant or the inferred atomic fuzzy expression. The Rule Weight (or simply Weight) is the value (real number) that refers to the confidence of the associated fuzzy rule.

2.3 Genetic Algorithms

Genetic Algorithm (GA) is an evolutionary search strategy, which is used for optimization of complex problems, especially when the objective function is not smooth, or there are multiple local optima and there are a large number of parameters. Genetic algorithms have shown proven ability in solving complex problems in machine learning, game theory, design automation, evolvable hardware, network security, and bioinformatics. It provides the mechanism to explore possible solutions in a wider solution space by applying crossover, mutation and selection operators to the individual solutions [Goldberg 1989].

Genetic algorithms mimic the way life and intelligence evolves in a natural environment. Like the evolution process in a natural environment, evolution in GA takes place as the result of natural selection and reproduction. GA follows the Darwinian principle of natural selection where only the fittest individuals evolve into the next generation while the unfit ceases to exist (the principle commonly known as the survival of the fittest)[Holland 1992].

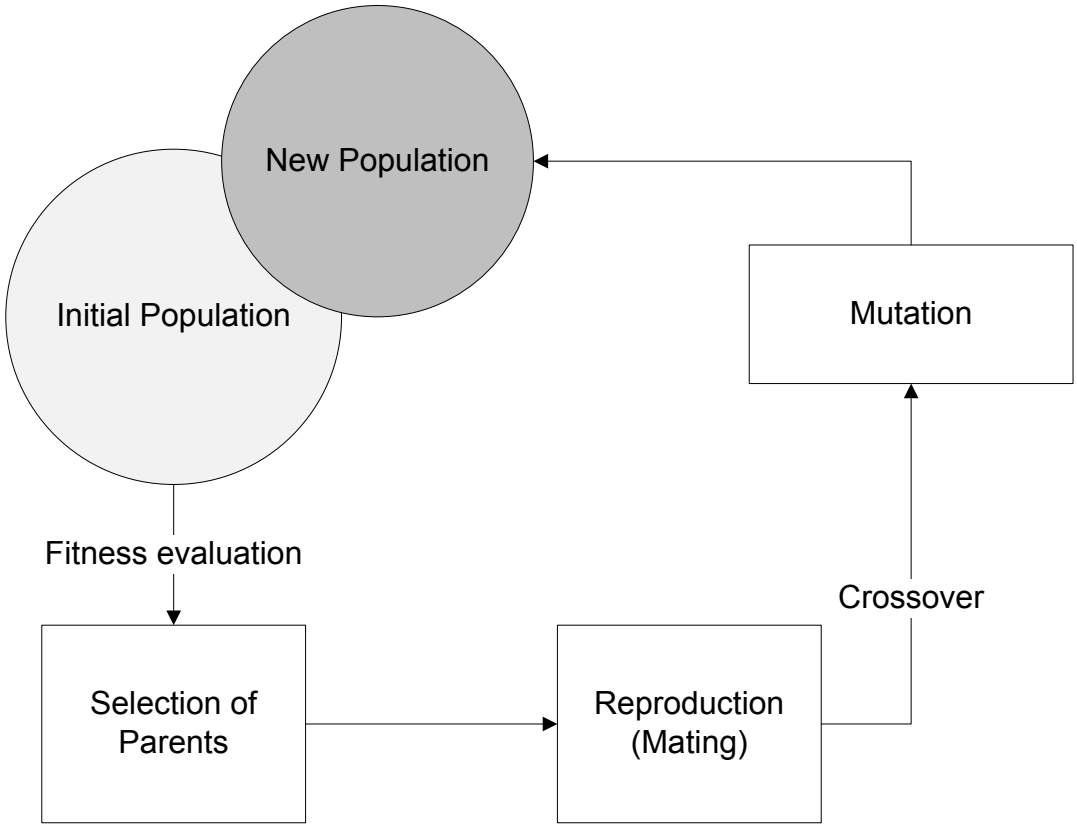

Figure 2.5: Operational diagram for genetic algorithm [Fulcher 2008, P.33]

The generic genetic algorithm involves representation mechanism, initialization of the population of individuals (candidate solutions), credit assignment/evaluation criteria, selection mechanism, reproduction mechanism, and replacement mechanism [Fulcher 2008]. Refer to Figure 2.5.

2.3.1 Representation Mechanism

In genetic algorithm, individuals, or chromosomes, are mostly represented as a string. The strings can be encoded in binary or other formats.

2.3.2 Reproduction Mechanism

This mechanism involves production additional individuals within the existing population of chromosomes. This may be simply by duplicating the current individuals within the population.

However, since the essential characteristic of GA is increasing the average fitness value of the population, the choice of population for reproduction will be based on the fitness value associated with the individual. Therefore, to help increase the average fitness value of the population over period of time, individuals with better fitness will have a better chance of being selected. In some problems, individual with the best fitness may not be the best individual to select.

2.3.3 Selection Operator

Selection operator selects individuals from the population of chromosomes based on their fitness value. Preference is given to the individuals with better fitness value. Individuals selected by the selection operator will go through process of genetic operations (crossover and mutation) to produce fitter offspring.

There are different ways in which individuals are selected in genetic algorithm. Some of the most common selection methods includes: elitist, fitness-proportionate, roulette selection and tournament selection. In elitist selection method, the fitter individuals in the population are guaranteed to be selected. In Fitness proportionate selection fitter individuals are more likely selected but not guaranteed. Fitness proportionate selection is also known as roulette-wheel selection. Tournament selection selects fitter individuals from subgroups where only one individual is selected from a single subgroup.

2.3.4 Genetic Operators

Genetic Algorithm involves two main genetic operators: Crossover and Mutation. The following section provides explanation in each of these operators.

Crossover Operator

In Crossover operation two individuals are first chosen by using the selection operator. Then, the crossover site is randomly chosen along the bit strings and the value of the chosen string is exchanged at the chosen crossover point. Finally, the two new offspring generated by this process are added to the population by replacing the unfit individuals.

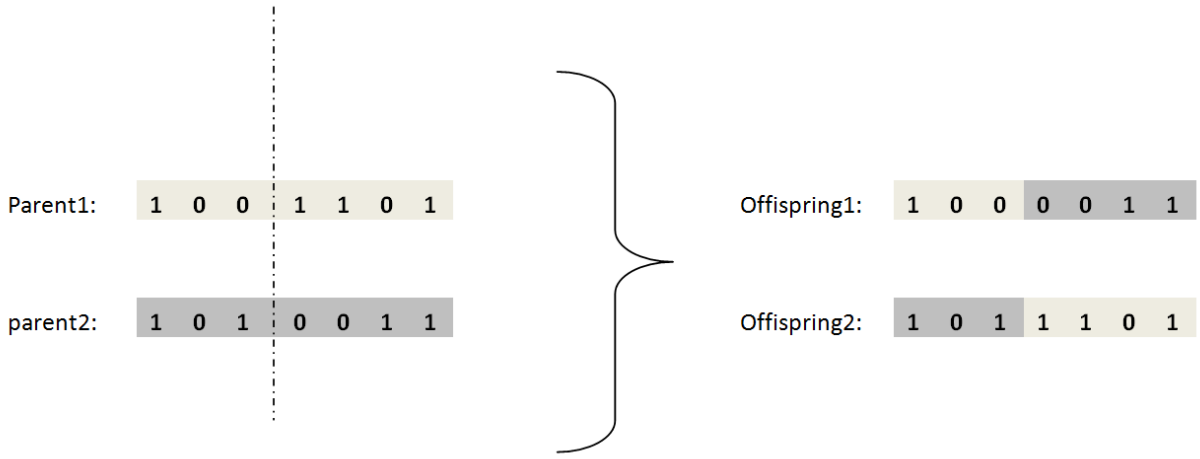

Figure 2.4: An example of a single point crossover

In a single point crossover, a line is drawn between two parents and the parent exchange genes at a crossover points (position three for both parent 1 and 2) as shown in Figure 2.4.

Mutation Operator

Mutation operation involves changing the value of the one or more bits within the bit string by a very small probability value. The main purpose of this operation is to maintain diversity within the population and inhibit premature convergence of the search algorithm. Mutation also allows the search algorithm to make a random walk among the population.

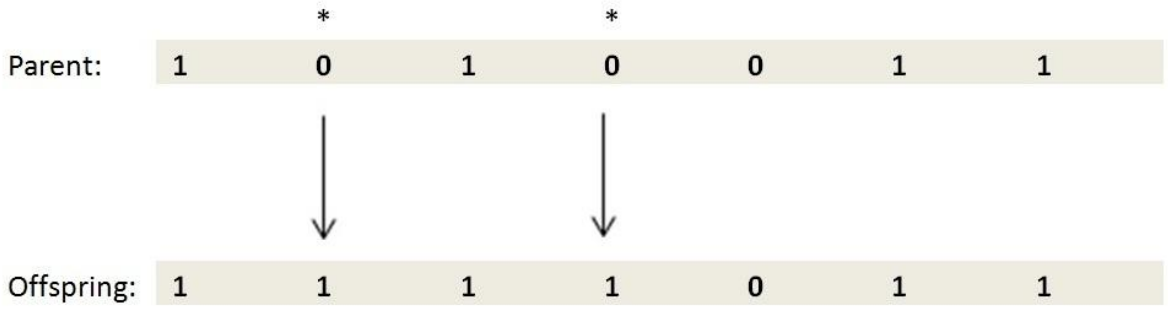

Figure 2.6: An example of mutation operation

Mutation of the type shown in Figure 2.6 where the bit values of the given chromosome are flipped is also known as bit flip mutations. It is the most common type of mutation, which is widely used in various problems.

2.3.4 Replacement

In this step the offspring replaces other individuals in the population depending up on the fitness criterion. In most cases worst individuals are replaced, however, there are also cases where the best, similar or immediate parents are replaced. Replacement may also be performed randomly in some problems. The following is the basic Genetic Algorithm procedure.

Algorithm 1: Genetic Algorithm

Step1: Generate initial population

Step2: **While** stopping criteria is not met **do**

(a) Select parents based up on the fitness rank

(b) Apply genetic operation on the Parents (crossover, mutation)

(c) Replacement: Update the current population by replacing the unfit individuals by the fitter new individuals

---

Chapter 3: RFID Security and Privacy Issues

Despite the enormous benefits it offers, RFID is also prone to security and privacy threats. The problem of security and privacy is worse in low-cost RFID devices because unlike other wireless communication systems, low-cost RFID tags have very low computational resources to execute standard cryptographic algorithm [Juels 2006]. Security and privacy models have been considered the limitations of the low cost RFID tags to have real world applicability. The following section discusses common type of attacks in RIFD systems as well as the challenges in preventing them.

3.1 Common Types of attacks in RIFD System

There are several identified attack types in RFID systems. Some of the attacks are well known attacks such as those found in other wireless systems. There are also other attacks which are prominent to RFID system due to the nature of the technology. The following subsections discuss some of the known attack s in RFID.

3.1.1 Eavesdropping

Eavesdropping involves unauthorized listening of communication between two or more parties without the knowledge of the participants. Eavesdropping can take place in two ways, active and passive. In active eavesdropping, the attacker probes the victim, which in most cases is the tag, and collects the response. In passive eavesdropping, the attacker does not do anything but listens and captures communication between the tag and reader. Passive eavesdroppers are difficult to detect as they are not involved in a communication, but are only listening to the channel.

Passive eavesdropping is prevented in traditional wireless systems by implementing strong encryption algorithms, which are computationally infeasible for the eavesdropper to be able to decrypt it.

3.1.2 Denial of Service

Denial of Service (DoS) attack occurs when the attacker uses a blocker tag or jams the normal communication between the legit reader and the tag. During blocking attack, the attacker uses fake tags and simulated the existence of several in the system there by making the reader to endlessly probe the several non existing tags, hence denying service to the legitimate tags.

In jumping attack, the attacker induces large amount of radio noises with the same frequency within the system causing the system to freeze hence successful launching of denial of service. However, denial of service such as blocking has been proposed in some literatures to be techniques to preserve the privacy of the tag bearer [Juels et al. 2003].

3.1.3 Unauthorized Reading

It has been shown that it is possible to maliciously extend the read range to much higher range than the standard read range given by the RFID application, such as those with near field communication [Kirschenbaum 2006]. This makes it possible for the attacker with the fake reader to be able to read information stored on the tag even if they are far at an invisible location.

3.1.4 Tag Data Modification

In an RFID system that has rewritable memory, attacker may be able to modify the content of the memory of the tag, such as tag id, and other valuable information to the other value or even delete it so that it will not be available to the legitimate users. Such attacks are usually be avoided by having strong pass-code or disable modification option on the tag.

3.1.5 Relay Attack

In relay attack the attacker makes connection with genuine parties (both the reader and the tag in our case) by using rob reader and tag. The rogue reader appears to the tag as the genuine reader and receives the message and sends it to the rogue tag which in turn passes the message on to the genuine reader appearing to be the genuine tag. The genuine reader then replies to the rob tag which again assumes the rob tag to be the genuine tag and passes the message on to the rob

reader and vice versa. The communication between the rob reader and the tag is normally so fast that they minimize the delay.

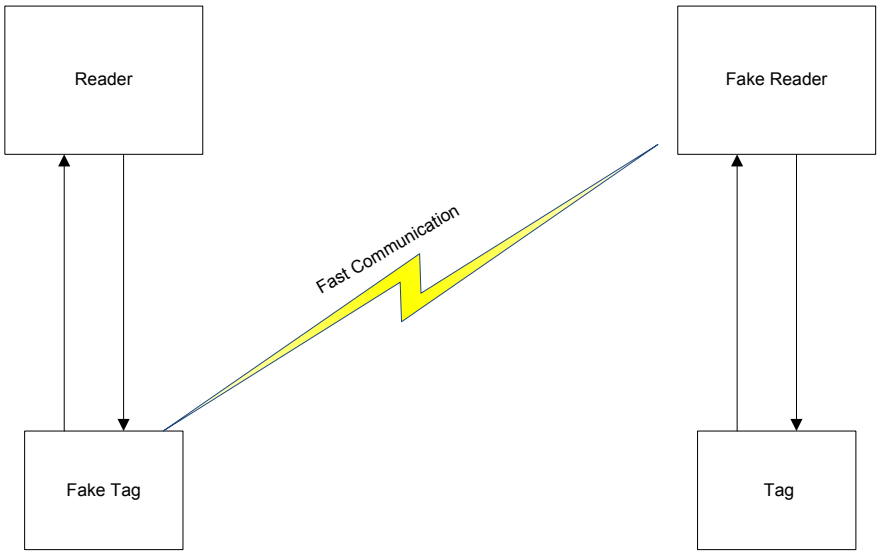

Figure 3.1: Relay Attack, modified from [Kirschenbaum 2006]

Figure 3.1 shows the practical scenario in how the relay attack can be launched on proximity based smart card [Juels et al. 2003; Hancke 2005]. In this way, the attacker can gain access to the system causing significant risks in various application areas, such as access control system.

3.1.6 Replay Attack

In Replay attack, the attacker first eavesdrop the communication between the genuine tag and reader and sends the same procedure, e.g. authentication sequence, to the genuine devices at latter times in an attempt to make system resources unavailable to the legit users (DoS attack). To detect and protect replay attack, different countermeasure procedures have been given and most of them are based adopting challenge response protocols. We will give more discussion on solutions to this attack by other literature in the later section.

3.1.7 Tag Cloning

Cloning attack is one of the major concerns in RFID system. Tag cloning occurs when the attacker makes a duplicate of the legitimate tag by being able to gain access to the ID and other data of the legitimate tags. The cloned tag will then be used by the attacker to gain an authorized access to a restricted area or launch any other malicious action by making they appear as the genuine tags. Table 3.1 summarizes the common types of RFID attacks and interfaces that can be exploited to launch them.

Table 3.1: RFID attack Types and Exploited Interfaces, Modified from [Rotter 2008]

Attack type	**Interface**
Eavesdropping	Communication Channel, Tag, Backend
Denial of Service	Communication Channel
Unauthorized Reading	Communication Channel, Tag
Tag Data Modification	Tag
Relay	Communication Channel
Replay	Communication Channel
Cloning	Tag
Malicious SQL injection	Middleware, Backend

3.2 RFID Privacy Concerns

Besides the security potential security threats discusses in section 3.1.1, RFID technology also introduces a number of privacy issues that have been one of the major factors slowing the wide spread use of the technology. For example, when a customer buys an items from a retail store, RFID tagged items may be used to locate the customer without their knowledge. This is because RFID tags, unlike other AUTO-ID technologies such as barcodes, can be read by a reader over a

longer distance without the knowledge of the users. In addition, RFID tags contain unique ID that helps to associate the tag carrier and the tag.

One of the technical solutions designed to combat RFID privacy issue is tag killing and sleeping approach [Juels 2006]. Tag killing is an approach which is used to permanently disable RFID tags. It is a command from the authorized user to the tag so that the tag in a way that makes the tag permanently inoperable.

Kill commands are usually protected and may prompt the reader for a pin or password for authentication. The command discards the tag so that no information can be inferred from it at later times. For example, whenever someone makes a purchase at retail store which uses an RFID to tag items, all the items the customer purchased will get their associated RFID tags disabled at the check-points.

RFID tag sleep command also works same way as RFID tag kill command; however, instead of permanently disabling the tag, it temporarily deactivates them. By temporarily deactivating the tags (tag sleeping), the customer privacy can be prevented in a way that the tag will not be read by any reader unless they are reactivated. Reactivating (waking up) a sleeping tag requires an authentication code just like the ones used to kill the tags. Other similar approach is to this is RFID tag renaming and relabeling [Juels 2006].

Another approach to combating RFID tag privacy issues is by using blocker tags [Juels 2006]. Blocker tags use an anti-collision algorithms protocols used by the RFID readers while communicating with a single tag in the presence of multiple tags within its read range. RFID reader recursively asks the tags in the range for their next bit and normal tags replies with the bit '0' or '1' but the blocker tags reply with both '0' and '1' which makes the readers think that all the possible tags are present [Juels et al. 2003]. The reader keeps scanning the blocker tags and eventually stalls as the number of possible tags which appears to be in the range is too large.

3. 3 Challenges in Securing RFID System

As discussed in the earlier section 3.2, the major issues that are holding back the wide spread adoption of RFID technology are security and privacy issues that RFID may introduce to the

adopting parties. This security and privacy concerns are posed by certain requirements while manufacturing RIFD tags.

Firstly, RFID tags are required to come at a low cost. This is because, for RFID tags to replace the existing labels such as barcodes, its cost has to be too lower. Reducing the cost however, is directly related to reducing the cost of hardware that makes up the tags. In other words, low cost passive UHF RFID is a desirable type in most applications. Low-cost passive UHF RFID tags have very limited memory size, no on-board power source and very limited processing capacity. Such tags are not capable to support computationally intensive algorithms.

Secondly, RFID tags are desired to be of small size in most applications. Having small size, however, comes with less computational capability because the lower the required size of the tag the less the memory size. In addition, it cannot support battery and other computational resources required for heavy weight cryptography. This forms another challenge in security low cost RFID tags (which is the key interest of this thesis).

Chapter 4: Literature Survey

In the previous chapter, we reviewed some of the common attacks on RFID systems and provided a table summarizing the types of attacks and the ways that they can be launched. We also discussed the major research challenges in securing RFID systems. In this chapter, we will discuss various solutions provided for some of the well known RFID attacks.

4.1 Overview

Security and privacy issues have been the bottlenecks for the fast adoption of RFID systems in various application domains. As discussed in the earlier section both manufacturers and consumers which intend to adopt the technology have raised serious security and privacy concerns. There have been several solutions given by researches in both industry and academia. The major challenge in securing an RFID system is attributed to the limitation on the tag in terms of its storage and computational capability to support computationally intensive algorithms. This section provides a review of various approaches taken by the researchers to overcome the challenges, specifically, in systems that adopt passive UHF low-cost RFID tags.

4.2 Approaches to Address RFID Security and Privacy Issues

There have been several literatures published over the past few years to address security and privacy threats in RFID systems. Most of the literatures are light weight cryptographic algorithms based on hash function [Dimitriou 2005; Avoine 2005; Juels 2007]. The light weight hash functions work well in a sense that the RFID tags can support the hash protocol which requires relatively low computational capabilities. However, standard hash functions such as MD5 and SHA-1 are still beyond the capability of the current Class-1 Gen-2 low-cost RFID tags [Dang et al. 2006].

4.2.1 Efficient Tag Hardware Design

One of the major challenges with regard to designing RFID tag hardware is to make it computationally powerful, small in size and low-cost. For a tag to be computationally powerful it

has to have sufficient processing and storage capacity. Having sufficient storage capacity means increasing the memory size of the tag which directly adds to the cost of the tag, thus, RFID tags with higher computational capacity comes with high cost. Therefore, research in manufacturing low-cost microchips is one of the major engineering challenges behind the success of the modern RFID system.

In addition to the cost, modern RFID systems also require the tag to be small in size. The size of a tag is mostly attributed to the antenna size which is required both to transmit/receive signals and in receiving power from the external sources by electromagnetic inductive coupling. For tags with their own on-board battery, the size of the battery is the main component contributing to the size of the tag.

As the scope of this thesis is limited to the algorithm perspectives of the tags than its engineering perspective, the reviews of such work are not presented here. In the next sections, reviews on some of the light weight cryptographic algorithms and intrusion detection approaches are presented.

4.2.2 Lightweight Cryptographic Algorithms

As implementing computationally intensive cryptographic algorithms is not feasible on low-cost RFID tags, the research in this category focuses on designing secure light-weight cryptographic algorithms that can be used on the tags with low computational capability. Most of the light weight cryptographic algorithms are based on the hash function; however, most of the well known hash functions are themselves computationally expensive. For example, cryptographic hash functions such as MD5 and SHA-1 cannot be implemented on the popular EPCGlobal Class-1 Gen-2 tags because of their low computational resources [EPC Global Inc. 2005].

Weis et al. [2003] proposed the hash-locking protocol. In this protocol, each reader knows the key ‘k’ for each tag in the system and each tag have an ID called MetaID which is the hash value of the key. When the tag is probed by the reader, it sends the MetaID instead of its actual ID and the reader computes the key of the tag. The reader then, sends the key back to the tag. The tag hashes the key received from the reader and compares it with the metaID it already has. This way, both the tag and the reader authenticates one another before the transmission of the actual

message. The main drawback of this protocol is that it has to keep the metaID fixed which can easily be exploited by the adversary.

As an extension to the hash-lock function Weis et al. [2004] introduced randomized hash-function. In randomized hash-function, the tag has pseudorandom number in addition to the hashed value. The tag computes the hash value based on the generated pseudorandom number. This technique, hence, keeps changing the metaID in the previous approach frequently making it difficult to be tracked by the adversaries. However, using a large pseudorandom number is not feasible on the low-cost RFID tag because of its limited resources.

Ohkubo et al. [2003] proposed a forward secure authentication scheme to solve the problem of eavesdropping. The paper introduces the technique called hash chaining. Hash chaining technique simultaneously uses two hash functions where the tag and the reader perform a series of transactions before they establish secure transaction. The problem with this scheme is that even if it protects the privacy it is not scalable.

Avoine et al. [2005] introduced a scheme that solves the scalability issues in Koutarou et al. [2003] discussed above. This scheme uses specific time-memory trade-off optimized to reduce complexity in the system. Being an extension of the earlier hash chaining mechanism, it is a forward secure technique in addition to being scalable without the need to implement asymmetric cryptography. This protocol, however, is still prone to attacks such as replay where the attacker queries the tag and uses the response from the tag for authentication.

Juels [2005] proposed a method for strengthening EPC tags against cloning based on challenge response authentication. This method leverages PIN-based access control system to achieve authentication through challenge response so that the PIN can be used not only for reader-to-tag authentication but also tag-to-reader authentication. This approach works well in preventing skimming attack; however, it does not protect the system from more sophisticated attacks.

Dimitriou [2005] proposed another authentication protocol based on hash function to prevent cloning. The scheme uses shared secret key between the tag and the backend database. The shared key is periodically changed to prevent tracking. However, this scheme is prone to attacks such as desynchronization.

Dang et al. [2006] proposed synchronization based approach for EPCGlobal Class-1 Gen-2 RFID tag with the capability to support cryptographic primitives like Pseudo-random Number Generator (PRNG) and Cyclic Redundancy Code (CRC). Every time the tag is queried by the reader, it generates different bit strings by using pseudorandom number generator to prevent tracking. The protocol also makes use of the CRC to ensure the integrity of the message at the backend system. The tag and the backend server are synchronized by sharing the same seed on PRNG in every session. Message XORing is used as an encryption/decryption mechanism.

4.2.3 Intrusion Detection Mechanisms

As discussed in earlier sections, tag resources are limited and implementing complex cryptographic algorithms on them is not feasible. For some applications implementing light-weight cryptographic protocols may be sufficient but in other applications which require a high level of security, their use may not be recommended as they do not provide a high level of security. Intrusion detection techniques are useful in such cases as they ignore the limitations on the tag and secure the system by detecting attackers even if they are able to obtain the secret keys. Such security protocols are implemented on readers or backend servers which have relatively higher computational capability. Components such as readers, middleware and backend enterprise servers have more computational resources than the tag.

In non-RFID systems, intrusion detection is an important mechanism that serves as an additional security layer to detect attacks in network systems where cryptographic algorithms may fail. In the same manner adoption of intrusion detection systems to RFID systems has also been shown to provide promising results [Mirowski and Hartnett 2007].

A method to detect transfer of RFID tag ownership, called Deckard Mirowski et al [2007], has been introduced to an RFID system. This method implements statistical intrusion detection technique proposed by Denning [1987]. In Deckard, the behavior of each user accessing RFID based proximity card system is modeled by using events generated by each user of the premises. The model characterizes the normal activity of the user and deviation from the known activity is reported as anomaly, triggering an alarm. Deckard achieves reasonably high detection rate, however, it was prone to high false positive rate (false alarms).

Lehtonen et al. [2009] introduced a method to detect cloning of a tag as soon as it is scanned by the system. The system makes use of the rewritable memory of a tag to replace the current secret number on a tag by a new randomly generated number every time the tag is scanned. These secret numbers are synchronized on the back-end system and the backend system keeps track of these numbers as well as check for synchronization errors. This way, the tags, which appear in the system with outdated synchronized secret, will trigger an alarm. This solution is less prone to the false positive rate than Mirowski et al. [2007] since each alarm is associated with irregularities in the system. This solution is limited by the strength of encryption algorithm used to encrypt the synchronized secret. In addition, the solution takes into account only the tags with rewritable memory and hence doesn't scale to passive UHF tags having read only memory.

4.2 Summary

In this chapter we started by giving an overview of security and privacy challenges presented in RFID systems and then we presented the directions being taken in research to secure low-cost tags. In general, securing low-cost RFID tags has been a remarkable challenge among the RFID research community mainly due to insufficient computation resources that the low-cost tags have in order to execute more secure but computationally intensive cryptographic algorithms.

The key research directions in securing low-cost RFID tags take three roots. The first deals with the issues related to the manufacturing small and efficient tag hardware. The second deals with methods of designing efficient security algorithms that requires less resource. In other words, they involve research that deals with the design of light weight cryptographic algorithms. There are a good number of researches published in this category. The third research root deals with the design of the protocol that ignores the limitations on tag resources, and implements the security protocols on the other system components such as reader, middleware or the backend system.

In the next chapter, we will provide the detailed description of the approach considered in this thesis.

Chapter 5: Solution Approach

In this chapter, we will discuss Hybrid Fuzzy GBML algorithm, the algorithm used to design our RFID anomaly detection system. We begin by giving an overview of the algorithm. We will discuss steps involved in implementing the algorithm, which includes the techniques used to generate rules from the raw data (RFID system event logs in the case of this thesis). We first discuss how antecedent fuzzy sets are generated and proceed to explain the processes involved in determining the most accurate consequent class for antecedents of the given rules and how a grade of certainty for the given rule is computed.

5.1 Overview

A Hybrid fuzzy genetics-based machine learning algorithm combines two genetics-based machine learning algorithms, known as Pittsburgh and Michigan approaches, for designing classification system that makes use of fuzzy rulebase [Ishibuchi et al. 1999]. It makes use of the benefits of both its constituent algorithms to come up with the classification system with better performance.

Pittsburgh approach considers a set of fuzzy rules as an individual and each fuzzy rule set in a population is evaluated by its overall classification performance over the training samples. The fitness of individual rule set is its classification performance (accuracy) over the training samples. During genetic operation, crossover takes place between two fittest individuals (rule sets) while mutation is applied on an individual rule set by modifying the rule set's elements (rules). Each fuzzy rule sets in Pittsburgh approach is coded as a binary string where each bit indicates the presence or absence of constituent fuzzy rule within the population.

Michigan approach considers each rule within the population as an individual and each fuzzy rule are coded by series of integer/character. The fitness of individual fuzzy rule is calculated by its classification performance over all the training samples. Genetic operations are performed on fuzzy rules in the population where crossover takes place between two best individual fuzzy rules and mutation is performed by modifying constituent antecedent fuzzy sets of the fuzzy rules depending on the given probability.

The advantage of Michigan approach over the Pittsburgh approach is that it has global search ability for a best fuzzy rule over the training samples. Its weakness is that it has low optimization ability of the rulebase used for the system. On the other hand, Pittsburgh approach has ability to optimize rule-base used for classification system. However, Pittsburgh approach has lower global search ability of fuzzy rules over the training samples.

The hybrid approach makes use of the advantage of both approaches. Genetic operators for generating new fuzzy rules are performed by using Michigan approach to heuristically search for the best fuzzy rules while the entire algorithm is based on Pittsburgh approach. Hybrid approach has been shown to have better performance ability than its constituents, Pittsburgh and Michigan approaches [Ishibuchi et al. 1999].

5.2 Initial Fuzzy Rule-Base Generation

Fuzzy rule bases are normally constructed by using knowledge from human experts who know a domain for which the expert system is being designed very well. Obtaining knowledge from experts, however, is tedious and expensive or can even be erroneous when designing complex system where the amount of information required is too large. To overcome this challenge, various methods have been proposed to automatically learn fuzzy rules from data [Abe and Ming-Shong 1995; Nauck and Kruse 1999; Nozaki et al. 1996]. For generating fuzzy rules from numerical data in this thesis, we utilize the method proposed by [Ishibuchi et al. 1992] for classification problems.

5.2.1 Antecedent Fuzzy set Design

For the classification problem, Nozaki et al. [1996] divides the process of constructing fuzzy rules from training data into two phases. In the first phase a pattern space is partitioned into fuzzy subspaces, as illustrated in Figure 5.1. Partitioning of a pattern space into fuzzy spaces is done in so that it does not result in too few or too many fuzzy partitions. Too fine and too coarse partitioning of a pattern spaces are both undesirable, as they would lead to the poor performance of the model. If the partition is too fine, the model would not generate rules for some of the partitions due to the lack of training data within them. If the partition is too coarse, then the model will generate few rules which are too general leading the model to poor performance.

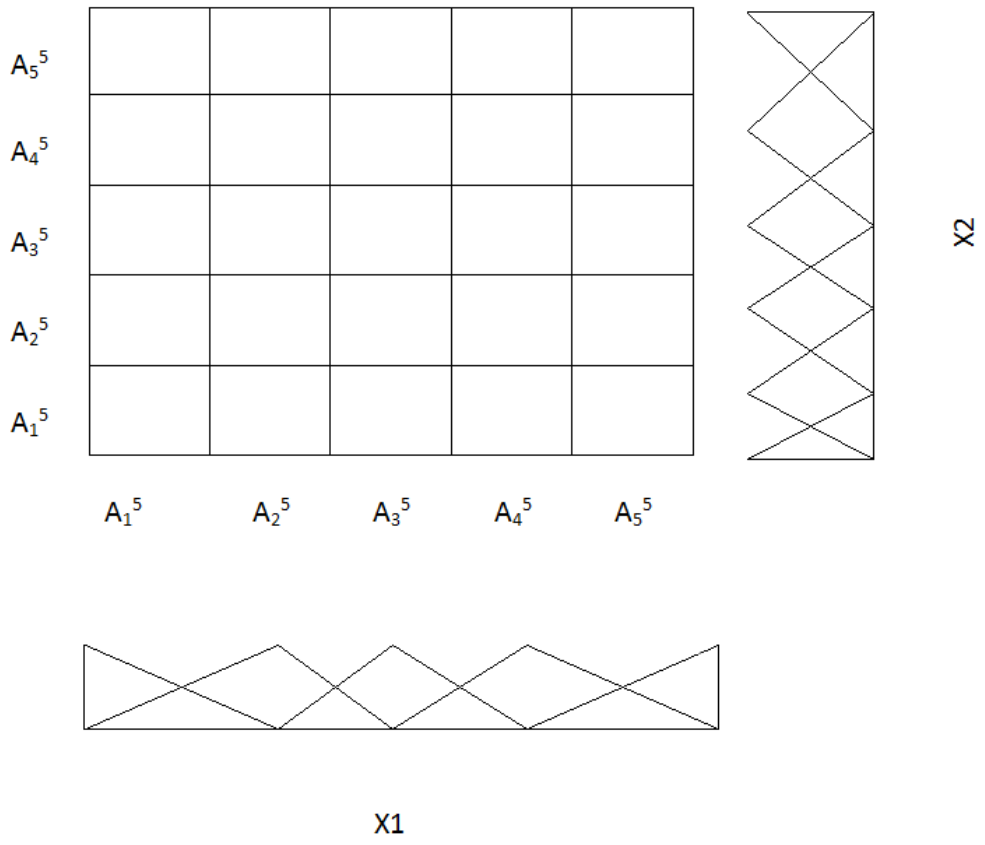

Figure 5.1 Fuzzy partitioning of the two dimensional pattern space [Nozaki, et al. 1996]

In the Figure 5.1, we have fuzzy sets 5×5 which equals 25 fuzzy partitions on two dimensional spaces. The superscript K, in an antecedent fuzzy set A_i^K, represents the maximum number of subspace in each axis and i represents the current partition on the given axis.

Nozaki et al. [1992] proposes two possible approaches to avoid the problem of having too few or too many fuzzy partitions. One of the methods considers applying heuristics technique based on the density of the dataset. The other method makes use of multiple fuzzy partitions simultaneously. By simultaneously utilizing multiple fuzzy partitions, the model considers all possibilities of partitioning the pattern space. This is illustrated in Figure 5.2. The latter approach has been used by Nozaki et al. [1996] for pattern classification problem. In this thesis, we make use of the latter approach where multiple fuzzy partitions are simultaneously utilized.

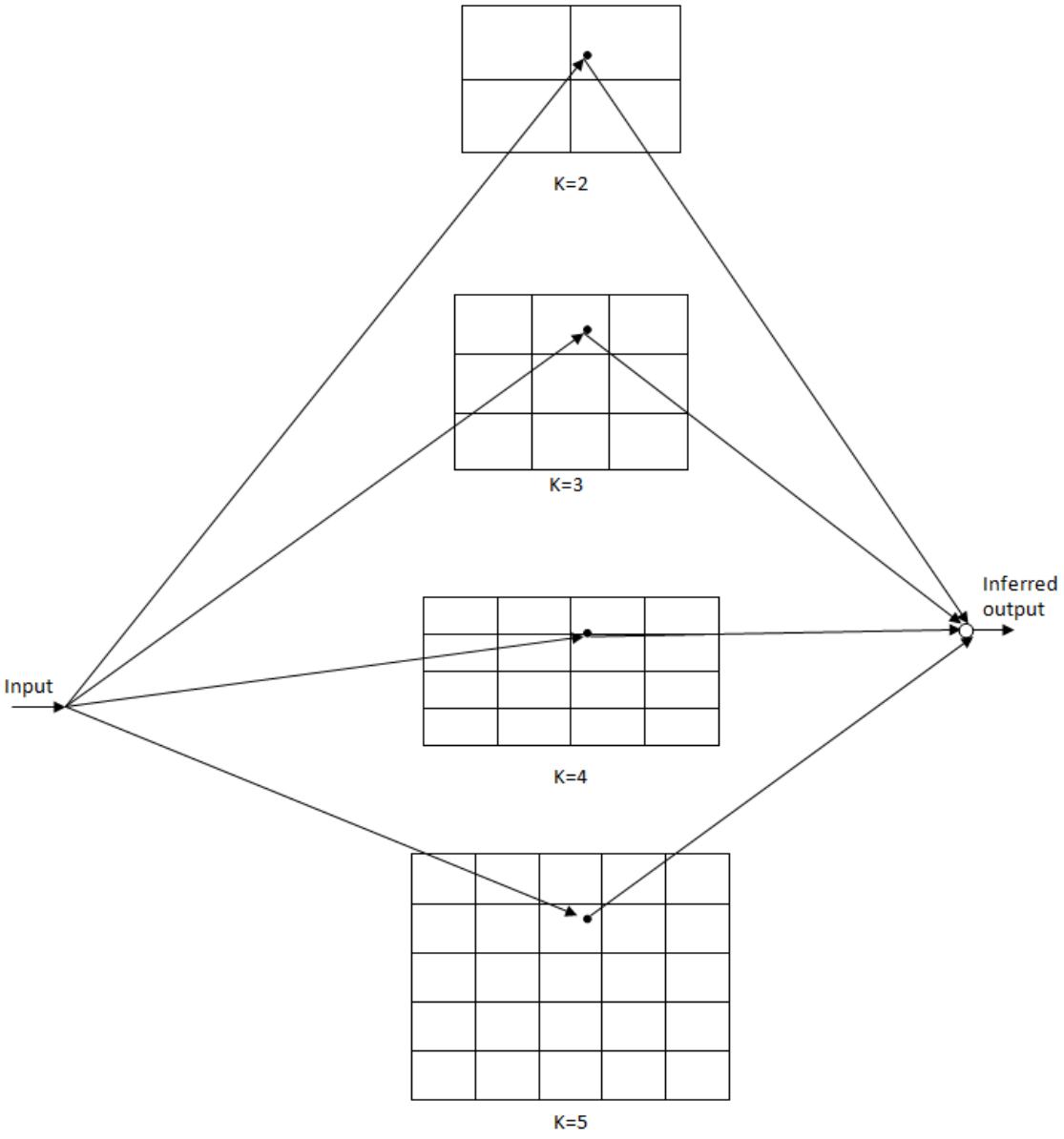

Figure 5.2: Simultaneous use of multiple pattern spaces for two dimensional patterns [Nozaki et al. 1996]

The approach that simultaneously utilizes multiple fuzzy partitions also has some drawbacks when working with high dimensional pattern classification problems. This is because the number of the generated fuzzy rules could be quite large leading to more complex model.

Once the pattern space is partitioned into subspaces and each fuzzy subspace is defined, the next step is to use the combination of each attributes as an antecedent part of the fuzzy rule. For

example, in Figure 5.2, we have $(14+1)^2$ combinations of the fuzzy sets as an antecedent part of the fuzzy rule. The total number of partitions for each attributes is 14 and 1 is for the "don't care" character. The superscript 2 is the number of attribute in a pattern (the dimension of the pattern classification space). In the partition above (see Figure 5.2), if there are n-dimensions in the problem, the total number of the generated fuzzy IF-THEN rules are 15^n.

For a low dimensional problem with a few attributes, the number of generated rules is reasonably small and they can all be used as candidate rules. On the other hand, for high dimensional pattern classification problem (large value of n), it is infeasible to examine all the generated rules [Ishibuchi 2007]. To select reasonable number of rules for the pattern classification problem with high dimension, in some approaches rules are randomly selected [Abadeh et al. 2007] while some other approaches use heuristic rule evaluation criteria to choose a smaller number of rules as candidate fuzzy rules [Ishibuchi 2007]. For the experiment presented in this thesis, there are only three attributes and it is reasonable to consider all the generated fuzzy rules as the candidate rules.

The consequent class for each fuzzy IF-THEN rule (as discussed in next sub-section) is determined by taking the majority class in the corresponding fuzzy subspace. The method is proposed in [Ishibuchi et al. 1992]. The same approach in used for the experiments in this thesis. The given axis the pattern space is divided into K fuzzy sets $\{\alpha_1^k, \alpha_2^k, \ldots, \alpha_k^k\}$. The membership function selected is triangular.

$$f_i^k(x) = Max\{1 - \frac{|x - a_i^k|}{b^k}, 0\}, \qquad i=1, 2, \ldots, k \qquad (1)$$

a_i^k of the symmetric triangular membership function in the equation(1) is computed as:

$a_i^k=(i-1)/(k-1)$, i=1,2,…,k. (2)

b^k in equation(1), representing the spread of the membership function is computed as:

$b^k = 1/(1-k)$ (3)

5.2.2 Determining the Consequent Classes and the Grade of Certainty

The consequent class of the given rule (antecedent) is determined by first calculating the total compatibility grade of the fuzzy rule with each training pattern and assigning the class with the maximum compatibility grade as the consequent class. The following procedure explains how to determine the compatibility grade fuzzy rule with each training pattern. Then the method for computing grade of certainty is given.

The Total Compatibility Grade

There are two operators used to calculate the compatibility grade of fuzzy rules with each training patterns, product and minimum operators. The product operator is more popular while the minimum operator is not as famous [Fulcher 2008]. In this thesis, we use the product operator to calculate the compatibility grade of the training pattern with antecedent fuzzy sets.

$$\alpha_q(X_p) = \alpha_{q1}(x_{p1}).\alpha_{q2}(x_{p2})\dots\alpha_{qn}(x_{pn}) \quad (4)$$

Since our patterns have only three attributes, the compatibility grade will be

$$\alpha_q(X_p) = \alpha_{q1}(x_{p1}).\alpha_{q2}(x_{p2}).\alpha_{q3}(x_{p3})$$

$\alpha_q(X_p)$ is the compatibility of Xp with the antecedent part $\alpha_q=(\alpha_{q1},\alpha_{q2},\ldots,\alpha_{qi})$ and $\alpha_{q1}(.)$ is the membership function of the antecedent fuzzy part.

The total compatibility grade of each class with antecedent vector is calculated as follows:

$$\beta_{C1} = \sum_{x_p \in C1} \alpha_q(X_p) \qquad (5)$$

And

$$\beta_{C2} = \sum_{x_p \in C2} \alpha_q(X_p) \qquad (6)$$

If $\beta_{C1} = \beta_{C2}$, then the fuzzy sets corresponding to the fuzzy subspace (α_i^k, α_j^k) is not generated. For $\beta_{C1} \neq \beta_{C2}$, the class of the partition space will be the majority class in that subspace.

$$\beta_{C1} > \beta_{C2} \quad \text{then } C_{ij}^k = C1 \qquad (7)$$

And

$$\beta_{C1} < \beta_{C2} \quad \text{then } C_{ij}^k = C2 \qquad (8)$$

The the grade of certainty for each rule is calculated as follows:

$$CFijk=|\beta_{C1} - \beta_{C2}|/(\beta_{C1} + \beta_{C2}) \quad (9)$$

From the computations in equations 5 and 6, the class with the larger total compatibility grade to the premises of the fuzzy rule set is considered to be the consequent class. The value of the certainty grade is between the interval [0, 1].

From the above procedure, large number of fuzzy rules can be extracted by assigning the consequent class and the rule weight for every possible combination of the antecedent fuzzy sets. However, all these large number of fuzzy rules are intractable for human users. In addition, long fuzzy rules with many antecedent fuzzy rule conditions are difficult for humans to understand. In other words, the generated rule by using only the above procedure suffers from low interpretability. To overcome this, only short fuzzy rules with limited length of antecedent part are normally used. In our experiments, since we only have three attribute, we use them as they are.

Chapter 6: Experimental Setup

In this section, we provide the description of the dataset used for evaluation of the proposed model and preprocessing performed on the dataset before the model is applied to it. The results of the experiments are presented in chapter 7.

6.1 Description of the dataset

The dataset used for the evaluation of the proposed approach is provided by the University of Tasmania's School of Computing and Information Systems. This dataset has been used by Mirowski et al. [2007] and Lehtonen et al. [2009] to evaluate cloning detection approaches in RFID systems.

The dataset consists of RFID read events, collected from proximity card system, over the period of three years. The events are generated by the system when the cardholder attempts to unlock the doorways of the building.

Since the dataset contains, information such as: "which user", "which doorways of the building", "at what time of the day", "how often", it can be used to characterize the normal behavior of each user accessing restricted areas of the building. These in turn would help to characterize the abnormal cases where the observation varies from the expected normal.

The proximity card system consists of: the proximity card (RFID tag), the proximity card reader (RFID reader), and the database system at the backend. There are originally six attributes associated to each read event in the database at the backend. Table 6.1 shows the attributes in the original dataset and their possible values.

Table 6.1: The Original dataset Attributes and their possible values [Mirowski et al. 2008]

Attributes	**Description**
Date	The date of a year, dd/mm/yyyy
Time	The time 12 hour format, hh:mm:ss
Time Period	AM, PM
Access/Decision	Granted
	Denied-noPermission
	Denied-foreignCard
	Denied-TimeZone
Reader Number	Readers numbered 25 to 28
Card Number	Tag ID number converted to integer values (100 – 413)

6.2 Data Preprocessing

For our experiment, we used events generated by randomly selected users of the proximity card system. The users are selected such that the number of events generated by them is beyond the certain minimum (120 in our case), as users with insufficient records cannot be properly characterized. The choice of 120 as the bottom-line is done by intuition; however, selecting users with larger number of records would provide better result.

The location frequency profile (LFP) and time frequency profile (TFP) was used to characterize every user's access behavior to the premises. All events with missing data were removed, since they do not have any impact on the learning process. In addition, events with access type *denied-timeZone, denied-noPermission* and *denied-foreignCard* have been removed since they can be automatically detected by the system and are too rare to characterize the normal behavior.

The attributes Time and Time Period of the original dataset were merged and converted into a 24 hour time format. A new attribute called day was computed and added to represent the day of the week (e.g. Monday, Tuesday).

We also added frequencies of identical events observed in the system within a day, as a new attribute, by computing it from the original dataset and the day attribute. A day (24 hours) and the user's access frequency attributes are normalized to the range [0, 1] by using the MIN-MAX

normalization method. Each day of the week (Sunday to Saturday) was assigned a numeric value 1 to 7 respectively.

Every user of the premises is characterized by the frequency of their normal access within a given range of time of a day. Whenever there are events generated by an unauthorized user(s) by obtaining the clone of one or more proximity cards, the total number of events generated by the same ID will raise above normal (the aggregate of the normal user and the non-legitimate user), signaling the presence of anomaly within the system.

After analyzing the normal access behavior for each user, we randomly added events to the dataset that will lead the system learn to recognize the occurrence of an anomaly. Figure 6.1 shows events generated during the normal activity and the raise in frequency (number of identical events) due to the presence of one or more unauthorized access/accesses to the premises.

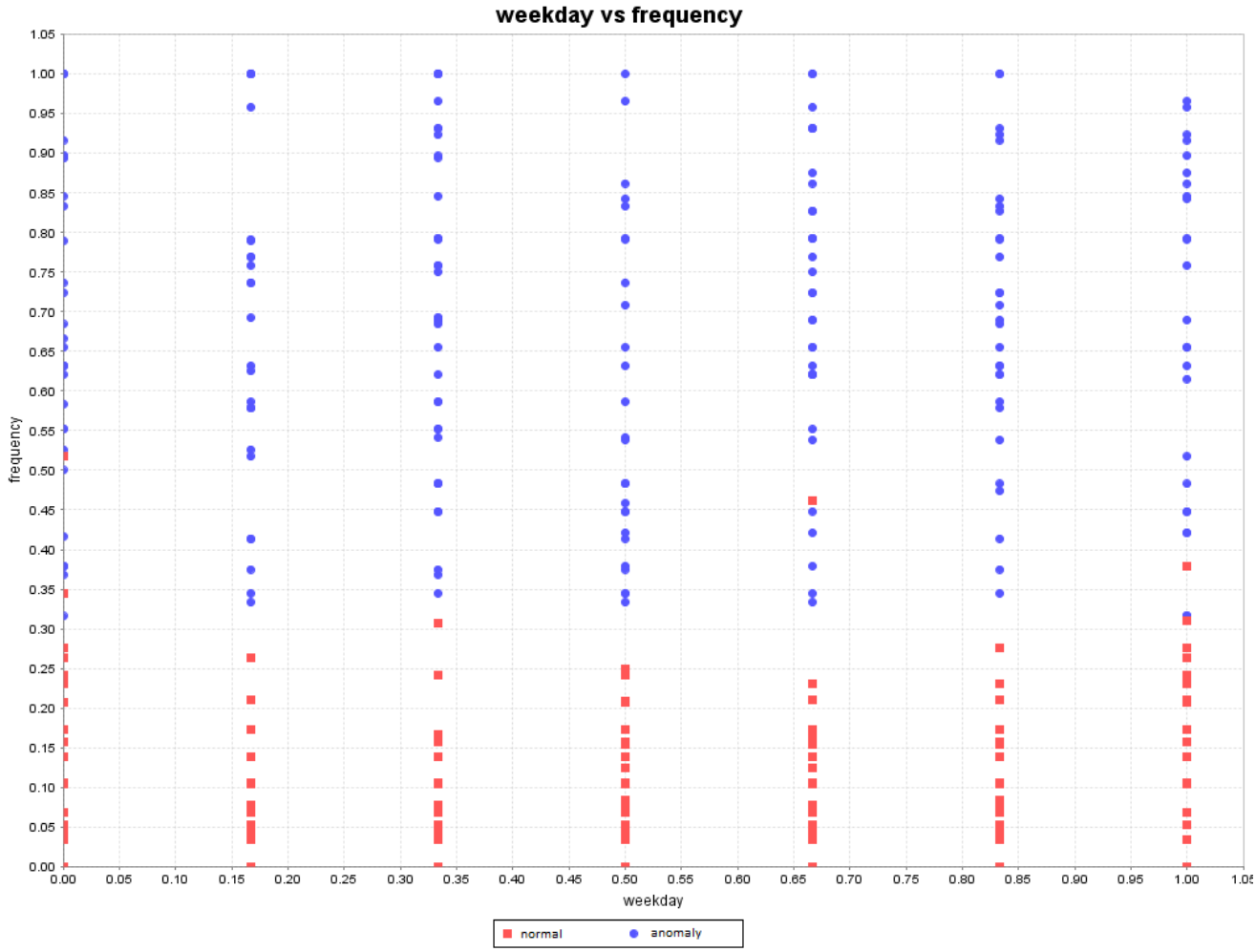

Figure 6.1: Events generated by the normal users and events generated in the presence an Intruder User(s)

Anomalous data of a given day for a given user (See Figure 6.2) was generated by using the following formula:

$$T_{(u,d)} = f_{max}(u, d) + randb(1, 3 * f_{max}(u,d))$$

Where

$T_{(u,d)}$: is anomalous data of the give user at a given day

$f_{max}(u, d)$: is the maximum use frequency of the given user at the given

$randb(1, 3 * f_{max}(u,d))$: is the random number between 1 and thrice the maximum frequency at a given day.

As discussed in the section above, the major limitation of the real world dataset is the fact that it is assumed to be attack-free. In other words, all the obtained events are taken to be generated by normal tags. Thus, synthesizing events to characterize attack behavior was the necessary step in our experiment. After adding simulated attacks to the original dataset, we ran the perceptron algorithm [Freund and Schapire 1999] to check whether the dataset was linearly separable or not.

Although the perceptron algorithm converges in a finite number of iterations, it has an error rate of 12.03% over all the training and testing sets. The main advantage of the approach taken in this thesis over the linear classifiers is its ability to provide linguistically interpretable rules that can be used to support decision making. The results obtained from our approach are presented in Chapter 7.

6.3 Fuzzy Partitioning of the Pattern Space

To avoid the problem of too coarse and too fine fuzzy partitioning during our experiment, we employed the technique given by [Ishibuchi et al. 1999]. Four pattern spaces with different granularity and fuzzy sets with triangular membership function were used. The total number of used fuzzy sets including the "don't care" are 15, see the following paragraph for detail. All attributes used in our experiment are normalized to the range [0, 1] by using MIN-MAX normalization technique. Figure 6.3 shows the fuzzy partitions used during our experiment.

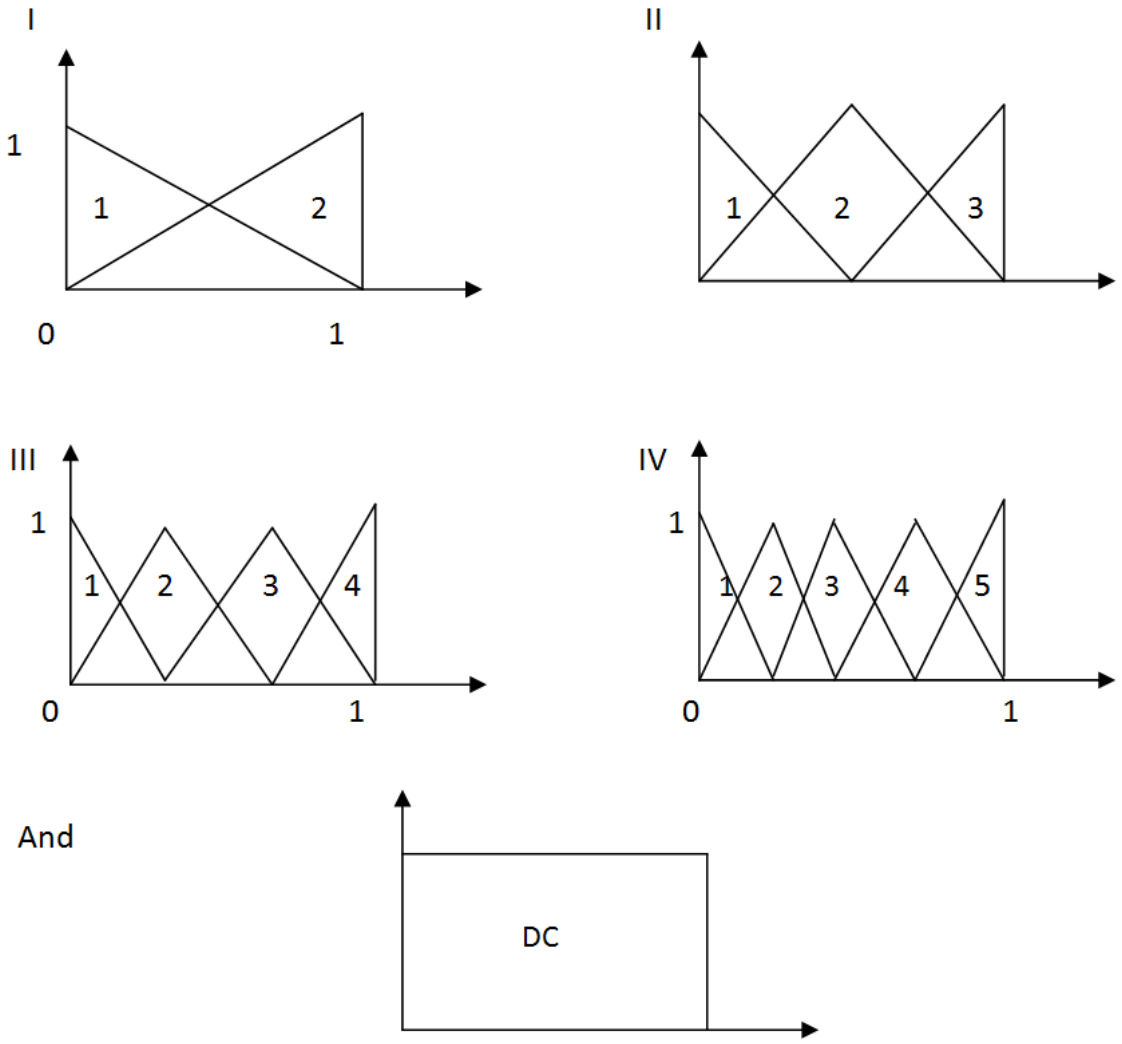

Figure 6.2: The Fuzzy Partitions used [Ishibuchi et al. 1999]

The linguistic variables used are small, medium and large. There is the total of 15 antecedent fuzzy sets used. Each fuzzy partition corresponds to their respective linguistic value where 1 (I) is small, 2 (I) is large, 1 (II) is small, 2 (II) is medium, 3 (II) is large, 1 (III) is small, 2 (III) is medium small, 3 (III) is medium large, 4 (III) is large, 1 (IV) is small, 2 (IV) is medium small, 3 (IV) is medium, 4 (IV) is medium large, 5 (IV) is large and DC is don't care.

6.4 Hybrid Fuzzy GBML Parameter Settings

Hybrid Fuzzy Genetics-based Algorithm was applied to the candidate rule set population containing N_{pop}=200 fuzzy rule sets where each fuzzy rule set consists of 10 rules. The probability for the Michigan iteration is set to 0.5 while crossover probability for both Pittsburgh and Michigan operation is set to 0.9. The probability of the "do not care" being 0.5. The total number of iteration in a single run is set to 1000. Table 5.4 summarizes the parameters setting used during the first run of our computational experiments.

Table 5.4: The Hybrid Fuzzy-GBML parameters used

Parameter	Value
Number of Fuzzy Rules	20
Number of Rule Sets	200
Crossover probability	0.9
Number of Generations	1000
Do not Care Probability	0.5
Probability for a Michigan Iteration	0.5

Chapter 7: Experimental Results

This chapter presents the results of the experiments performed in this thesis according to the procedures presented in chapter 5. We begin by presenting the rules generated at the end of the training of the model by using 10-fold cross validation. Then, we present the performance of the model on the training and unseen datasets.

The following are the first 10 rules generated after the training of the model according to the setup provided in chapter 6:

1: weekdays IS D.C. AND frequency IS L_0(4) AND Id IS D.C.: granted with Rule Weight: 1.00

2: weekdays IS L_1(4) AND frequency IS L_0(5) AND Id IS D.C.: granted with Rule Weight: 1.00

3: weekdays IS D.C. AND frequency IS L_2(5) AND Id IS L_1(3): denied with Rule Weight: 0.98

4: weekdays IS L_0(3) AND frequency IS L_3(4) AND Id IS D.C.: denied with Rule Weight: 1.00

5: weekdays IS L_0(5) AND frequency IS L_2(5) AND Id IS L_4(5): denied with Rule Weight: 0.97

6: weekdays IS L_3(4) AND frequency IS L_1(4) AND Id IS L_2(5): granted with Rule Weight: 0.51

7: weekdays IS L_1(3) AND frequency IS L_0(5) AND Id IS L_0(2): granted with Rule Weight: 1.00

8: weekdays IS L_1(5) AND frequency IS L_0(3) AND Id IS L_3(4): granted with Rule Weight: 0.99

9: weekdays IS D.C. AND frequency IS L_3(5) AND Id IS L_3(5): denied with Rule Weight: 1.00

10: weekdays IS L_2(5) AND frequency IS L_0(5) AND Id IS L_0(4): granted with Rule Weight: 1.00

7.1 Post Training Validation

In this validation phase, our training dataset is divided into 10 equal partitions where one of them (10% of the training dataset) is used as a test set and the remaining 9 (90% of the training dataset) are used as the training set. On the next round, one of the nine training partitions is picked as the test set while the test partition in the previous step is merged to the remaining 8 partitions to form a newer training set. This process is repeated 10 times until each of the partitions has been used as the test set. Finally, the results obtained in every process are averaged to obtain the global classification performance of the model over both the training and test sets. This method is commonly known as 10-fold cross validation. By using the 10-fold cross validation, the classification performance of the model was computed by a single run of a hybridized fuzzy genetics-based machine learning algorithm. A total of 1077 records were used and the model was able to achieve classification accuracy of 99.50% and 99.16% on the training and test partitions respectively.

7.2 Evaluation on the Unseen Dataset (Test set)

After performing the post-training validation discussed in section 7.1, we evaluated the model's ability to predict over the unseen data. A total of 732 records were picked from the samples to make up the dataset used at this phase. This dataset was not used during the training of the model and it is therefore an unseen, although it is from the same domain as from the training set. The classification accuracy of the model on the unseen dataset was 88.45%.

The model's Detection Rate (DR) and False Positive Rate (FPR) were computed as follows:

$$DR = \frac{TP}{TP+FN} \quad (10)$$

$$= 99.5\%$$

where TP is the number of true positive events, actual attacker events correctly detected as an attack by the model and FN is the number of false negative events or the number of normal events detected as an attack by the model.

$$FPR = \frac{FP}{FP+TN} \quad (11)$$

$$= 4.67\%$$

FP is the number of false positive or the number of normal events predicted by the model as an attack and TN is the number of true negatives, which means normal events predicted as a normal by the model.

The model has achieved the detection rate of 99.5 and the false positive rate of 4.67%. The false positive rate of 4.67% could still be high for some applications; however, compared to linear approaches, the main advantage of this approach is that it provides highly interpretable rules which could be used to during decision making.

In order to understand the effect of changing the number of fuzzy rules over the detection rate of model, we run a series of experiments by varying the number of rules from 5 to 50, see Figure 7.1. The model was found to achieve the highest classification performance both on the training and test partitions when the number of rules is 20. Another observation was that as the number of rules get larger, the performance of the model increases on the training set but decreases on the test set. This is because as the number of rules gets large, the model gets complex leading to problem of overfitting.

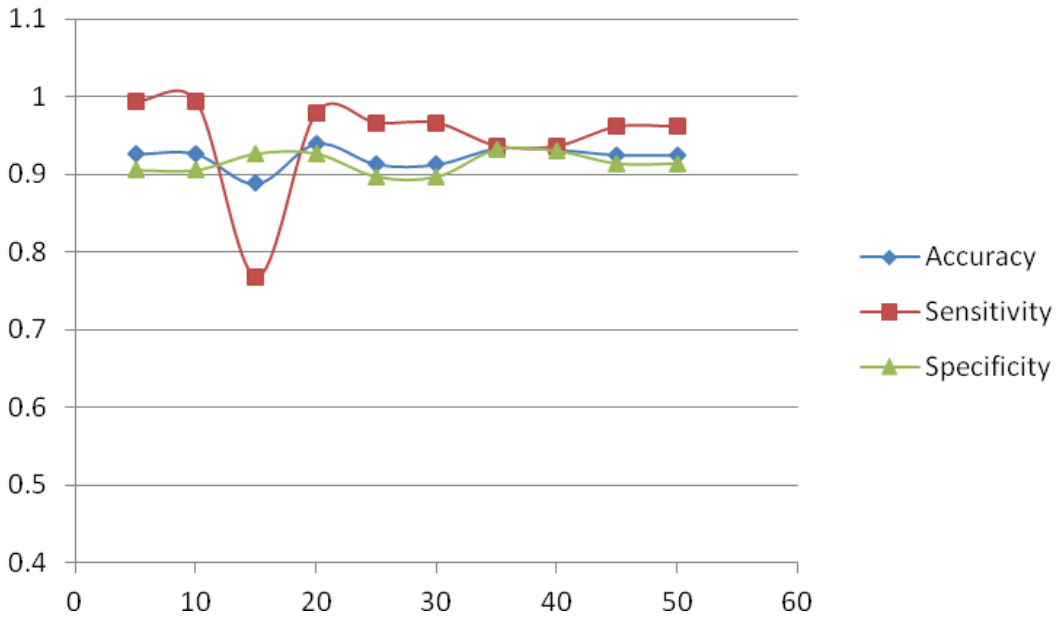

Figure 7.1: The model's performance for various number of rules

One of the limitations of the model is that although it effectively reports the presence or absence of an anomaly in the system, it does not automatically tell between who is carrying cloned tags. However, for the specific problem addressed by the computational experiment in this thesis, the system can be extended to take further action by automatically changing the card number. This can be achieved by updating the legitimate users with the new PIN number (e.g. once only use PIN) via e-mail or text message to their mobile phones while putting access restriction on the old card or trigger an alarm as soon as the old card is scanned in the system.

The other limitation of this approach is that all the data collected from the system is considered to be anomaly free. This can affect the design of the detection system in that the normal users cannot be correctly characterized allowing the undetected intruder users to continue accessing the system. We suggest that this can be addressed by automatically updating all normal user's ID by sending out PIN numbers (e.g. once only use PIN) every few weeks to every user and continuing to train the model until the behavior of every user is sufficiently characterized.

Chapter 8: Conclusion

8.1 Conclusion

Security and privacy issues in RFID systems have been widely studied by research communities in industry and academia. Low-cost RFID tags, EPCGlobal Class-1 Gen-2, are the widely studied type of tags. This is because they are the highly desirable tag types in many applications including supply chain management, manufacturing industries and retail stores.

Intrusion detection approaches are a less studied approach in securing RFID systems. In this thesis, we made use of the hybrid fuzzy genetics-based machine learning algorithm to design an anomaly intrusion detection system for RFID systems by making use of RFID generated events in the past. The results from our experiment indicate that a high detection and low false positive rate can be achieved by the proposed model. The model also has an advantage in that it yields highly interpretable fuzzy rules that enhance human decision making when there is no clear distinction between the normal and anomalous situations.

The result from our experiments show that the hybrid fuzzy genetics-based machine learning algorithm can achieve high performance in detecting anomaly intrusions in RFID systems, specifically cloning intrusion caused by one or more cloned RFID tags in the system. Over all the detection rate of the model was 99.5% and the false positive rate was 4.67% over the training and test dataset used in the computational experiments of this thesis. The results indicate that such models can successfully be adapted to other applications such as supply chain management where location and time attributes could be used to model the normal behavior of an item in the system [Lehtonen et al. 2007].

Fuzzy systems such as the one discussed in this thesis are quite useful in that they have the ability to precisely indicate the situations where there is no clear distinction between the normal and anomalous cases. They achieve this by attaching a level of certainty to every rule used during decision making.

8.2 Future Work

One of the future extensions of this thesis could be applying the method to other RFID application areas, such as supply chain management, by assessing the most relevant features to characterize the system. In this thesis, we were able to detect the presence or absence of an anomaly in the system along with a given degree of confidence; however, we did not differentiate between the normal and intrusive events in the system. Extending the method so that it would differentiate between the normal and intrusive events would be another useful task to consider. In addition, assessing the applicability of this approach to identify anomalies caused by other RFID attacks (e.g. denial of services) would be a useful endeavor for future research.

As the number of users in the system increase, the hybrid fuzzy GBML algorithm requires a larger number of fuzzy rules so as to characterize the behavior of objects in the system. This would add complexity to the model and would result in decreased interpretability and detection rate over the test partition. In such cases, applying multi-objective optimization techniques, to search for an optimum detection rate and interpretability level, could be a direction to take for designing the intrusion detection model. Hence, studying the model with a larger number of tagged objects would be another important extension to this research.

Bibliography

ABADEH, M.S., HABIBI, J. and LUCAS, C., 2007. Intrusion detection using a fuzzy genetics-based learning algorithm. *J.Netw.Comput.Appl.,* **30**(1), pp. 414-428.

ABE, S. and MING-SHONG LAN, 1995. Fuzzy rules extraction directly from numerical data for function approximation. *Systems, Man and Cybernetics, IEEE Transactions on,* **25**(1), pp. 119-129.

ABERDEEN GROUP, I., 1996-2011-last update, Aberdeen Group Research Library [June/27, 2011].

AVOINE, G. and OECHSLIN, P., 2005. A Scalable and Provably Secure Hash-Based RFID Protocol. *Pervasive Computing and Communications Workshops, IEEE International Conference on,* **0**, pp. 110-114.

DAMITH C. RANASINGHE, DANIEL W. ENGELS, PETER H. COLE, *Low-Cost RFID Systems: Confronting Security and Privacy.*

DANG NGUYEN DUC, JAEMIN PARK, HYUNROK LEE, KWANGJO KIM, 2006. Enhancing Security of EPCglobal Gen-2 RFID Tag against Traceability and Cloning.

DAVE FRIEDLOS, 2011-last update, South Korean Consortium Launches EPC Gen 2 Reader for Mobile Phones [Homepage of RFID Journal LLC], [Online]. Available: http://www.rfidjournal.com/article/view/8155/1 [June/28, 2011].

DENNING, D.E., 1987. An Intrusion-Detection Model. *Software Engineering, IEEE Transactions on,* **SE-13**(2), pp. 222-232.

DIMITRIOU, T., 2005. A Lightweight RFID Protocol to protect against Traceability and Cloning attacks, *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on* 2005, pp. 59-66.

EPC GLOBAL INC., 15 December 2010, 2010-last update, The EPCglobal Architecture Framework. Available: http://www.gs1.org/gsmp/kc/epcglobal/architecture/architecture_1_4-framework-20101215.pdf [June/16, 2011].

EPC GLOBAL INC., 2005. *Specification for RFID Air interface: Radio Frequency Identification protocols Class-1 Generation-2 UHF RFID protocol for Communications at 830MHz to 960MHz.*

FREUND, Y. and SCHAPIRE, R.E., 1999. Large Margin Classification Using the Perceptron Algorithm. *Machine Learning,* **37**(3), pp. 277-296.

FULCHER, J. and JAIN, L.C., 2008. *Computational Intelligence: A Compendium.* 1st edn. Springer Publishing Company, Incorporated.

GERSHENFELD, N., 2004. The Internet of things. *Scientific American,* **291**(4), pp. 76.

GOLDBERG, D.E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning.* 1 edn. Addison-Wesley Professional.

HANCKE, G., 2005. *A practical relay attack on ISO 14443 proximity cards.*

HOLLAND, J.H., 1992. *Adaptation in natural and artificial systems.* Cambridge, MA, USA: MIT Press.

ISHIBUCHI, H., 2007. Evolutionary Multiobjective Design of Fuzzy Rule-Based Systems, *Foundations of Computational Intelligence, 2007. FOCI 2007. IEEE Symposium on* 2007, pp. 9-16.

ISHIBUCHI, H., NAKASHIMA, T. and KURODA, T., 1999. A hybrid fuzzy genetics-based machine learning algorithm: hybridization of Michigan approach and Pittsburgh approach, *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on* 1999, pp. 296-301 vol.1.

ISHIBUCHI, H., NOZAKI, K. and TANAKA, H., 1992. Pattern classification by distributed representation of fuzzy rules, *Fuzzy Systems, 1992., IEEE International Conference on* 1992, pp. 643-650.

JIM HURLEY, J.H., May 13 , 2003, 2003-last update, Identity Theft: A $2 Trillion Criminal Industry in 2005 [June/27, 2011].

JUELS, A., 2006. RFID security and privacy: a research survey. *Selected Areas in Communications, IEEE Journal on,* **24**(2), pp. 381-394.

JUELS, A., 2005. Strengthening EPC tags against cloning, *Proceedings of the 4th ACM workshop on Wireless security* 2005, ACM, pp. 67-76.

JUELS, A., RIVEST, R.L. and SZYDLO, M., 2003. The blocker tag: selective blocking of RFID tags for consumer privacy, *Proceedings of the 10th ACM conference on Computer and communications security* 2003, ACM, pp. 103-111.

JUELS, A. and WEIS, S.A., 2007. Defining Strong Privacy for RFID, *Pervasive Computing and Communications Workshops, 2007. PerCom Workshops '07. Fifth Annual IEEE International Conference on* 2007, pp. 342-347.

KARMAKAR, N., 2010. The Evolution of RFID. *Handbook of Smart Antennas for RFID Systems.* pp. 1-12.

KING, B. and ZHANG, X., 2007. Securing the Pharmaceutical Supply Chain using RFID, *Proceedings of the 2007 International Conference on Multimedia and Ubiquitous Engineering* 2007, IEEE Computer Society, pp. 23-28.

KIRSCHENBAUM, I. and WOOL, A., 2006. How to build a low-cost, extended-range RFID skimmer, *In Proceedings of the 15th USENIX Security Symposium* 2006, pp. 43-57.

KOUTAROU, M.O., SUZUKI, K. and KINOSHITA, S., 2003. Cryptographic Approach to Privacy-Friendly Tags, *In RFID Privacy Workshop* 2003.

L. MIROWSKI ET AL., 2008. *A RFID Proximity Card Data Set.* 2008: School of Computing and Information Systems, University of Tasmania, Hobart, Australia.

L. MIROWSKI, AND J. HARTNETT, 2007. Deckard: a system to detect change of RFID tag ownership. **7**, pp. 89-98.

LAHIRI, S., 2006. *RFID Sourcebook.* IBM Press.

LANDT, J., 2005. The history of RFID. *Potentials, IEEE,* **24**(4), pp. 8-11.

LEE, S., JOO, S. and LEE, C., 2005. An Enhanced Dynamic Framed Slotted ALOHA Algorithm for RFID Tag Identification, *Proceedings of the The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services* 2005, IEEE Computer Society, pp. 166-174.

LEHPAMER, H., 2007. *RFID Design Principles.* Norwood, MA, USA: Artech House, Inc.

LEHTONEN, M., MICHAHELLES, F. and FLEISCH, E., 2007. Probabilistic Approach for Location-Based Authentication.

LEHTONEN, M., OSTOJIC, D., ILIC, A. and MICHAHELLES, F., 2009. Securing RFID Systems by Detecting Tag Cloning, *Proceedings of the 7th International Conference on Pervasive Computing* 2009, Springer-Verlag, pp. 291-308.

MARK ROBERTI, May 30, 2011, 2011-last update, Will NFC Dominate Mobile Payments? [Homepage of RFID Journal LLC], [Online]. Available: http://www.rfidjournal.com/article/view/8473 [June/28, 2011].

MARY CATHERINE O'CONNOR, Aug. 23, 2010, 2010-last update, Printed-Electronics RFID Tags Debut [Homepage of RFID Journal LLC.], [Online]. Available: http://www.rfidjournal.com/article/purchase/7821 [June/2011, 2011].

MOTOROLA INC, 2007-last update, Business Benefits from Radio Frequency Identification (RFID). Available: http://www.motorola.com/web/Business/Products/RFID/RFID%20Reader%20Antennas/AN200/_Documents/Static%20Files/RFID_BBRFID_TB_0907_New.pdf [10/17, 2011].

NAUCK, D. and KRUSE, R., 1999. Obtaining interpretable fuzzy classification rules from medical data. *Artificial Intelligence in Medicine,* **16**(2), pp. 149-169.

NOZAKI, K., ISHIBUCHI, H. and TANAKA, H., 1996. Adaptive fuzzy rule-based classification systems. *Fuzzy Systems, IEEE Transactions on,* **4**(3), pp. 238-250.

ONDRUS, J. and PIGNEUR, Y., 2007. An Assessment of NFC for Future Mobile Payment Systems, *Management of Mobile Business, 2007. ICMB 2007. International Conference on the* 2007, pp. 43-43.

RICHARD MACMANUS, April 12, 2010, 2010-last update, DASH7: Bringing Sensor Networking to Smartphones, Available: http://www.readwriteweb.com/archives/dash7_bringing_sensor_networking_to_smartphones.php [June/28, 2011].

ROTTER, P., 2008. A Framework for Assessing RFID System Security and Privacy Risks. *Pervasive Computing, IEEE,* **7**(2), pp. 70-77.

SARMA, S., 2002. *white paper: Towards the 5¢ Tag.*

STAAKE, T., THIESSE, F. and FLEISCH, E., 2005. Extending the EPC Network – The Potential of RFID in Anti-Counterfeiting, *ACM Symposium on Applied Computing — SAC 2005* 2005, ACM Press, pp. 1607-1612.

STOCKMAN, H., 1948. Communication by Means of Reflected Power. *Proceedings of the IRE,* **36**(10), pp. 1196-1204.

WANG, L.-. and MENDEL, J.M., 1992. Generating fuzzy rules by learning from examples. *Systems, Man and Cybernetics, IEEE Transactions on,* **22**(6), pp. 1414-1427.

WEIS, S.A., SARMA, S.E., RIVEST, R.L. and ENGELS, D.W., 2003. Security and Privacy Aspects of Low-Cost Radio Frequency Identification Systems, 2003, Springer-Verlag, pp. 201-212.

WEIS, S., SARMA, S., RIVEST, R. and ENGELS, D., 2004. Security and Privacy Aspects of Low-Cost Radio Frequency Identification Systems. **2802**, pp. 50-59.

WORLD HEALTH ORGANIZATION, EXECUTIVE BOARD, 18 December 2008, 2008-last update, Counterfeit medical products [June/27, 2011].

WU, N.C., NYSTROM, M.A., LIN, T.R. and YU, H.C., 2006. Challenges to global RFID adoption. *Technovation,* **26**(12), pp. 1317-1323.

ZADEH, L.A., 1975. Fuzzy logic and approximate reasoning. *Synthese,* **30**(3), pp. 407-428.

ZADEH, L.A., 1965. Fuzzy sets. *Information and Control,* **8**(3), pp. 338-353.

ZHANG, S., MCCULLAGH, P., NUGENT, C., ZHENG, H. and BLACK, N., 2011. Reliability of Location Detection in Intelligent Environments. **92**, pp. 181-188.