Detection of Malicious Activities Against Advanced Metering Infrastructure in Smart Grid

by

Paria Jokar

B.Sc., University of Science and Technology, Tehran, Iran, 2003
M.Sc., University of Science and Technology, Tehran, Iran, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2015

© Paria Jokar, 2015
Abstract

In this thesis we investigate security challenges in smart grid and propose several algorithms for detecting malicious activities against AMI. Our work includes two parts. In the first part, we focus on the problem of intrusion detection in ZigBee HANs. We study the requirements and challenges of designing intrusion detection systems for HANs, and suggest application of model based intrusion detection and automatic intrusion prevention techniques. Accordingly we design algorithms for detecting and preventing spoofing attacks as an important attack type against wireless networks. We extend this work to design an intrusion detection and prevention system for ZigBee HANs, HANIDPS, which is able to detect and automatically stop various attack types. Through extensive experiments and analysis we show that the proposed method is able to detect and stop the attacks with high precision, low cost and short delay, which makes it suitable for HANs. Considering that in HANIDPS the prevention operation is performed automatically, costs of false positives are low and limited to some network overhead. Also the delay in stopping the attacks is significantly shortened compared to when human intervention is required. This reduces the damages caused by possible attacks.

In the next part, we focus on detection of cyber intrusions that affect the load curve. We suggest that by monitoring abnormalities in customers’ consumption pattern these attacks are detectable. We introduce a consumption pattern based electricity theft detector, CPBETD, which unlike previous techniques is robust against nonmalicious changes in consumption pattern and provides a high and adjustable performance without jeopardizing
Abstract
customers’ privacy. Extensive experiments on real dataset of 5000 customers show the effectiveness of our approach. We also introduce instantaneous anomaly detector, IAD, which by monitoring the usage patterns effectively detects attacks against direct and indirect load control which are some of the major concerns in AMI.
Preface

Hereby, I declare that I am the author of this thesis. Chapters 2-5 encompass work that has been published or is under review. The corresponding papers were co-authored by Prof. Victor C. M. Leung who supervised me through this research. The papers corresponding to Chapters 2, 3, and 5 were also co-authored by Nasim Arianpoo, and one of the papers related to Chapter 4 was co-authored by Hasen Nacanfar. They provided valuable comments on these works. The following publications describe the work completed in this thesis.

**Journal Papers, Published**


- Paria Jokar, Nasim Arianpoo and Victor C. M. Leung, “Spoofing Detection in IEEE 802.15.4 Networks Based on Received Signal Strength,” *Elsevier Ad Hoc Networks*, vol. 11, no. 8, pp. 2648-2660, November 2013. (Chapter 3)


**Journal Paper, Submitted**
• Paria Jokar and Victor C. M. Leung, “Intrusion Detection and Prevention for ZigBee-Based Home Area Networks in Smart Grids,” August 2015 (Chapter 4)

Conference Papers, Published

• Paria Jokar, Nasim Arianpoo and Victor C. M. Leung, “Spoofing Prevention Using Received Signal Strength for ZigBee-Based Home Area Networks,” in Proc. of IEEE International Conference on Smart Grid Communications (SmartGridComm), Vancouver, Canada, October 2013. (Chapter 3)

• Paria Jokar, Nasim Arianpoo and Victor C. M. Leung, “Intrusion Detection in Advanced Metering Infrastructure Based on Consumption Pattern,” in Proc. of IEEE International Conference on Communications (ICC), Budapest, Hungary, June 2013. (Chapter 5)

• Paria Jokar, Hasen Nicanfar and Victor C. M. Leung, “Specification-Based Intrusion Detection for Home Area Networks in Smart Grids,” in Proc. of IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, Belgium, October 2011. (Chapter 4)
Table of Contents

Abstract ................................................................. ii
Preface ................................................................. iv
Table of Contents ..................................................... vi
List of Tables ........................................................... xi
List of Figures ........................................................... xii
List of Abbreviations .................................................. xiv
Acknowledgments ....................................................... xviii

1 Introduction .......................................................... 1
   1.1 Detection of Malicious Activities Against Advanced Metering Infrastructure
       (Trends, Challenges and Needs) ................................ 3
   1.1.1 Intrusion Detection Techniques ............................. 4
   1.1.2 Advanced Metering Infrastructure Threats ................. 6
   1.1.3 Requirements and Challenges of Intrusion Detection in Advanced
         Metering Infrastructure ...................................... 9
   1.2 Summary of Results and Main Contributions .................. 11
   1.3 Thesis Organization ............................................. 14
# Table of Contents

2 Security Issues in Smart Grids ................................. 16
   2.1 Introduction ........................................... 16
   2.2 Smart Grid Architecture ................................. 17
      2.2.1 Advanced Metering Infrastructure .................. 18
      2.2.2 Distributed Generation ............................ 19
      2.2.3 Wide Area Measurement and Control ................ 20
   2.3 Security Threats ........................................ 20
   2.4 Privacy ............................................... 22
   2.5 Cryptography ......................................... 24
   2.6 Key Management ..................................... 29
   2.7 Security Risk Assessment and Management ............... 30
   2.8 Secure System Architecture ............................ 33
   2.9 Summary ............................................. 34

3 Spoofing Detection and Prevention System for ZigBee-Based Home Area Networks .......................... 35
   3.1 Introduction ........................................... 35
   3.2 A Survey on RSS Based Spoofing Detection Methods ........ 38
   3.3 Home Area Network Architecture ........................ 42
   3.4 Threat Model ......................................... 43
   3.5 Spoofing Detection Algorithm ............................ 44
      3.5.1 Operation Phase .................................. 45
      3.5.2 Training Phase ................................... 47
      3.5.3 Multiple Air Monitors ......................... 48
   3.6 Spoofing Prevention Algorithm .......................... 49
      3.6.1 Static Threshold ................................ 50
Table of Contents

3.6.2 Dynamic Threshold ........................................ 50

3.7 Performance Analysis ........................................ 51
  3.7.1 Performance of the Spoofing Detection Algorithm ........ 51
  3.7.2 Performance of the Spoofing Prevention Algorithms ........ 60

3.8 Experiments ................................................ 61
  3.8.1 Testbed ................................................ 61
  3.8.2 Performance of the Spoofing Detection Algorithm ........ 62
  3.8.3 Performance of the Spoofing Prevention Algorithms ........ 66

3.9 Discussions and Comparison .................................. 68

3.10 Summary .................................................. 69

4 HANIDPS, An Intrusion Detection and Prevention System for ZigBee
  HANs ....................................................... 72
  4.1 Introduction ............................................. 72
  4.2 Related Work and Comparison ................................ 73
  4.3 Home Area Network Security Threats .......................... 75
    4.3.1 Case Studies ........................................ 77
  4.4 Home Area Network Intrusion Detection and Prevention System .... 80
    4.4.1 Architecture ........................................ 80
    4.4.2 Features and State Space .............................. 81
    4.4.3 Actions ............................................. 85
    4.4.4 Learning Algorithm .................................. 86
  4.5 Performance Analysis ...................................... 89
    4.5.1 Performance of the Detection Module .................... 89
    4.5.2 Performance of the Prevention Module ................... 90
  4.6 Evaluations .............................................. 92
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6.1 Home Area Network Intrusion Detection and Prevention System</td>
<td>92</td>
</tr>
<tr>
<td>against IEEE 802.15.4 attacks</td>
<td></td>
</tr>
<tr>
<td>4.6.2 Experiment 1</td>
<td>94</td>
</tr>
<tr>
<td>4.6.3 Experiment 2</td>
<td>96</td>
</tr>
<tr>
<td>4.7 Summary</td>
<td>96</td>
</tr>
<tr>
<td>5 Detection of Malicious Activities in AMI Using Customers Consumption</td>
<td>98</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>98</td>
</tr>
<tr>
<td>5.2 Related Work</td>
<td>102</td>
</tr>
<tr>
<td>5.3 Threat Model</td>
<td>106</td>
</tr>
<tr>
<td>5.4 Consumption Pattern Based Electricity Theft Detection Algorithm</td>
<td>107</td>
</tr>
<tr>
<td>5.4.1 Training Phase</td>
<td>107</td>
</tr>
<tr>
<td>5.4.2 Application Phase</td>
<td>111</td>
</tr>
<tr>
<td>5.5 Performance Analysis</td>
<td>113</td>
</tr>
<tr>
<td>5.5.1 Required Performance</td>
<td>113</td>
</tr>
<tr>
<td>5.5.2 Classification Method</td>
<td>115</td>
</tr>
<tr>
<td>5.5.3 Clustering Method</td>
<td>116</td>
</tr>
<tr>
<td>5.6 Performance Evaluations</td>
<td>116</td>
</tr>
<tr>
<td>5.6.1 Experiment 1: One-class Support Vector Machine</td>
<td>119</td>
</tr>
<tr>
<td>5.6.2 Experiment 2: Multi-class Support Vector Machine</td>
<td>120</td>
</tr>
<tr>
<td>5.6.3 Experiment 3: Multi-class Support Vector Machine with New Attacks</td>
<td>122</td>
</tr>
<tr>
<td>5.6.4 Overall Performance of Consumption Pattern Based Electricity Theft</td>
<td>123</td>
</tr>
<tr>
<td>Detection algorithm</td>
<td></td>
</tr>
<tr>
<td>5.6.5 Effect of Sampling Rate on Performance</td>
<td>124</td>
</tr>
</tbody>
</table>
# Table of Contents

5.6.6 Discussion and Comparison .............................................. 125  
5.7 Instantaneous Anomaly Detection Algorithm ............................... 127  
5.8 Performance Evaluation of Instantaneous Anomaly Detection ............... 129  
  5.8.1 Dataset ............................................................................. 129  
  5.8.2 Test Results ...................................................................... 132  
5.9 Conclusion ............................................................................. 134  

6 Conclusions and Future Work .................................................... 135  
  6.1 Summary of Accomplished Work .............................................. 135  
  6.2 Suggestions for Future Work .................................................. 137  

Bibliography .................................................................................. 139
List of Tables

3.1 Comparison between $\lambda$ of RSS and SDC ........................................ 64
3.2 Prevention performance for dynamic threshold spoofing prevention .......... 68
3.3 Comparison of different RSS-based spoofing detection techniques (NA: not applicable, - : was not provided in the paper, *: average of the results of 4 AMs) ................................................................. 68
4.1 Example of attacks against ZigBee HANs ................................................. 78
4.2 Requirements of HAN according to the U.S. department of energy guideline. 83
4.3 Performance for dynamic threshold spoofing prevention. ......................... 96
5.1 Experiment results .............................................................................. 122
5.2 BDR for different values of $m$ ............................................................. 124
5.3 Effect of sampling rate on detection performance .................................... 125
5.4 Comparison among energy theft detection methods ............................... 127
5.5 Appliance load profile ....................................................................... 130
5.6 Cross classification and detection performance of IAD .......................... 133
List of Figures

1.1 Architecture of AMI. .......................................................... 3

2.1 Integration of a data layer to different parts of power system in smart grid al-
allows automated transmission and distribution as well as energy conservation
and generation based on demand.(Image by Southern California Edison) . 18

3.1 Architecture of a CP-HAN. .................................................. 43

3.2 a) Attack scenario. b) CDF of the RSS values of the attacker and the genuine
nodes. ..................................................................................... 52

3.3 Effect of ratio of malicious traffic on performance of the first step of the
spoofing detection algorithm ................................................. 55

3.4 Effect of the difference between RSS mean values on performance of the first
step of the spoofing detection algorithm .................................. 56

3.5 Discrete wavelet transform decomposition algorithm .......................... 56

3.6 Effect of increase in the number of AMs and non-centrality parameter ($n$
represents the number of AMs and $k$ is the scale of non-centrality) ........ 59

3.7 Theoretic ROC for static threshold spoofing prevention ($m$ and $m'$ are the
mean RSS values of the genuine and attacker nodes) ....................... 61

3.8 Expected number of tries for successful transmission vs. threshold ........ 61

3.9 Expected number of tries for successful transmission vs. FPR ................ 62

3.10 Testbed setting (the distance between consecutive grid dots is 50 cm) . . . 63
3.11 ROC curve of the spoofing detection algorithms for a) AM1, b) AM2, c) AM3, d)AM4 ......................................................... 66
3.12 Average ROC of the spoofing detection algorithm. ............... 67
3.13 ROC of spoofing detection in AM4 based on a) R, b) SDC, c) proposed algorithm.) .................................................. 70
3.14 Experimental ROC for static threshold spoofing prevention. ... 71
3.15 Average experimental ROC of V1, V2, V3 and V4 for static threshold spoofing prevention. .................. 71
4.1 ROC of the detection module. ........................................... 95
5.1 Pseudo-codes of application phase .................................... 114
5.2 An example of the daily consumption. ................................. 118
5.3 An example of the daily consumption. ................................. 119
5.4 ROC curves for three customers with the worst, average and the best detection performance using one-class SVM. .................. 120
5.5 Silhouette plots for customers with \(k=1\) and \(k=2\). ............ 121
5.6 ROC curves for three customers with the worst, average, and the best detection performance using multi-class SVM. ............. 121
5.7 ROC curves for three customers with the worst, average and the best detection performance using multi-class SVM with undefined attacks. .......... 122
5.8 Relationship between BDR and FPR. ................................. 123
5.9 Detection accuracy of IAD for false price attack as a function of difference between the usage probabilities during high and low price. ........ 134
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>Advanced encryption standard</td>
</tr>
<tr>
<td>AGC</td>
<td>Automatic generation control</td>
</tr>
<tr>
<td>ALG</td>
<td>Application layer gateway</td>
</tr>
<tr>
<td>AM</td>
<td>Air monitor</td>
</tr>
<tr>
<td>AMI</td>
<td>Advanced metering infrastructure</td>
</tr>
<tr>
<td>C-IDPS</td>
<td>Central IDPS</td>
</tr>
<tr>
<td>CAK</td>
<td>Cumulative attestation kernel</td>
</tr>
<tr>
<td>CBR</td>
<td>Constant bit rate</td>
</tr>
<tr>
<td>CMAC</td>
<td>Cipher-based message authentication code</td>
</tr>
<tr>
<td>CP-HAN</td>
<td>Consumer private HAN</td>
</tr>
<tr>
<td>CPDETD</td>
<td>Consumption pattern-based energy theft detector</td>
</tr>
<tr>
<td>CS</td>
<td>Central server</td>
</tr>
<tr>
<td>CSMA-CA</td>
<td>Carrier sense multiple access - collision avoidance</td>
</tr>
<tr>
<td>D</td>
<td>Datagram</td>
</tr>
<tr>
<td>DER</td>
<td>Distributed energy resources</td>
</tr>
<tr>
<td>DHWT</td>
<td>Discrete Haar wavelet transform</td>
</tr>
<tr>
<td>DR</td>
<td>Detection rate</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete wavelet transform</td>
</tr>
<tr>
<td>DoS</td>
<td>Denial of service</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ECDSA</td>
<td>Elliptic curve digital signature algorithm</td>
</tr>
<tr>
<td>ED</td>
<td>Energy detection</td>
</tr>
<tr>
<td>EMS</td>
<td>Energy management system</td>
</tr>
<tr>
<td>EPRI</td>
<td>Electric power research institute</td>
</tr>
<tr>
<td>ESI</td>
<td>Energy service interface</td>
</tr>
<tr>
<td>ETDS</td>
<td>Energy theft detection systems</td>
</tr>
<tr>
<td>FDI</td>
<td>False data injection</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier transform</td>
</tr>
<tr>
<td>FPR</td>
<td>False positive rate</td>
</tr>
<tr>
<td>GTS</td>
<td>Guaranteed time slot</td>
</tr>
<tr>
<td>HAN</td>
<td>Home area network</td>
</tr>
<tr>
<td>HFID</td>
<td>High-frequency identifier</td>
</tr>
<tr>
<td>HTTPS</td>
<td>Hyper text transfer protocol with secure sockets</td>
</tr>
<tr>
<td>IAD</td>
<td>Instantaneous anomaly detector</td>
</tr>
<tr>
<td>IBE</td>
<td>Identity-based encryption</td>
</tr>
<tr>
<td>IBS</td>
<td>Identity-based signcryption</td>
</tr>
<tr>
<td>ICS</td>
<td>Industrial control systems</td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>IDPS</td>
<td>Intrusion detection and prevention system</td>
</tr>
<tr>
<td>IDS</td>
<td>Intrusion detection system</td>
</tr>
<tr>
<td>IED</td>
<td>Intelligent electronic devices</td>
</tr>
<tr>
<td>IPS</td>
<td>Intrusion prevention system</td>
</tr>
<tr>
<td>IT</td>
<td>Information technology</td>
</tr>
<tr>
<td>KGS</td>
<td>Key-generating server</td>
</tr>
<tr>
<td>LFID</td>
<td>Low-frequency identifier</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>LQI</td>
<td>Link quality indicator</td>
</tr>
<tr>
<td>LUT</td>
<td>Look up table</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium access control</td>
</tr>
<tr>
<td>NA</td>
<td>Node availability</td>
</tr>
<tr>
<td>NALM</td>
<td>None-intrusive appliance load monitoring</td>
</tr>
<tr>
<td>NAN</td>
<td>Neighborhood area networks</td>
</tr>
<tr>
<td>NILM</td>
<td>Nonintrusive load monitoring</td>
</tr>
<tr>
<td>PDC</td>
<td>Phasor data concentrators</td>
</tr>
<tr>
<td>PER</td>
<td>Packet error rate</td>
</tr>
<tr>
<td>PHY</td>
<td>Physical</td>
</tr>
<tr>
<td>PKI</td>
<td>Public key infrastructure</td>
</tr>
<tr>
<td>PMU</td>
<td>Phasor measurement units</td>
</tr>
<tr>
<td>PR</td>
<td>Prevention rate</td>
</tr>
<tr>
<td>RSS</td>
<td>Received signal strength</td>
</tr>
<tr>
<td>SCADA</td>
<td>Supervisory control and data acquisition</td>
</tr>
<tr>
<td>SDC</td>
<td>Summation of detailed coefficients</td>
</tr>
<tr>
<td>SDPS</td>
<td>Spoofing detection and prevention system</td>
</tr>
<tr>
<td>SEP</td>
<td>Smart energy profile</td>
</tr>
<tr>
<td>SN</td>
<td>Sequence number</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission control protocol</td>
</tr>
<tr>
<td>TR</td>
<td>Traffic rate</td>
</tr>
<tr>
<td>UE-HAN</td>
<td>Utility enabled HAN</td>
</tr>
<tr>
<td>CP-HAN</td>
<td>Customer private HAN</td>
</tr>
<tr>
<td>WAMAC</td>
<td>Wide area measurement and control</td>
</tr>
</tbody>
</table>
List of Abbreviations

WAN Wide area networks
Acknowledgments

First of all, I would like to express my deepest sense of gratitude to my supervisor Prof. Victor C. M. Leung for his continuous support, patience, technical and non-technical guidance that helped me accomplish this research.

I am also thankful to my friends Maziar Tabarestani, Hasen Nicanfar, Nasim Arianpoo and Pedram Samadi for their always being supportive and accountable.

Most importantly, I feel indebted to my mother and father for their everlasting love, support, and protectiveness not only in years of doing PhD but also throughout my entire life.

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.
Chapter 1

Introduction

Smart grid is a vision to modernize the electricity transmission and distribution systems by incorporating computer intelligence into the power system, and providing two way energy flow and data communication. Advanced metering infrastructure (AMI) is a subsystem within the smart grid which provides two-way data communication between the smart meters and the utility company. AMI enables real-time transmissions of power consumption and pricing information, as well as control commands.

Unlocking the tremendous potentials of smart grid such as resilience, high power quality, and consumer participation, strongly depends on security of this system. Integration of a data layer to the power system can expose it to many cyber security threats. Without strong security measures in place, not only the smart grid will inherit the vulnerabilities of the legacy power system, but also new vulnerabilities will be added due to the proliferation of new technologies. From the advent of the smart grid concept, security has always been a primary concern. In the 2009 White House cyberspace policy review [1], federal government was asked to ensure that security standards are developed and adopted to avoid creating unexpected opportunists to penetrate these systems or conduct large-scale attacks.

Along with security mechanisms that must be designed into the smart grid with the goal of reducing the vulnerabilities and mitigating their consequences, such as cryptographic algorithms and secure protocols, appropriate intrusion detection systems (IDSs) and intrusion prevention systems (IPSs) with the ability to detect and prevent malicious activities resulting from exploiting the vulnerabilities in the system should also be in place. The
need for research on intrusion detection for embedded processors within the smart grid was emphasized in the United States (US) National Institute of Standards and Technology (NIST) guidelines for smart grid cyber security [2], in that smart grid contains a large number of processors with limited resources and strict timeliness requirements. Although AMI deployment is still in pilot phase in many jurisdictions, several security flaws have already been reported for multiple metering devices [3],[4]; this emphasizes on the need for developing appropriate IDSs for AMI communication networks.

In this thesis, we study security issues in AMI and requirements of IDS/IPS for AMI networks. Accordingly, we propose various algorithms for detecting and preventing malicious activities against AMI. The proposed algorithms are tailored for unique characteristics of AMI networks. Mathematical tools and machine learning techniques are applied in designing the algorithms, and their performances are evaluated through extensive analysis, simulations and experiments. The rest of this chapter is organized as follows. In Section 1.1 we explain general techniques and trends for intrusion detection. Then, we describe new threats introduced against the power system due to addition of a data layer. Further, we overview requirements and challenges of detecting malicious activities against AMI. Section 1.2 provides a summary of the main contributions in this thesis, and discusses the significance and novelty of the proposed mechanisms. Finally, the organization of the thesis is described in Section 1.3.
1.1 Detection of Malicious Activities Against Advanced Metering Infrastructure (Trends, Challenges and Needs)

AMI includes several communication networks as shown in Figure 1.1. Home area networks (HANs) are responsible for demand-side management. Home electric devices such as appliances and thermostats, communicate with the smart meters through HANs to exchange price and usage data as well as control commands. Smart meters’ data in each neighborhood is collected by data aggregators before being sent to the utility. Communication between smart meters and data aggregators is provided by neighborhood area networks (NANs). Wide area networks (WANs) enable the communication between data aggregators and the utility networks. AMI infrastructure assets are divided into private and public domains. The private domain includes systems that are similar to standard information technology (IT) assets. These systems contain a large amount of critical data; yet they are located in data centers which are secure environments. The public domain assets, on
the other hand, are located in physically insecure environments which makes them more vulnerable to cyber threats. This necessitates development of appropriate IDSs/IPSs.

1.1.1 Intrusion Detection Techniques

As it was defined in [5], “intrusion detection is the process of monitoring the events that occur in a computer system or network and analyzing them for signs of possible incidents.” There are three key metrics commonly used for measuring the performance of an IDS, detection rate (DR), false positive rate (FPR) and Bayesian detection rate (BDR). DR is defined as the number of intrusion instances detected by the system divided by the total number of intrusion instances present in the test set. FPR is defined as the number of normal patterns classified as attacks divided by the total number of normal patterns. BDR is the probability of occurrence of an intrusion once an attack is detected by the IDS. Assuming that \( A \) represents an intrusion alarm by the IDS, and \( I \) shows the occurrence of an intrusion these metrics are formulated as follows:

\[
DR = P(A|I) \quad (1.1)
\]

\[
FPR = P(A|\bar{I}) \quad (1.2)
\]

\[
BDR = P(I|A) \quad (1.3)
\]

Usually there is a trade-off between DR and FPR which is represented through receiver operating characteristic (ROC) curves. An ROC curve shows the values of DR for different values of FPR.

In general, there are three types of IDSs based on the method they use for recognizing malicious activities.

- Signature-based IDSs have a database of predefined attack patterns, known as signa-
Chapter 1. Introduction

tures, and detect intrusions by comparing the system behavior with the signatures.

- Anomaly-based IDSs detect malicious activities as deviations from statistically normal behavior of the system.

- Specification-based IDSs also recognize intrusions as deviation from normal behaviors; however, instead of statistical methods, normal behaviors are defined based on manually extracted specifications of the system.

Signature-based IDSs have low FPR, yet they are incapable of detecting unknown attacks and their databases need to be updated frequently. Existing anomaly-based IDSs, on the other hand, suffer from high FPR and require long training and tuning time, yet they are able to detect new attacks. Specification-based IDSs potentially have low FPR and the ability to detect new attacks. However, the strength of this type of IDS depends on the accuracy and efficiency of selected specifications. Considering that many of the AMI equipment apply new technologies, an exhaustive database of known attacks is not available. Thus, signature-based IDSs are not appropriate in the context of AMI.

In this work, we design new anomaly-based and specification-based algorithms for detecting malicious activities against AMI. Specifically, we are focused on intrusion detection and prevention in HANs, as well as detection of newly introduced attacks against the power system such as attacks against direct and indirect load control and energy theft attempts.

The concept of intrusion detection was first introduced in 1980 in [6], where audit trails were suggested as valuable information that can be utilized to detect anomalous behavior as sign of malicious activities. In 1984 the first model of intrusion detection, Intrusion Detection Expert System, was born [7]. Until 1990, the majority of IDSs were host-based [8] where individual host level audit records were analyzed. In 1990 the concept of network IDS [9] was introduced in which instead of host behavior, network traffic was monitored for signs of intrusion. The first system under this category was Network Security Monitor.
Chapter 1. Introduction

[10]. IDS gradually entered the commercial market with products such as NetRanger and RealSecure [11]. One can look at intrusion detection as a classification problem in which the signatures or features of the system are classified under benign or malicious categories. Since 1995 various classification techniques were used to achieve a higher detection performance, including neural networks [12], genetic algorithms [13], state transition [14], fuzzy logic [15], and decision tree [16].

Recent research on IDS is directed to target specific systems or attack types rather than general solutions. This approach allows to design IDSs tailored for requirements, limitations and characteristics of a given system or attack type which yields a better detection performance. In the context of IDS for AMI a few work have been done over the last few years. Based on the algorithms introduced in each chapter of this thesis, we provide a literature review of related works in the corresponding chapter separately.

1.1.2 Advanced Metering Infrastructure Threats

Increase in use of information technology for demand side management introduces new types of cyber-intrusions. AMI threats are categorized into three primary groups [17]: customer attacks, insider attacks, and terrorist or nation-state attacks. These threats can cause cyber impacts such as loss of availability and integrity to the AMI system or the bulk electric controls. Consequences of these cyber effects on power system, range from increase in peak usage to widespread outages. The lowest level, highest probability threat to AMI is the unethical customer. Next highest in probability is the insider who has financial motivation. Beyond these threats are the high level, low probability threats of nation-state or terrorist groups.
Unethical customer threat

One of the major concerns in AMI is energy theft. Energy theft causes billions of dollars of financial loss to the utilities every year. Traditional energy theft techniques require mechanical manipulation of analog electric meters. However, employment of smart meters, introduces numerous new vectors for energy theft. In AMI, the usage data can be tampered with before recording, in the smart meter or during the transmission to the utility company. Software-based attacks usually need less knowledge and expertise, therefore, are more probable to become widespread. Utilities rely on blink count for detecting theft attempts, in which theft is detected by counting the number of times the meter has been de-energized as a sign of customer attempt for tampering with the meter. This, however, does not cover many attacks that can happen over the communication links or through manipulation of wires ahead of meters. New techniques need to be developed to efficiently detect energy theft attempts.

The unethical customer requires low funding, has extended physical access to the AMI devices in his premise, has a long time and moderate commitment to achieve his goal, and needs low to high cyber skills. The threat might arise from small groups who have the knowledge to develop tools to penetrate advanced metering devices, and become widespread when these tools go commercial. The primary cyber effect of customer attack will be under-report of customer usage. A secondary and probably unintended cyber effect can be failure in reporting correct outage or status information.

The consequences of customer attacks can be low if only a few customers modify their metering data, to severe when the attacks become widespread and available to customers with moderate technical knowledge. The immediate consequence of customer attacks is decreased profitability. In long-term utilities will not be able to plan the infrastructure based on demand, since the load information is not valid. Besides, invalid outage information
increases the operational costs and required time for disaster recovery.

**Insider threat**

One of the AMI goals is to manipulate the load curve to reduce peak usage. This helps to reduce the cost of energy for both the utility and the end-customer. A generation provider who sells energy to the utility may collaborate with an insider to attack the AMI in order to make money. For instance, by modifying the pricing calculation software in the head-end server the insider can use AMI to increase peak usage. This increases the demand for generating high price electricity; therefore, the generation provider will make more money. The insider requires low funding and cyber skills, has low commitment to the goal, high stealth, high physical access, and long time for implementing the attack. The insider takes advantage of access to systems at opposite end of the AMI from the customer end-point; some examples include AMI head-end, the system that provides pricing information (such as energy management system or inter control center protocol server), and the network infrastructure for these systems. The impacts of insider attack on electric system include higher peak electricity usage and artificially high usage reporting for planning.

**Nation threat**

This is a high-level threat which targets AMI to cause damages outside the AMI system, for instance to affect the bulk electric grid. Unauthorized access to sensitive AMI devices from the customer endpoint as well as mass load manipulation constitute the major concerns that enable nation threat.

AMI potentially allows access to the electric grid from the customer endpoints, since there is an inherent connection from the endpoint to sources of pricing information in the utility. By penetrating to the customer endpoint or cracking the wireless communication between the AMI meter and other endpoint equipment or from AMI to the local aggrega-
tors, the adversary can gain access the head end equipment. Details of this attack strongly depend on the AMI implementation.

The primary goal of AMI as a demand-response system is reduction of peak load through managing the customers’ usage. This happens through direct and indirect load control programs. By tampering the pricing information or direct load control commands an adversary can control the load curve to cause damage to the power system equipment. For instance, in an attack against direct load control, the attacker sends turn of messages to all controllable customer equipment. After passing a long enough time which guarantees most of the equipment would turn on when allowed, the attacker sends turn-on permission messages. This results in a sudden increase in the load, which can affect the bulk electric grid.

1.1.3 Requirements and Challenges of Intrusion Detection in Advanced Metering Infrastructure

In designing intrusion detection mechanisms for AMI, specific characteristics and constraints of this system must be considered. It has been repeatedly argued that existing IDS solutions designed for traditional IT systems are not applicable in AMI [18]. AMI is a large scale system including several communication networks with different technologies. Traditional IDSs including a number of lightweight agents reporting to a central management server is not applicable in such a system. AMI networks contain millions of nodes. With a central approach for monitoring and intrusion detection, the traffic load, required storage and computational capabilities at the central server will be overwhelming.

One of the most vulnerable systems in AMI is HAN. HANs are located in public domain, and therefore are easier targets for attackers. In North America and many other countries, wireless is the dominant HAN technology. Ease of accessibility of HAN devices
and application of wireless technology, which uses a shared medium, makes them vulnerable to cyber attacks. At the same time, due to the resource and computational constrains of HAN devices, implementation of strong security mechanisms is a challenging issue.

There are many unanswered questions in the area of designing IDSs for HANs. For instance, who should be responsible for managing the IDS alarms? What is the cost of false positives and how much false positives are bearable for a HAN IDS? What are the differences between HANs and other existing sensor networks, and how these differences affect the process of monitoring and intrusion detection? As it was described in [2] considering the limited resources of sensor nodes, the IDSs/IPSs must not impose high computational and storage load on network nodes. At the same time, due to the large scale, FPR for an IDS must be very low without sacrificing the DR. Otherwise, it will introduce a large operational cost. Designing an IDS with high performance and low resource usage is very challenging. In the area of IDS for sensor networks, focusing on specific attack types rather than general solutions was suggested [19],[20],[21],[22],[23]. Still performances of the proposed methods need to be significantly improved to be used in practice. Although a lot of work has been done in the area of intrusion detection, the area of intrusion prevention has been neglected, while automatic prevention can overcome many challenges of managing the IDS alarms. As many of the AMI deployments are located in areas far from the utility, receiving IDS alarms by the utility and acting upon them introduces a large operational cost and delay in stopping the attacks. In traditional IDSs, once an attack is detected, an alarm is sent to a network operator who is responsible for finding the roots of the attack and triggering response operations varying from remote diagnosis to on-site inspection. Considering the large scale of the smart grid, when human response is expected, a small percentage of false alarms results in a high operational cost. Therefore, in the context of AMI, intrusion prevention mechanisms which not only detect but also automatically stop
1.2 Summary of Results and Main Contributions

This thesis investigates security challenges in smart grid and proposes several algorithms for detecting and preventing malicious activities against AMI. The main contributions in this thesis are as follows:

- We investigate cyber security and privacy issues in smart grid and challenges of securing this system. We survey existing solutions to enhance the security and privacy of smart grid and provide directions for further research.

- Identity spoofing is an important class of attacks which can be used as a basis for several other attack types to penetrate or disturb the operation of a ZigBee HAN. Considering the openness property of the transmission medium and inadequate resources for implementing strong security measures, HANs are highly vulnerable to this attack. To determine whether an identity belongs to a legitimate entity or has been counterfeited by a malicious node, forge-resistant parameters are employed. One property that has recently attracted the attention of researchers for detecting spoofing attacks in ZigBee and WiFi networks, is received signal strength (RSS) values. Several RSS-based techniques have been proposed over the last few years [19],[20],[21],[22],[23]. Yet, existing methods have limited performances, require multiple air monitors (AMs) to provide an acceptable detection accuracy, and are vulnerable to environmental changes. Furthermore, there has been no attempt regarding automatic spoofing prevention using RSS values in the past. As we discussed in Section 1.1.2, automatic prevention in networks like HAN is of great importance.

In this thesis we propose a novel high performance RSS-based spoofing detection
for ZigBee-based HANs. By extracting magnitude and frequency related features of RSS stream and adaptive learning of the distribution of RSS values, the proposed algorithm provides a superior performance compared to previous works and is robust against environmental changes. We also introduce two techniques for preventing spoofing attacks using RSS values. Extensive analysis and experiments show a high performance for the proposed methods. The results of this work were published in [24] and [25].

• Several IDSs tailored for AMI networks have been proposed over the last few years. Some of the suggested algorithms [26],[27], apply the same method for detecting attacks in all parts of the AMI. Using the same solution for different AMI networks including HAN, NAN and head end which use different protocols and have different traffic features makes these methods inefficient. Some works are focused on specific attack types like false data injection [28],[29], and some cover specific parts of AMI, like NAN [30],[31] and supervisory control and data acquisition (SCADA) [32],[33].

In this thesis we focus on designing a novel high performance intrusion detection and prevention system for ZigBee-based HANs, HANIDPS, as one of the most vulnerable AMI networks. HANIDPS utilizes a model-based IDS along with a dynamic machine learning-based prevention technique to detect and prevent malicious activities without prior knowledge of attacks. Considering that in HANIDPS the prevention operation is performed automatically, the costs of false positives are low and limited to some network overhead. Also the delay in stopping the attacks is significantly shortened compared to when human intervention is required. This reduces the damages caused by possible attacks. Through analysis and experiments we show that HANIDPS is able to detect various attack types with a high performance. The results of this work was submitted to [34] and published in [35].
• As described in Section 1.1.2, energy theft is one of the major threats against AMI. Current AMI energy theft detection systems (ETDSs) are mainly categorized into three groups, state-based, game theory-based and classification-based. State-based detection schemes [28],[36],[37] employ specific devices, like wireless sensors and radio-frequency identifiers (RFID), to provide a high detection accuracy. This, however, comes with the price of extra investment required for the monitoring system including device cost, system implementation cost, software cost and operating/training cost. In game theory-based methods [38],[39], the problem of electricity theft detection is formulated as a game between the electricity thief and the electric utility. These methods may present a low cost and reasonable, though not optimal, solution for reducing energy theft. Yet, how to formulate the utility function of all players, including thieves, regulators and distributors, as well as potential strategies is still a challenging issue. Classification-based approaches [40],[41],[42],[43],[44],[45],[46],[47] take advantage of the detailed energy consumption measurements collected from the AMI. Under normal condition customers consumption follow certain statistical pattern; irregularities in usage pattern can be a sign of some malicious activities. Since these techniques take advantage of the readily available smart meter data, their costs are moderate. However, there are several shortcomings in existing classification-based schemes which limits their DR and causes a high FPR, including use of imbalanced data, dependency on high sampling rates which threatens customers privacy, vulnerability to non-malicious changes in consumption patterns and contamination attacks.

In this thesis we present a novel consumption pattern-based energy theft detector (CPBETD) which is robust against non-malicious changes in usage pattern and provides a high and adjustable performance with a low sampling rate. Therefore, the proposed method does not invade customers’ privacy. Extensive experiments on real
Chapter 1. Introduction

dataset of 5000 customers show a high performance for the proposed method. We also introduced instantaneous anomaly detector (IAD), an algorithm for detecting attacks against direct and indirect load control. IAD monitors the consumption pattern of a group of customers within a NAN, and detects abnormal behaviors with short delay. We have provided a synthetic dataset to evaluate the performance of IAD and observed satisfactory results. The results of these works were published in [48] and [49].

In summary this thesis proposes several novel algorithms for detecting malicious activities against AMI. The thesis includes two parts. In the first part we target the area of intrusion detection in ZigBee-based HANs which are among the most vulnerable AMI sub-systems. We investigate challenges and requirements of intrusion detection and prevention for HANs and accordingly present a solution that meets these requirements. In the second part, we suggest that the fine-grained usage measurements can be used to detect suspicious behaviors in AMI that can be originated from any of the AMI networks including HANs, NANs and WANs. We develop algorithms which by monitoring irregularities in customers consumption patterns detect major types of malicious activities against AMI, including energy theft and attacks against direct and indirect load control mechanisms.

1.3 Thesis Organization

The rest of the thesis is organized as follows.

In Chapter 2, we introduce several security issues in the smart grid. We survey existing literature on different security aspects of the smart grid, and provide directions for further research. In Chapter 3, we propose an algorithm for detecting spoofing attacks in ZigBee HANs in the future smart grid. The proposed method aims to detect spoofing attacks with
a high performance and low resource usage. Further we introduce algorithms for automatically stopping spoofing attacks once they are detected. We evaluate the performance of the proposed methods through extensive experiments and analysis. In Chapter 4, we propose an intrusion detection and prevention system for ZigBee based HANs which is able to detect several attack types without prior knowledge of the attacks. We extract features of HAN network traffic using the corresponding standards to design a specification based IDS. We also introduce some defense mechanisms for mitigating the attacks and suggest a machine learning approach to find the best strategy against a given attack. In Chapter 5, we propose algorithms for detecting malicious activities against AMI that work based on monitoring abnormalities in customers’ consumption patterns. We introduce CPBETD for detecting energy theft attempts with high precision, and IAD to detect attacks against direct and indirect load control. Through extensive experiments we show the effectiveness of the proposed methods. Finally, the thesis is concluded and some potential future work is introduced in Chapter 6. Each of the main chapters in this thesis is self-contained and included in separate journal articles or conference papers.
Chapter 2

A Survey on Security Issues in Smart Grids

2.1 Introduction

The existing power system, which provides one-way electricity distribution from central power plants to the consumers, is inefficient, unreliable and outdated. This legacy system is not capable of responding to the increasing demand for energy in the near future, nor to satisfy the requests of today’s modern life. According to [50] power outages resulting from power system vulnerabilities are estimated to cost about 100 billion dollars per year in the United States (US). In August 2003 a cascading outage of generation and transmission facilities in the North America, caused a massive blackout in Northeastern and Midwestern US and Ontario Canada, which resulted in a 6 billion dollar loss. At the same year, major blackouts also happened in several European countries including Denmark, Sweden, and Italy [51].

The vision of smart grid has the potential to bring reliability, efficiency, flexibility, resilience, robustness, and consumer participation to the electricity system by adding a cyber layer to the power grid and providing two-way energy flow and data communications. The two-way energy flow allows the easier integration of the renewable energy, such as solar and wind, into the electricity system and enables distributed energy generation. The widespread usage of plug-in hybrid electric vehicles (PHEVs) requires a smart grid where
customers are aware of the usage prices; accordingly they charge their vehicles in the low tariff period when the consumption is low [52]. Intelligent monitoring and control of the power system as well as demand side management relies on the two-way communication among the smart grid deployments.

Addition of the 2-way data communication layer, however, can expose the power system to many cyber security threats. Beside security vulnerabilities introduced by the expansion of information system, complexity, highly time sensitive operational requirements and large number of stakeholders will introduce additional risks to the power system.

In this chapter we introduce several cyber security and privacy issues in smart grid. We explain the major security challenges that must be considered in the context of the smart grid. Moreover, we investigate the solutions that have been proposed by researchers, and provide directions for further research to address the existing security problems. The rest of this chapter is organized as follows. In Section 2.2, we briefly introduce the architecture of the smart grid. Section 2.3 explains the significant causes of smart grid security concerns. Section 2.4 addresses the privacy issues. Cryptography solutions and key management challenges are covered in Section 2.5 and 2.6 respectively. Section 2.7 investigates security risk assessment and management issues. Section 2.8 addresses the secure architectures for the smart grid, and Section 2.9 summarizes the chapter.

### 2.2 Smart Grid Architecture

According to a definition by electric power research institute (EPRI) smart grid is ”a power system made up of numerous automated transmission and distribution systems, all operating in a coordinated, efficient and reliable manner. This power system serves millions of customers and has an intelligent communication infrastructure enabling the timely, secure, and updatable information flow needed to provide power to evolving digital
Chapter 2. Security Issues in Smart Grids

Figure 2.1: Integration of a data layer to different parts of power system in smart grid allows automated transmission and distribution as well as energy conservation and generation based on demand. (Image by Southern California Edison)

economy”. Figure 2.1 shows integration of the data layer to different parts of power system in smart grid. A conceptual reference model has been presented in the US NIST smart grid interoperability record [53]. In this model the smart grid is divided into seven domains including generation, transmission, distribution, markets, operations, service provider, and customer. This reference model explains the interfaces among domains, networks and actors within each domain, and communication between domains and their gateway. From the operational point of view, smart grid is divided into three domains: AMI, distributed generation, wide area measurement and control (WAMAC).

2.2.1 Advanced Metering Infrastructure

AMI provides two-way communications between smart meters and utility companies which enable real-time transmission of power consumption and pricing data, as well as control commands. This helps customers to adjust their consumption according to the pricing information and helps the utilities to deal with the overload caused by peak electricity demand.
In order to provide a security profile for AMI, the UtiliSec working group has presented a logical architecture for AMI in [54]. The authors believe that although technology choices and features of an AMI system might vary among utilities, at the logical level they are very similar. They have extracted a logical architecture including the logical components of an AMI system and their communications, based on the detailed scenarios and use cases provided by AMI Enterprise community. This work is valuable in that it provides an inclusive source of information about several components of an AMI system, their information and control signal flow, and their connections, which can be used as a basis for security analysis and developing security solutions for AMI systems.

2.2.2 Distributed Generation

One of the major goals of smart grid is to increase the reliability of the power system. This can be achieved by employment of distributed energy resources (DER). DERs are small-scale power generators. A DER can be a fossil fuel generator like a natural gas turbine, a renewable energy generator such as a wind turbine or a roof top solar, as well as a battery that can save energy over the low-price periods and sell it back during peak usage hours. Unlike traditional power grid that relies on centralized generation and provides one way power flow from the central generators to the customers, smart grid utilizes two-way power flow. This enables the customers to sell back the excessive power that has been generated by their local microgrids, to the utilities. DER can relieve grid congestion by diverting power from the power surges; and therefore increases the grid robustness against shut downs. Intelligent integration and management is required in order to efficiently combine these devices into the distribution grid. Intelligent integration and management relies on the real-time two-way communication network between DER elements and utilities.
2.2.3 Wide Area Measurement and Control

WAMAC is applied to improve the reliability and visibility of the power system by synchronously measuring the instantaneous state of the system. Unlike traditional power system where decision making and executing is performed in the range of multi-seconds to multi-minutes, WAMAC enables the future smart grid to take action within 100 milliseconds time frame. WAMAC consists of phasor measurement units (PMUs) or syncrophasors, which are digital instruments responsible for measuring the parameters of electrical networks, phasor data concentrators (PDCs) responsible for collecting measured information, and supervisory control and data acquisition (SCADA) responsible for central control. The real-time operation and high level of interconnectivity, however, might expose the system to outside attacks.

2.3 Security Threats

The future smart grid is expected to enhance the security and reliability of the existing power system. Without strong security measures in place, however, not only the smart grid will inherit the vulnerabilities of the legacy power grid, but also new vulnerabilities will be added due to application of the new technologies. Many security threats have been reported for the legacy power system until now. In March 2007, the Department of Energy’s (DOE) Idaho National Laboratory (INL) conducted an experiment named "Aurora Generator Test". In this experiment, the exploitation of a security vulnerability in SCADA system caused physical damage to a diesel generator [55]. Later, in 2008, the INL published a report in which several vulnerabilities of the SCADA system were categorized and described [56]. Still there are many flaws in the legacy power system which are not publicly announced. Considering that the smart grid is built on top of the
existing power grid it is crucial to improve the security of the legacy system.

The supporting technology for the smart grid includes several devices located in physically insecure environments, such as smart meters, intelligent appliances, distributed generation and storage equipment. These devices have 2-way communication to the electric system, and therefore add numerous entry points to the grid. Because of their unprotected locations, it is easier for attackers to exploit the vulnerabilities of these devices and either cause local damages or benefiting from the 2-way communication, gain access to the more critical parts of the network. In [3], the authors explain how important data like authentication keys can be extracted from the memory of the smart meters and malicious codes can be inserted into these devices to launch attacks against other parts of the grid. Considering the large scale of the smart grid, one single software vulnerability in a device, like a smart meter, can be used to compromise millions of devices.

Wireless technologies are widely used in the smart grid deployments, because of their low-cost, low power consumption, ease of installation, etc. On the other hand, wireless networks are inherently more vulnerable to several types of passive and active attacks, such as eavesdropping and denial of service, compared to wired networks, in that they communicate through the shared medium of air. ZigBee standard which is the dominant technology for HANs in North America is still in early stages and its security has not been evaluated broadly. Still serious vulnerabilities have been reported in Zigbee protocol [57],[58],[59],[60].

Smart grid is an attractive target for different attackers with various motivations. Unethical customers, publicity seekers, curious or motivated eavesdroppers, etc. might target the grid for a variety of malicious reasons. Smart grid is a critical infrastructure that many other utilities depend on; therefore not only does attract normal hackers with less harmful intentions, but also terrorists might want to disrupt the grid. When many individuals with
high motivations and rich resources target the system, the risk of finding and exploiting
the vulnerabilities and penetrating to the system increases.

2.4 Privacy

Customers’ concern about their privacy is one of the major obstacles public adoption of
smart grid. Unlike the traditional power grid where metering data is read monthly, in the
smart grid detailed energy usage data are collected through smart meters at much shorter
time intervals (about every 15 minutes or less). While these precise data are critical to
efficient electricity distribution, demand-side management, load management, etc. they
might reveal a great amount of valuable and intimate information about customers, rang-
ing from the energy usage patterns and the types of household devices and appliances, to
information about the number of individuals in a house and their specific activities [61].
A lot of research has been done on nonintrusive appliance load monitoring (NALM) tech-
niques. NALM techniques can identify the individual household appliances by comparing
the energy usage pattern to libraries of known usage patterns (signatures) of different ap-
pliances [62]. NALM algorithms can use the detailed data obtained from smart meters
to reveal intimate information about customers’ habits. Another challenging issue is the
ownership of the valuable collected data. Several entities can benefit from the data includ-
ing utility companies and appliance manufacturers. The ownership and accessibility of the
data should be identified clearly.

One solution to mitigate the privacy problems in the smart grid is to anonymize data
whenever possible, so that the meter readings cannot easily be associated to a specific
customer. Although some of the collected data should be attributable, for example for
billing purposes, most of the high-frequency readings, used for power generation and dis-
tribution network control, are not required to be attributable to an identified customer;
instead, knowing the sub-station or the neighborhood of the gathered data will suffice. Anonymization is a rich research topic and several algorithms for different applications have been proposed. Along with adopting the existing algorithms, innovative methods which better fit the characteristics and requirements of the smart grid are required. In [22], a method for anonymization of the smart meter data was proposed. In this approach the meter readings are divided into two categories, low frequency data and high frequency data. Each data group is transferred with a specific identifier (ID), low-frequency ID (LFID) and high-frequency ID (HFID). The former is attributable, while the latter is anonymous. HFID is hard-coded inside the smart meter and known only by the manufacturer and a trusted third party responsible for ID verification of data at the aggregators. The disadvantage of this method is that the HFID is a fixed value; considering that the low frequency and high frequency data are correlated, it is possible to correlate the HFID and LFID pairs.

Appropriate system design is another way to reduce the privacy invasion risks. For instance, if optimization of energy consumption is performed locally through intelligent appliances and customer Energy Management System (EMS), the necessary granularity of data transfer will be reduced [2]. In [63], a local power management system model was proposed. This model uses a rechargeable battery to moderate the exposure of load signature. The idea is to resist the changes in power load so that the metered load stays constant. The battery discharges and recharges when the energy usage increases and decreases respectively; therefore, the cumulative energy usage stays almost the same over time. Although this method provides a high level of protection, it involves some practical challenges. One significant problem with the proposed method is that the algorithm depends on the usage of a rechargeable battery, which might not be available in all HANs in the future smart grid. In addition, applying this method to provide privacy might conflict with power usage
efficiency, which is one of the main goals of the smart grid. For instance, maintaining the constant load might force the battery to charge during the peak usage period.

2.5 Cryptography

Smart grid collects data from a large number of devices, such as smart meters and smart appliances. The collected data are used to manage the demand-response and integrate the distributed electrical energy resources. The data are normally transferred through wireless links which are not secure in nature. Therefore, strong security measures must be in place to protect the critical communication assets. Encryption and authentication - although not enough - can play a significant role in improving the integrity and confidentiality of the data. Constraining issues that should be considered in designing cryptographic algorithms for the smart grid include:

- Some smart grid devices, such as home appliances and residential meters, are employed in very large scales. In order to be cost effective, they have limited computational power, memory and storage. Although it is expected that most future smart grid devices support basic cryptographic capabilities, in designing the cryptographic algorithms these constrains must be considered.

- Some communication channels in the smart grid are designed to transmit short messages and therefore, have low bandwidths. Integrity protection mechanisms such as Cipher-Based Message Authentication Code (CMAC) [64] add 64 to 96 bits to every message [2]. This leads to a high overhead and might cause latency which is not affordable in many applications in the smart grid. Bandwidth limitations should also be considered in designing authentication algorithms for applications in which data are transferred with high frequency, such as wide area protection.
Chapter 2. Security Issues in Smart Grids

- Unlike many other IT systems where confidentiality and integrity are the most important cyber security properties, in many parts of the smart grid availability gets the highest priority. It is more important to make sure power is available rather than making sure the data regarding power flow are confidential. Therefore, in designing security measures such as authentication schemes, it should be considered that degrading the availability is not an acceptable cost.

- Smart grid equipment have much longer (about 20 years) life compared to typical IT systems, which might outlast the cryptography algorithms lifetimes. Test and replacement of these devices are often longer and more costly, due to the large-scale and availability requirements. In designing the cryptographic modules for such devices, evolvability and upgradability must be considered to enable future changes. Also using cryptographic schemes that exceed current security requirements is encouraged to postpone the possible need for future upgrades.

- Real time operation is necessary for many smart grid deployments. Cryptography and authentication algorithms for such systems require minimum computational cost and packet buffering.

Beside several existing standard encryption algorithms and authentication schemes that can be adopted in different parts of the smart grid, research on new methods that best meet the unique characteristics of the smart grid systems is required.

Identity-based encryption (IBE) is a new public key cryptographic system in which the public key is generated based on pre-agreed information bound to the user identity. The sender can encrypt the message with the receiver’s public key, without communicating to any trusted third party. Instead, the receiver obtains its private key from a key-generating server (KGS) only if it wants to decrypt the message. Once the receiver obtained its private key, it might not need to communicate to the KGS as long as the private key validity
Chapter 2. Security Issues in Smart Grids

 period is not expired. Some advantages of IBE over standard public key infrastructure (PKI) include simple system setup, reduced key exchange data traffic, and asymmetric key management requirements. These qualities are beneficial in many parts of the smart grid. For instance, in HAN the smart appliances’ sensors might be subject to energy constrains which is not the case for smart meters. Using IBE can off-load the key message traffic from the sensors to the smart meter with less resource constrains. However, in IBE scheme the KGS should be absolutely trustable which might be impractical in large deployments. In such scenarios hybrid approaches which combine PKI and IBE are preferable.

So et al. [65] have proposed an identity-based signcryption (IBS) scheme for end-to-end data encryption, authentication, and message integrity in the AMI communication network. Their scheme includes two phases: registration phase and data transmission phase. In registration phase, when a device wants to decrypt a received message or sign a message to be transferred, registers with a KGS to obtain its private key. The KGS holds the master key of the system and generates the private key based on this master key. Once the device obtained its private key, it can communicate with other devices in the network without contacting the KGS again. In data transmission phase, when device A wants to transfer a message to device B, it calculates the public key of B, using B’s unique ID. Then A encrypts the packet using an advanced encryption standard (AES) block chipper with a unique key generated based on B’s public key. Upon receiving the encrypted packet, B decrypts the packet with its own private key and verifies the content of the packet using A’s public key. So et al. [65] have used Tate pairing [66] on elliptic curve cryptosystem to generate a shared secret key pair between the message sender and receiver. This shared secret is used for authentication and encryption purposes. They have further improved the throughput of their signcryption scheme, by proposing a key caching mechanism to reduce the number of Tate pairing calculations. Simulations in [66]
confirm that the processing time and computational overhead of the proposed method meets the smart grid requirements. One advantage of elliptic curve is the adjustable level of security that this method can provide. More level of security needs more calculation time and therefore more delay. One can make a trade-off between security level and time delay, by adjusting the algorithm parameters, according to the specific requirements of different applications in the smart grid. Some other advantages of the proposed scheme include simple key management mechanism and independency from pre-device software configuration by users. More research, however, is required to evaluate the robustness of such an algorithm against different cyber attacks, as well as research on key revocation and key update mechanisms.

In-network aggregation methods can be applied to improve the efficiency of many-to-one communications in the smart grid. In these methods the computational load of data aggregation is distributed among the intermediate nodes of the network. In doing so, the intermediate nodes need to perform computational operations on the data packets in transit, which is impossible if the data is encrypted by traditional cryptographic algorithms. Homomorphic encryption schemes, on the other hand, allow some mathematical computations on the cipher text.

In [67], Li et al. presented a distributed data aggregation method. In their approach, data aggregation is performed through the smart meters participating in routing the data from the source smart meter to the data collector. Each intermediate smart meter adds its own data to the received data packet coming from the previous smart meter in a tree structure. The aggregator then receives the accumulated data of each tree branch. Authors of [67] applied Pillier homomorphic encryption scheme [68], so that the middle nodes can add their data to the encrypted data of previous smart meters without accessing the plain text information. Compared to traditional aggregation methods, the proposed scheme reduces
the computational load of data aggregation, while it slightly increases the computational load of the smart meters. In traditional approaches, each smart meter encrypts its message with its public key, and then the collector has to decrypt all the received messages from all smart meters separately. In [67], on the other hand, in addition to encrypting messages with homomorphic encryption, each smart meter has to do some mathematical operation to aggregate the messages. Instead, the collector only needs one asymmetric decryption for all of the smart meters in its tree. Considering the significant decrease of the required computation in the aggregator, computational overhead of the smart meters is acceptable. This scheme, however, is not resistant against false data injection. Having the cipher and public key, an illegitimate smart meter can falsify the aggregation result by injecting false cipher which is decryptable to a meaningful data. Research on mechanisms to detect malicious manipulation of aggregated data, as well as homomorphic encryption algorithms that match the requirements of in-network aggregation in the smart grid is required.

Applications and requirements of multicast authentication in the smart grid was investigated in [69]. Authors of [69] proposed a new one-time signature scheme with lower storage cost and smaller signature size compared to the existing methods, at the cost of increasing the computational requirements for signature generation and verification. However, the increased computation effect has been mitigated by flexible allocation of computations between sender and receiver regarding their computing capabilities. Multicast authentication has different requirements in different points of the smart grid. While bandwidth is important in wide area protection, due to high frequency of data transmission, it might not be a crucial factor in HANs. Instead storage limitations might be a concerning issue for home appliances. The scheme in [69] is particularly appropriate for applications with high multicast frequency and small message size such as demand-response and wide area protection.
Fouda et al. [70] proposed a light-weight message authentication algorithm for smart grid authentication. They adopted Deffi-Helman key establishment protocol [71] in their proposed authentication method. Also integrity of messages was assured by applying a Hash-based message authentication code. Fouda et al. [70] compared the performance of their scheme to elliptic curve digital signature algorithm (ECDSA) through computer simulations. ECDSA is considered to be a secure protocol for the smart grid demand response communication. They concluded that their approach has higher scalability, less decryption/verification delay, and less memory usage. Low latency and few message exchanges for authentication, makes this algorithm suitable for mutual authentication among smart meters. More research is required to evaluate the resistant of the proposed mechanism against passive and active cyber attacks.

### 2.6 Key Management

Another challenging issue in utilizing cryptographic algorithms for the smart grid is key management. Smart grid deployments, such as distributed automation, AMI, distributed generation, substations, include a large number of devices with many resource constrains. These devices are responsible for providing remote sensing and remote control. Research on economical, large-scale and flexible key management schemes that require less centralization and persistent connectivity, on a scale of possibly millions of keys and credentials is needed [2].

A key management scheme tailored for the smart grid deployments was proposed by Wu et al. [72]. They presented an authentication scheme based on elliptic curve public key cryptography [73] and Needham Schrouder symmetric key authentication protocol [74]. The public key was used to securely establish the symmetric keys of the agents. Then the on-the-fly generated symmetric key is employed in the authentication protocol.
The authors investigated the robustness of their proposed mechanism against a variety of security attacks and concluded that their scheme is resistant against replay and man-in-the-middle attacks. Simplicity and scalability are some benefits of their proposed approach. Though, no experimental proof regarding the performance of the algorithm, especially in large scale was provided.

Group key management for advanced distribution automation system issues was investigated by Sun et al. [75]. They, further, presented a novel key management scheme based on a decentralized three-tier hierarchal network model. Their scheme uses one-way key derivation algorithm in order to minimize the number of rekeying messages. Efficient key update and key storage, along with scalability are the advantages of their proposed method.

Authors of [76] discussed the application of public-key and symmetric-key schemes for key management in electricity transmission and distribution substations. They have investigated possible attacks and failure modes of these two schemes and compared their advantages and disadvantages [76].

2.7 Security Risk Assessment and Management

Risk management is the practice of identifying the critical assets, finding their vulnerabilities, assessing and prioritizing the risks, and establishing appropriate risk mitigation tactics. Considering the evolving nature of the smart grid vulnerabilities, appropriate risk management frameworks must be designed to improve the security and reliability of the power system against deliberate attacks, inadvertent errors, equipment failures, or natural disasters. Risk assessment is a critical step in the risk management process. In order to decide how much security is required in a specific point of the smart grid, a risk assessment methodology is needed. Risk assessment recognizes the system vulnerabilities and
the threats that might exploit these vulnerabilities, and estimates the extent of harm that these threats might cause. This enables the security personnel to prioritize the risks and accordingly implement more efficient and cost-effective security measures. Considering that the smart grid is a large-scale system which integrates physical and cyber systems, implementing a systematic risk assessment methodology is a challenging task.

One of the initial works in the context of cyber security assessment of the power grid was done by DOE’s infrastructure assurance outreach program [77]. A few works have been done with the focus of risk assessment and management methodologies in the context of the smart grid.

Ray et al. [78] investigated different approaches for security risk management of the smart grid. They believe that although the future smart grid entails both industrial control systems (ICS) and IT systems, which have different performance requirements and might be the target of different types of attacks, unified risk management approaches are required to enhance the security and reliability of the smart grid information exchange system. The authors propose threat and vulnerability modeling schemes to categorize the threats, analyze their impacts, and accordingly prioritize them. They use these schemes to provide quantitative methodologies for risk assessment. In quantitative risk assessment approaches, security metrics for assets, including systems and applications, are defined; then the threat risk which is the probability of exploiting a vulnerability to cause adverse effects is calculated by performing mathematical computations on the security metrics. Although quantitative approaches are accurate and consistent, quantification of attack probabilities, especially for a complicated system like the smart grid which lacks a large database of prior attacks is very difficult.

Kundur et al. [79] presented a framework for cyber security attack impact analysis of the electric smart grid. They have used directed graphs to provide a unified model for
both cyber and physical grid elements relationship. The graph nodes represent the grid elements, while the graph edges show the relationship among different grid components. The state information of each of the graph nodes is modeled by a dynamic system equation, which shows the physics of the interaction of electrical grid elements and the functionality of the cyber grid components.

In [80] two methods for cyber security vulnerability assessment of power industry was proposed. One method is a probabilistic assessment in which a vulnerability index is assigned to each of the cyber systems of the power grid by the use of probabilistic methods. To achieve the above goal, the probability of occurrence of cyber security events, the probability of the accidents resulting from these security events, and the loss caused by these accidents are calculated. This information is used to calculate the vulnerability index of each of the cyber systems. The authors believe that this method is more appropriate for control systems of the power industry. However, identifying a probabilistic distribution for the cyber security events can be a challenging issue. Some works in this regard were done in [81] and [82]. The other method is an integrated risk assessment in which first the level of cyber security risk is categorized into five group, and then a risk matrix is set up. The risk matrix includes the percentage of cyber security risk relating to each category, the probability of accidents caused by the security events and the influence of the accidents on the power industry. The vulnerability index is calculated based on this information. The authors believe that this method is more appropriate for management information system. This work is valuable in that it provides two procedural methods for vulnerability assessment of the cyber systems in the power industry, which can be extended for use in the smart grid. However, no proof regarding the efficiency of these two methods was presented in [80].
2.8 Secure System Architecture

Intelligent electronic devices (IEDs) will be widely used in the future smart grid to participate in critical tasks such as tripping circuit breakers in case of anomalies, raising or lowering voltage levels, etc. IEDs are expected to be remotely controllable and updatable, which might expose them to a variety of cyber attacks, including illegal command execution by a spoofed remote device, or attacks by software and firmware updates. Therefore, improving the security of IEDs is of great importance. Considering the scale and resource limitations of such devices, the need for research into cost-effective and mass-producible architectures that are highly tamper resistant and remotely recoverable, as well as secure firmware/software upgrade methods was emphasized in [2].

The need for code-auditing capabilities of the AMI deployments, which can be performed through remote attestation, was emphasized by the UtiliSec AMI security task force [83]. According to [84] ”remote attestation is the process whereby a remote party can obtain certified measurements of parts of the state of a system”. LeMay et al. [84] have presented an architecture named cumulative attestation kernel (CAK). This architecture can provide secure remote firmware auditing capability for networked embedded systems. In their approach an unbroken sequence of applications firmware revisions, is recorded in the kernel memory as audit logs. During the attestation process, a signed version of this audit log is provided for the verifier. In designing such architecture they considered the limitations of the embedded systems that are typically applied in AMI. Furthermore, they developed a prototype CAK that satisfies the requirements of advanced electric meters, and they investigated the security and fault-tolerance properties of their proposed architecture. However, in the threat model for this architecture the authors excluded extraordinary environmental phenomena, such as cosmetic radiation. Moreover, they assumed that physical attacks on microcontrollers, for instance probing, silicon modification, and fault analysis
are blocked by independent tamper-resistance techniques built into the device’s package. Although this prototype has been designed for smart meters, it is also extensible for use in IEDs.

2.9 Summary

In this chapter, we have focused on security aspects of the smart grid. We have described some security challenges that are introduced due to the special characteristics, architecture, constrains, and goals of the smart grid. We have investigated existing solutions to address these security issues and reviewed some works that have been done in this area over the last few years. Moreover, we have provided directions for further research. Although there are still some uncertainties regarding the architecture and function of the smart grid systems, it is crucial to design security solutions before widespread deployment. In addition to adoption of traditional security measures, research on new methods that meet the especial requirements of the smart grid is needed.
Chapter 3

Spoofing Detection and Prevention System for ZigBee-Based Home Area Networks

3.1 Introduction

In this chapter we focus on designing a novel spoofing detection and prevention system for ZigBee-based HANs. Identity spoofing is an important class of attacks which by exploiting the openness property of transmission medium is launched more easily in wireless networks compared to wired ones. In a spoofing attack, an adversary masquerades as one or more legitimate nodes to inject malicious traffic on their behalf. Spoofing is a basis for several other attacks, including various types of denial of service (DoS), session hijacking, etc. Therefore, designing appropriate spoofing detection and prevention mechanisms is of crucial importance and an open research topic.

Most existing systems rely on cryptographic methods to prevent spoofing attacks. However, the long history of breaking the authentication and encryption mechanisms employed in wireless networks shows the inadequateness of such approaches in guaranteeing a spoof-free network [85]. Furthermore, a variety of DoS attacks work in layers 1 and 2 of the network protocol stack, while encryption usually covers the upper layers. In addition, re-
source limitations of wireless sensors, hinder the implementation of strong cryptographic schemes.

To determine whether an identity belongs to a legitimate entity or has been counterfeited by a malicious node, forge-resistant qualities are utilized. One commonly used characteristic is sequence number of data-link layer frames [86]. Sequence numbers are linear chains of numbers assigned to frames by network cards. It is assumed that since sequence numbers are allotted by network cards, attackers cannot create a stream of packets that match the sequence number of the legitimate traffic. Therefore, the gap between sequence numbers could be employed to detect the presence of sybil nodes. However, nowadays myriad of free packet generator tools exist which enable the attackers to manipulate the desired fields of every frame.

Another property that has recently attracted the attention of researchers is RSS. According to the laws of physics, signal strength at a receiver antenna is proportional to the spatial distance between the receiver and the sender. Assuming that the sybil and legitimate nodes are located in different places, the RSS spatial correlation can be used to discriminate the entities applying the same identity. Beside distance, RSS depends on wireless environment features, such as absorption and multipath effect, which makes it hard to predict the power level of frames collected by a given receiver. Thus, sybil nodes cannot simply adjust their power levels to match the RSS of the legitimate nodes. Considering that RSS is a physical property that is hard to forge and is highly correlated to the transmitter’s location, using RSS for spoofing detection has the potential to provide a much better performance compared to when a nonphysical parameter like sequence number is used which is much easier to forge. This is why using RSS for spoofing detection has recently attracted a lot of attention and we have used it in our work. On the other hand, RSS is a random variable with Gaussian distribution [19]. Variance of RSS values confines the resolution of
Chapter 3. Spoofing Detection and Prevention System for ZigBee-Based Home Area Networks

RSS-based detection methods. To address this problem, multiple air monitors (AMs) are employed [19],[20],[21],[22]. AMs are responsible for listening to the network traffic and analyzing the RSS values of received frames. Increasing the number of AMs facilitates finer differentiation between entities located in closer distances and/or have close RSS values. The downsides of using multiple AMs are the extra cost required for excessive devices, as well as secure and reliable connections between several AMs and a central server. Moreover, relying on multiple AMs complicates the development of preventive measures. In this chapter, we introduce a novel RSS-based algorithm which provides a high detection accuracy even with a single AM. Unlike most existing solutions that directly process the RSS values of the packet stream, we employ feature extraction techniques to reduce data redundancy, and obtain a better representation of data. We extract two features of RSS streams, ratio of out-of-bound frames and the summation of detailed coefficients (SDC) in discrete Haar wavelet transform (DHWT) of the RSS streams. The former is effective in detecting spoofing attacks where sufficiently far distances exist between the attacker and the genuine nodes, or when the traffic rate of malicious node significantly outweighs that of the genuine node. SDC is useful when the legitimate and malicious nodes are located in close proximity or have close RSS mean values. Moreover, we suggest adaptive learning of RSS mean values, which reduces the false positives imposed by environmental changes. In addition to spoofing detection algorithm, we introduce two algorithms for automatically distinguishing and filtering malicious packets once a spoofing attack is detected. The contributions of this chapter are summarized as follows:

- We survey existing RSS-based spoofing detection mechanisms, and discuss their weaknesses.
- We design a novel robust RSS-based spoofing detection mechanism with low computational and resource overhead. While existing methods rely on multiple AMs for
accurate attack detection, the proposed approach provides a high detection performance with a single AM, and a superior performance over other existing methods using multiple AMs.

• For the first time, we introduce two methods for spoofing prevention using RSS values, static threshold and dynamic threshold. The former has very low computational requirements, yet due to high FPR introduces some network overhead. The latter needs more computations. However, it has a higher accuracy and a very low network overhead.

• Extensive performance analysis and experiments prove that the proposed algorithms can detect and stop spoofing attacks with high accuracy and low overhead.

The remainder of this chapter is organized as follows: In Section 3.2 we survey the existing RSS based spoofing detection methods. HAN Architecture is described in Section 3.3, and threat model is provided in Section 3.4. Section 3.5 and 3.6 explain the proposed spoofing detection and prevention algorithms orderly. In Section 3.7 the performance of our approach is analyzed. Experiment results are described in Section 3.8. Section 3.9 includes a discussion on the proposed method and comparison with previous approaches. Finally Section 3.10 summarizes the chapter.

### 3.2 A Survey on RSS Based Spoofing Detection Methods

In [19] a method for detecting spoofing attacks in wireless networks based on signalprints was proposed. Signalprint was defined as a vector containing RSS readings in multiple AMs. Signalprints of the traffic generated by a single node are expected to be similar. Dissimilar
samples suggest the presence of an attacker. As a dissimilarity measure, number of vector elements differing from a mean value more than a predefined threshold was counted. The threshold value was defined based on the variance of RSS values. When the out-of-bound elements of a vector exceeded a specific number, an attack alert was raised. For an IEEE 802.11 testbed and 6 AMs, the authors reported 95% DR without mentioning the rate of false positives. This approach requires a high number of AMs to achieve a desirable performance. The authors did not provide any updating mechanism for mean values of the RSS stream in AMs, which may cause a high FPR in the long term due to environmental changes.

Spoofing detection in IEEE 802.11 transmitters with antenna diversity was targeted in [21]. The authors showed that as a result of antenna diversity, the RSS distribution function tends to a multi-Gaussian model, instead of the single Gaussian assumed in other literature. They further showed that the difference between the mean RSS values of the traffic generated with different antennas of the same node is more than 5dB (5dB is the variance of the RSS Gaussian model used in other literature, which is an important factor in defining threshold values for classifiers.). For each AM, an RSS profile was built; then for a sequence of RSS samples, likelihood-ratio test was performed to detect deviations from the AM profiles. One or two times updating of RSS profiles per day was suggested to deal with the effect of environmental changes on distribution function. In an IEEE 802.11 testbed in an office building, using a single AM they achieved 73.4% DR with 3% FPR. For the same FPR, by increasing the number of AMs to 20, detection rate improved to 97.8%. This work is valuable in that it is the only method effective for multi-antenna transmitters. However, this approach is not efficient for single antenna, since it requires a high number of AMs, high computation and resources. Besides, one or two time profile updates might not be adequate to avoid false positives.
In [22] a technique for detecting spoofing attacks and localizing the position of adversaries was introduced. The authors used k-means clustering [22] for attack detection. For each frame, an \( n \)-dimensional vector of RSS readings in \( n \) different AMs was defined. Then, utilizing k-means algorithm, \( m \) vectors corresponding to a stream of \( m \) frames were divided into two clusters. Assuming a Gaussian distribution with 5dB standard deviation, a threshold was defined for the distance between the centers of the clusters under normal condition. When the distance exceeded the threshold value, a spoofing alert was raised. Performance of the method was tested in both IEEE 802.11 and IEEE 802.15.4 network testbeds, each with four AMs. For 10% FPR, [22] achieved 95% DR. Moreover, authors of [22] studied effect of the distance between the spoofing and original nodes on detection performance, and concluded that the further away is the spoofer from the original node, the higher is the detection rate. For IEEE 802.11, the detection rate was reported to be more than 90% when the distance is about 13 feet, while for IEEE 802.15.4 the same detection rate was obtained for distances about 20 feet.

The most recent work in the area of RSS-based spoofing detection is [20] which presents methods for spoofing detection, finding the number of attackers, and locating multiple adversaries. For detection phase, they used partitioning around mediod (PAM) algorithm. PAM clustering is similar to k-means, yet it is more robust against noise and outliers. For discovering the number of attackers, two methods were suggested, Silhouette plot and SILENCE. Both methods were based on finding the number of clusters in a clustering problem, where each cluster contains samples of a same distribution. This approach is effective, as long as the adversary node does not change its transmission power. A single attacker utilizing different power levels, can present multiple clusters. Performance of this method was assessed in IEEE 802.11 and IEEE 802.15.4 testbeds, with 5 and 4 AMs, respectively. For 5% FPR, DR was above 90%, when the distances between the malicious
and genuine nodes are less than 15 feet and 20 feet for IEEE 802.11 and IEEE 802.15.4 networks, respectively.

The major drawback of clustering-based approaches such as k-means and PAM is that when the ratio of malicious traffic significantly outweighs the benign traffic, benign frames are treated by the clustering algorithm as outliers. In this case malicious traffic is divided into two clusters; since both clusters belong to the same origin, the attack will not be detected. Therefore, clustering-based methods can not detect high traffic rate spoofing attacks which include most denial of service (DoS) attacks, such as back-off manipulation attack. In addition, the attacker and the genuine nodes do not necessarily communicate with the victim at the same time. Attack can happen when the genuine node is silent or have a very low traffic rate.

The only work that by converting the time series of RSS values into frequency domain, tries to provide a more proper representation of data is [23]. In [23], signal strength Fourier analysis (SSFA) was utilized. The intuition behind this method was that under normal condition only low-frequency oscillations exist. On the other hand, during spoofing attacks, the genuine frames are interleaved with malicious frames which generate high-frequency components. Using fast Fourier transform (FFT), the energy of high-frequency components were compared to a threshold. Passing the threshold value was interpreted as a spoofing attack. The advantage of this method is that using a single AM it can achieve a better detection performance compared to other methods. However, when the traffic rate of the original and spoofing nodes surpasses a specific range, this method will not be effective. Besides, relying on high frequency component of the Fourier transform introduced 0.2 second delay in attack detection.

The goal of this chapter is to provide a resource and time efficient algorithm to detect a vast range of spoofing attacks. While most of the previous works, [19],[20],[21] and [22],
Chapter 3. Spoofing Detection and Prevention System for ZigBee-Based Home Area Networks

tried to improve the detection performance by applying different classification techniques
on the raw RSS values, and achieved almost similar results, we focus on providing a thor-
ough and more distinctive representation of the data. We show that projecting the data
into a feature space that includes both magnitude and frequency related components, can
overcome the limitations of [19],[20],[21] and [22], and allows a high detection performance
even with one AM.

Our work is motivated by the proposed method in [23] in leveraging the fluctuations in
RSS stream for attack detection. Yet, we take a different approach; instead of FFT and
energy of high frequency component, we employ DHWT which is more time and resource
efficient. Accordingly we introduce SDC parameter which is highly separable for benign
and malicious traffic. In [23] only high frequency components are used for attack detection,
which not only imposes a high detection delay, but also is ineffective in detecting high rate
attacks. In addition to the fact that DHWT is faster than FFT, by avoiding the reliance
on frequency feature for highly separable attacks (high differences between RSS values or
high attack ratio), we further improve the detection delay.

Another advantage of the proposed method compared to [19],[20],[21] and [22] is ro-
 bustness against environmental changes achieved through adaptive learning of legitimate
RSS values. Overall, as we show in the rest of the chapter, the proposed algorithm provides
a significantly better performance in terms of resources and detection accuracy compared
to previous works.

3.3 Home Area Network Architecture

A ZigBee HAN primarily contains two types of devices, a smart meter which connects
a HAN to a NAN, and home devices inside the HAN. Three different HAN topologies
were defined in SEP 1.X [87] specification. In a utility enabled HAN (UE-HAN), all HAN
devices are under the control of the utility. In this architecture the smart meter acts as ZigBee MAC layer coordinator, trust center, and energy service interface (ESI). Another topology is consumer private HAN (CP-HAN) where an application layer gateway (ALG) connects the smart meter to home devices which are completely governed by the customer. The smart meter passes the usage data and pricing information to the ALG which plays the role of the network coordinator and also can be the security center. The third topology is a hybrid of UE-HAN and CP-HAN, where some devices are directly controlled by the utility, while others are managed by the customer. Typical devices in a CP-HAN are shown in Figure 3.1.

### 3.4 Threat Model

In a spoofing attack, at least three entities are involved, a legitimate node, an attacker and a victim. The legitimate node is allowed to exchange information and command massages with the victim. The victim uses the identity of the legitimate node to decide whether packets are coming from a genuine node or not. For instance, in IEEE 802.15.4 networks node identities are defined by node ID. The attacker eavesdrops network traffic to extract
identity of the legitimate node; then, by forging its identity, sends malicious traffic to
the victim on its behalf. Using a spoofing attack, as the basis for other attack types,
an adversary can disturb normal operation of the network, gain control of home devices,
remove some devices from the HAN, and affect the load curve. The adversary goal can
differ from simply unsettling a neighbor, to more malicious intentions such as inducing an
inefficient usage pattern, which can cause financial loss and damage to the customer and
the utility.

Through spoofing, an adversary can directly send control commands such as turn on/off
to the smart appliances by forging the identity of authoritative nodes like the energy
management system (EMS). Adversary might also send false data to the governing nodes
to indirectly control network nodes. For instance, by masquerading the identity of the
smart meter adversary can send false price and usage data to the EMS and customer
display, in return these devices control the HAN as the attacker wants.

In order to do a spoofing attack, the attacker node needs to be within the network
communication range to be able to exchange messages with the network nodes. The at-
tacker might also need to know the authentication credentials and encryption keys where
encryption is supported. This is especially true for upper layer attacks.

3.5 Spoofing Detection Algorithm

Our proposed spoofing detection and prevention system constitutes two modules. A spoof-
ing detection part which is installed on the security center and is responsible for detecting
spoofing attempts, as well as prevention agents installed on each sensor node to filter
malicious traffic. Both modules work based on the RSS values of received frames. The de-
tection system keeps track of RSS values of all network nodes, and analyzes them for signs
of spoofing attempts. Once an attack is detected, the detection module sends a secured
message to the victim, informing it about the presence and identity of an attacker. The intrusion prevention agent then distinguishes and filters malicious frames. The detection algorithm is as follows.

Spoofing detection can be formulated as a statistical significance testing problem. The null hypothesis is defined as: $H_0: \text{benign traffic (no attack)}$

Test statistics are then used to decide if the observed data belongs to the null hypothesis. In order to achieve high detection performance in terms of number of AMs, false positive/negative and detection rate, we utilize two parameters to represent the features of the stream of RSS values. We use the ratio of out-of-bound frames, which deals with the magnitude of RSS values. Further, we apply DHWT on time series of RSS values and use SDC to measure the oscillations in the data stream. In presence of a spoofing attack, RSS time series have more fluctuations since the legitimate packets are interleaved by forged packets with possibly different RSS values. In Section 3.7 we show that under a variety of attack scenarios, SDC provides a more separable distribution function (compared to RSS) which allows an accurate attack detection even when the genuine and attacker nodes are in close proximity.

### 3.5.1 Operation Phase

The data stream is divided into windows containing $2^n$ frames. Selection of the window size is related to the required number of samples as inputs of a DHWT. Following [19], [20], [21] and [22] we assume a Gaussian distribution for RSS. In these papers the RSS distribution was examined in indoor environments with areas about 2000 $ft^2$, which is similar to the physical environment that we target in this work.

**Step 1:** For each captured frame, the RSS is compared with the mean value, $\mu_{global}$, of the Gaussian distribution. A counter keeps track of the number of RSS values differing
from the $\mu_{global}$ more than a threshold, $th_{RSS}$. $th_{RSS}$ is related to the variance ($\sigma$) of the Gaussian distribution. At the end of each window, the ratio of out-of-bound frames is calculated, $R = \frac{n_{out}}{n}$, where $n_{out}$ is the number of out-of-bound frames. If $R$ lies in the range $R_{min} < R < R_{max}$, the algorithm stops at this step and raises an alarm declaring the presence of a spoofing attack. $R$ greater than $R_{min}$ shows that the number of frames having an out-of-bound RSS value is more than normal; this with a high probability is due to the presence of another entity caliming the same identity. However, $R_{max} < R$ might be the result of alteration of the mean value of the RSS distribution due to environmental changes rather than a malicious activity. If an attack is not detected at this step, the algorithm continues in Step 2.

**Step 2:** Next step is evaluation of the SDC. SDC is calculated using DHWT; in Section 3.7 we will briefly introduce DHWT and explain the reason behind using this transformation for feature extraction. Like RSS, SDC has a Gaussian distribution. Knowing the mean value and variance of SDC for a given node under normal condition, for each window SDC is compared with a threshold, $th_{SDC}$; if the threshold is exceeded, an attack alert is triggered. In addition to SDC, DHWT calculates the mean value of the frames inside the window, $\mu_w$.

**Step 3:** As the final step, if $R_{max} < R$ and step 2 did not detect an attack, an extra check is performed. $R_{max} < R$ can be the result of two conditions, shift of the mean value due to the environmental changes, or presence of an attacker with a much higher traffic rate compared to the genuine node. Therefore, in order to decide whether an attack alarm should be raised, or the mean value requires update, detection system sends a sequence of packets to the receiver, for instance a stream of data packets that require acknowledgment or internet control message protocol (ICMP) massages such as PING, if ICMP is supported. Using this method, the detection system forces the genuine node to transmit packets with
Chapter 3. Spooﬁng Detection and Prevention System for ZigBee-Based Home Area Networks

the traffic rate comparable to the traffic rate of the attacker. The detection system then can be sure that a sufﬁciently large portion of the received frames comes from the genuine node. After passing an expected time for receiving the replies, the algorithm is repeated from the ﬁrst step. This time if still \( R_{\text{max}} < R \) and an attack was not detected until this step, the algorithm decides that the traffic is benign and the mean value is updated by replacing \( \mu_{\text{global}} \) with \( \mu_w \).

3.5.2 Training Phase

To effectively detect spoofing attacks, the algorithm uses four threshold values, \( th_{\text{RSS}} \), \( th_{\text{SDC}} \), \( R_{\text{min}} \), \( R_{\text{max}} \). Except \( R_{\text{max}} \) which is a ﬁxed parameter close to one, other thresholds are learnt through a training phase in which the network is assumed to be spoof-free. First the mean, \( \mu_{\text{RSS}} \), and variance, \( \sigma_{\text{RSS}} \), of the RSS stream over the whole training duration are calculated. \( th_{\text{RSS}} \) is deﬁned as \( th_{\text{RSS}} = \sigma_{\text{RSS}} \).

In the next step, the RSS stream is divided into windows of \( 2^n \) frames. For each window, \( R \) (using \( \mu_{\text{RSS}} \) and \( \sigma_{\text{RSS}} \)) and \( SDC \) are calculated. According to the distribution of \( R \) and \( SDC \) over several windows, the following parameters are extracted:

- Mean value and variance of \( R \) (\( \mu_R \) and \( \sigma_R \)).
- Mean value and variance of \( SDC \) (\( \mu_{\text{SDC}} \) and \( \sigma_{\text{SDC}} \)).
- 10 largest values of \( R \) (\( R_{\text{MAX}} = \{R_{\text{MAX}1}, \ldots, R_{\text{MAX}10}\} \) in descending order)
- 10 largest values of \( SDC \) (\( SDC_{\text{MAX}} = \{SDC_{\text{MAX}1}, \ldots, SDC_{\text{MAX}10}\} \) in descending order)

Numerical values are assigned to \( R_{\text{min}} \) and \( th_{\text{SDC}} \) using the above parameters. \( R_{\text{min}} \) and \( th_{\text{SDC}} \) dictate the trade-off between DR and FPR. When very low FPR is required,
Chapter 3. Spoofing Detection and Prevention System for ZigBee-Based Home Area Networks

$R_{MAX}$ and $SDC_{MAX}$ are used (application of $R_{MAX1}$ and $SDC_{MAX1}$ results in close to 0% FPR). Otherwise, the threshold values can be defined according to the parameters of Gaussian distributions. One option is the combination in (3.1) which minimizes the FPR at the first step of the algorithm and provides a good detection rate for the second step; however, depending on the application and security policy, appropriate balance between DR and FPR are achievable by adjusting the thresholds:

$$
\begin{align*}
R_{min} & \in R_{MAX} \\
\text{th}_{SDC} &= \mu_{SDC} + 3\sigma_{SDC}
\end{align*}
$$

(3.1)

### 3.5.3 Multiple Air Monitors

When the detection system contains more than one AM, the RSS readings of the AMs are transferred to a central server (CS). All computations are performed in the CS and the AMs are responsible for reading the RSS values of the receiving frames from the target nodes, as well as sending packets to a given node for mean update. Based on the AM reports, CS makes a global decision about the health or malice of traffic. Using the RSS readings of different AMs, CS makes an RSS vector for each frame. Reports of different AMs might be received by CS with different time delays. CS assumes that reports within a predefined time interval (according to the delay estimation) belong to the same frame. If a report from a specific AM is not received by CS in the expected time, the mean RSS of that AM is used in the RSS vector. Detection algorithm for multi-AM scheme is summarized as follow:

- For each AM, the thresholds and algorithm parameters are learned through a training phase similar to what was described for single AM.

- During the operation phase, for each window, CS calculates the ratio of out-of-bound
frames. This time a frame is considered to be out-of-bound if (3.2) is true,

$$\|\overrightarrow{RSS} - \overrightarrow{\mu_{RSS}}\| > \|\sigma_{RSS}\|$$

where $\overrightarrow{RSS}$, $\overrightarrow{\mu_{RSS}}$ and $\overrightarrow{\sigma_{RSS}}$ are vectors with $n$ components, containing the RSS readings, mean and variance of RSS (learnt in the training phase) of $n$ AMs. If $R_{min} < R < R_{max}$, an attack is detected and the algorithm ends for the current window. $R_{min}$ is learnt in the training phase similar to the single AM method.

- For each AM, $SDC$ of the current window is calculated. An attack is detected if (3.3) is true.

$$\|\overrightarrow{SDC} - \overrightarrow{\mu_{SDC}}\| > 3\|\sigma_{RSS}\|$$

where $\overrightarrow{SDC}$, $\overrightarrow{\mu_{SDC}}$ and $\overrightarrow{\sigma_{SDC}}$ are vectors with $n$ components, containing the $SDC$, mean and variance of $SDC$ for $n$ AMs.

- If $R > R_{max}$ a notification is sent to the AMs. In response, each AM transmits a packet stream to the target node and the mean update mechanism is initiated.

Communication between AMs and CS can be wired or wireless; either way it must be highly secured, for instance through implementation of secure authentication and encryption algorithms.

### 3.6 Spoofing Prevention Algorithm

We propose two methods for differentiating and filtering malicious packets.
3.6.1 Static Threshold

Considering that RSS values follow a Gaussian distribution, the agent calculates the mean and variance of RSS values of received frames for each communicating node. These values are calculated in an online manner to reduce the size of required storage. When the node receives a mean update message from the security center, it discards the saved mean value and starts calculating it again. When an intrusion alarm is received, for each incoming frame the agent compares the RSS value with the mean RSS and if the difference exceeds a pre-defined threshold, the packet will be dropped. The threshold is a fixed value, selected according to the variance of RSS distribution. The main advantage of this method is its simplicity and low computation and storage requirements. The downside of this approach, however, is the network overhead due to dropping the legitimate packets. The threshold determines the trade off between DR and FPR. The smaller the threshold, the more is the chance of dropping illegitimate packets; at the same time it results in drop of more genuine packets. In Section 3.7 we will investigate this effect in more detail.

3.6.2 Dynamic Threshold

In this method the threshold value is assigned dynamically based on the distance between RSS distribution of the genuine and attacker nodes. In order to measure the distance, k-means clustering is employed. The number of operations in a one-dimensional two cluster k-means problem is in the order of $O(n^3 \log n)$. When the agent receives an attack alarm, a queue containing the RSS values of $n$ last frames is formed. The agent uses 2-cluster k-means algorithm, and RSS values are divided into two clusters. Then the center and distance between two clusters are calculated. If one of the calculated centers is close to the mean value of the genuine node, the measured distance is used to adjust the threshold; since in this case with a high probability one cluster belongs to the genuine frames and
the other to the attacker. However, the genuine and attacker nodes do not necessarily communicate with the victim simultaneously. To address this situation, if none of the measured mean values is close to the mean value of the genuine node, the average of the mean values of the two clusters is calculated and its difference with the mean value of the genuine node determines the distance.

The dynamic method allows a larger threshold for far RSS values which results in a lower FPR. In addition, when RSS values are close, a small threshold is used; therefore, attacks are still preventable, while in static approach in order to maintain the balance between prevention rate (PR) (which is the number of intrusion instances prevented by the system divided by the total number of intrusion instances) and FPR, attacks with close RSS values are not stoppable. Although dynamic threshold reduces network overhead as a result of the low FPR, this method requires more computations and introduces a short delay for queuing and clustering the frames; however, considering that these effects only happen during an attack, their costs are acceptable.

3.7 Performance Analysis

3.7.1 Performance of the Spoofing Detection Algorithm

In the proposed approach, we exploit the spatial correlation of RSS values to determine whether the incoming frames, carrying the same identity, belong to a single genuine node or are originated from different sources. The RSS of a frame measured in a given location is affected by parameters such as environmental condition, random noise and multipath effect; still it strongly depends on the distance between the sender and the receiver. As a result, RSS of devices located at different physical places are expected to be distinctive.
The spatial dependency of RSS is formulated as (3.4).

\[
RSS = P_0 - 10\gamma \log \left( \frac{d_1}{d_0} \right) + X
\]  

(3.4)

where \(P_0\) is the transmission power in a reference point, \(d_0\) and \(d_1\) are the distances from the sender to the reference point and to the receiver, respectively, \(\gamma\) is the path loss exponent, and \(X\) is the shadow fading with a Gaussian distribution \(N(0, \sigma)\). Having the configuration in Figure 3.2 and assuming an equal transmission power for nodes 1 and 2, the difference between RSS values of the two nodes, sensed by node 3 is:

\[
\Delta RSS = 10\gamma \log \left( \frac{d_1}{d_2} \right) + \Delta X
\]  

(3.5)

where \(\Delta X\) has a Gaussian distribution \(N(0, \sqrt{2}\sigma)\).

RSS values in a landmark also follow a Gaussian distribution \(N(\mu, \sigma)\); while \(\sigma\) depends on environmental condition, its average in an indoor location is reported to be about 5dB [19],[20],[21],[22]. Knowing the above physical properties, in the rest of this section we prove the efficiency of the proposed algorithm through mathematical analysis.

**Step 1:** As the first detection step, assuming the distribution function of \(N(\mu_g, \sigma_g)\) for
Chapter 3. Spooﬁng Detection and Prevention System for ZigBee-Based Home Area Networks

the genuine node, for each window the number of frames differing from the mean value, \( \mu_g \), more than \( \sigma_g \) is counted as \( n_{out} \). Then parameter \( R = \frac{n_{out}}{n} \) is compared with a threshold \( \tau \) (0 < \( \tau \) < 1). The value of \( R \) under normal condition is proportional to the area denoted by dots in Figure 3.2.b. Presence of an attacker increases the value of \( R \), since in this case \( n_{out} \) will be related to the summation of dotted and crossed areas. Smaller value of \( \tau \) results in a higher DR, yet it increases the FPR. We formulate the FPR as (3.6), which is the probability of \( R \) exceeding the threshold when there is no attacker.

\[
FPR = P(R > \tau|\text{normal}) = P(n_{out} > \tau n|\text{normal}) \tag{3.6}
\]

\[
P_{\text{normal}}(RSS) = N(\mu_g, \sigma_g) \tag{3.7}
\]

The probability that from \( n \) frames \( n_{out} \) are out of a boundary, follows a binomial distribution.

\[
P(n_{out}|\text{normal}) = \binom{n}{n_{out}} P_{out}^{n_{out}} (1 - P_{out})^{n - n_{out}} \tag{3.8}
\]

\[
P_{out} = P_{\text{normal}}(|RSS - \mu_g| > \sigma_g) = 2\Phi(\mu_g - \sigma_g; \mu_g, \sigma_g) \tag{3.9}
\]

where \( P_{out} \) is the probability of a frame having an out-of-bound RSS under normal condition. In (3.8), \( \Phi(.) \) is the cumulative distribution function (CDF) of the Gaussian distribution. Considering (3.6) and (3.8) the FPR is:

\[
FPR = \sum_{n_{out} = \tau n}^{n} P(n_{out}|\text{normal}) = 1 - F(\tau n; n, P_{out}) \tag{3.10}
\]

where \( F(.) \) is the CDF of the binomial distribution. As the above formula suggests, FPR
is inversely related to $\tau$. On the other hand DR is formulated as below:

$$DR = P(R > \tau|\text{attack}) = P(n_{out} > \tau n|\text{attack})$$  \hspace{1cm} (3.11)

Considering a Gaussian distribution for the attacker, $N(\mu_a, \sigma_a)$, PDF of the RSS values under attack is:

$$P_{\text{attack}}(RSS) = (1 - \eta)N(\mu_g, \sigma_g) + \eta N(\mu_a, \sigma_a)$$  \hspace{1cm} (3.12)

where $\eta$ is the ratio of the spoofed frames. Similar to the explanation for FPR, $n_{out}$ has a binomial PDF.

$$P(n_{out}|\text{attack}) = \binom{n}{n_{out}} (P'_{out})^{n_{out}}(1 - P'_{out})^{n - n_{out}}$$  \hspace{1cm} (3.13)

where $P'_{out}$ is the probability of a frame having an out-of-bound RSS under attack condition, considering (3.12):

$$P'_{out} = 2(1 - \eta)\Phi(\mu_g - \sigma_g; \mu_g, \sigma_g) + \eta Q(\mu_g + \sigma_g; \mu_a, \sigma_a)$$  \hspace{1cm} (3.14)

where $Q$ is the Q-function (the complement of the CDF) of the Gaussian distribution. Finally, DR is summarized as:

$$DR = 1 - F(\tau n; n, P'_{out})$$  \hspace{1cm} (3.15)

From the above equations and considering Figure 3.2.b, one can conclude that DR is a function of $\tau$, $\eta$, $|\mu_g - \mu_a|$, $\sigma_g$ and $\sigma_a$. The influence of $\eta$ and $|\mu_g - \mu_a|$ on detection performance, in terms of receiver operating characteristic (ROC), is depicted in Figures 3.3 and 3.4. $\tau$ defines the trade off between FPR and DR. Figure 3.3 shows that as the ratio of malicious frames increases the detection performance improves.
As expected, Figure 3.4 shows that the detection performance improves when the distance between mean values increases. On the other hand, enlargement of variance degrades the detection performance. In summary, the more separable are the distribution functions of the attacker and genuine node, the better is the detection performance.

In step two of the algorithm, by applying DHWT on RSS stream, we achieve a parameter with more separable distribution for a given RSS stream.

**Step 2:** In this step, for each window, DHWT is utilized to provide a measure of oscillations in RSS values. While fast Fourier transform (FFT) have been widely used to extract frequency components of time series, discrete wavelet transform (DWT) is proved to be a superior alternative in many applications [88]. DHWT has the desirable features of wavelet transform. It not only contains the frequency content of the input, but also shows the temporal order. Another advantage of DHWT is the low number of required operations, which makes it time and resource effective. Computing DHWT of $N$ points takes $O(N)$
Chapter 3. Spoofing Detection and Prevention System for ZigBee-Based Home Area Networks

Figure 3.4: Effect of the difference between RSS mean values on performance of the first step of the spoofing detection algorithm

![Graph showing the effect of RSS mean differences on DR](image)

Figure 3.5: Discrete wavelet transform decomposition algorithm

![Diagram of wavelet transform](image)

Arithmetic operations, which is much less than $O(N \log N)$ required for FFT. Resource and time efficiency are the major reasons why we employed DHWT in our detection algorithm. Figure 3.5 shows the decomposition process in a wavelet transform. In the figure, $g[n]$ and $h[n]$ are low-pass and high-pass filters which must be quadratic mirror. At each level, the input stream is decomposed into low and high frequencies. The outputs of low-pass and high-pass filters are called approximation coefficients and detail coefficients, respectively.
In summary, DHWT pairs up the input values, stores the differences, and passes the sums to the next level. The process is repeated until finally $2^n - 1$ differences and a mean value remain [89].

Let us assume that $m$ AMs are monitoring the RSS values of the frames with identity of a legitimate node. $\Delta RSS$, which is the total RSS deviation from the mean values in $m$ landmarks is calculated as:

$$\Delta RSS^2 = \sum_{i=1}^{m} (RSS_i - \mu_{gi})^2$$  \hspace{1cm} (3.16)

where $RSS_i$ is the value of RSS in landmark $i$, and $\mu_{gi}$ is the mean RSS of genuine node in landmark $i$. As it was shown in [19], when the two nodes are co-located (there is no attack), $X = \Delta RSS^2$ the random variable has a central Chi-square distribution $\chi^2(m)$, where $m$ is the degree of freedom which is equal to the number of AMs. On the other hand, when wireless nodes are at different locations, $X$ follows a non-central Chi-square distribution $\chi^2(m, \lambda)$, where $m$ is the degree of freedom and $\gamma$ is the non-centrality parameter, which in this case is:

$$\lambda = \sum_{i=1}^{m} \left( \frac{\mu_{gi} - \mu_{ai}}{\sigma} \right)^2$$  \hspace{1cm} (3.17)

In (3.17), $\mu_{gi}$ and $\mu_{ai}$ are the mean values of RSS stream of the genuine and attacker nodes in $i^{th}$ AM. The variance is assumed to be the same for both nodes, $\sigma$. Therefore, the DR and FPR are calculated using (3.18) and (3.19).

$$DR = P(x > \tau|\text{attack}) = 1 - F_{\chi^2(m, \frac{\lambda}{2\sigma^2})} \left( \frac{\tau}{2\sigma^2} \right)$$  \hspace{1cm} (3.18)

$$FPR = P(x > \tau|\text{normal}) = 1 - F_{\chi^2(m)} \left( \frac{\tau}{2\sigma^2} \right)$$  \hspace{1cm} (3.19)

where $F_{\chi(.)}$ is the CDF of $\chi$, and $\tau$ is a threshold. When $\tau$ is exceeded, a spoofing
attack is detected. While $\tau$ defines the trade-off between DR and FPR, DR is affected by $m$, $\sigma$, and $\lambda$. To achieve a higher DR, previous works increased the number of AMs ($m$). To further improve the detection performance, we suggest application of frequency components; instead of $X = \Delta RSS^2$, we define the random variable $X$ as $X = \Delta SDC^2$.

$$\Delta SDC^2 = \sum_{i=1}^{m} (SDC_i - \mu_{SDC_i})^2$$

(3.20)

where $\mu_{SDC_{gi}}$ is the mean of SDC of the genuine node in $i$th AM and,

$$SDC_i = \sum_{j=1}^{n-1} dc_i[j]$$

(3.21)

In (3.21), $n$ is the window size and $dc_i[j]$ is the $j$th detail coefficient, starting from the high frequencies, for the $i$th AM. Assume that the legitimate node sends a frame stream with RSS values $S_g = \{S_{g1}, ..., S_{gn/2}\}$, where $S_g \sim N(\mu_g, \sigma_g)$. At the same time the attacker sends the stream with RSS values of $S_a = \{S_{a1}, ..., S_{an/2}\}$, $S_a \sim N(\mu_a, \sigma_a)$. For simplicity of analysis we consider an ideal case when the ratio of malicious traffic is 0.5, and each pair of legitimate frames is interleaved by one malicious frame. Then the RSS stream in an AM is: $S_g = \{S_{g1}, S_{a1}, ..., S_{gn/2}, S_{an/2}\}$. By applying DHWT on $S$, level 1 detail coefficients are: $\left\{ \frac{S_{g1} - S_{a1}}{2}, \frac{S_{g2} - S_{a2}}{2}, ..., \frac{S_{gn} - S_{an}}{2} \right\}$ For simplicity we ignore higher level detail coefficients (Including higher level coefficients will have a positive effect on separability of $SDC$).

$SDC$ of the first level detail coefficients is $SDC_{S1} = \sum_{i=1}^{n/2} \frac{s_{gi} - s_{ai}}{2}$. Considering the summation property of Gaussian variables ($\sum_i a_i N(\mu_i, \sigma_i)) = N(\sum_i a_i \mu_i, \sqrt{\sum_i a_i}^2 \sigma^2$), and assuming the same variance for both attacker and genuine nodes

$$SDC_{S1} \sim N\left(\frac{n}{4}(\mu_g - \mu_a), \frac{\sqrt{n}}{2} \sigma\right), \text{ while } SDC_{S_{g1}} \sim N\left(0, \frac{\sqrt{n}}{2} \sigma\right) \text{. Therefore, } \Delta SDC \sim N\left(\frac{n}{4}(\mu_g - \mu_a), \frac{\sqrt{n}}{2} \sigma\right), \text{ while } \Delta RSS \sim N(\mu_g - \mu_a, \sigma)$$

Thus, for $n = 64$ as it can be calculated from (3.17), $\lambda$ of $\Delta SDC$ is 16 times the $\lambda$
Figure 3.6: Effect of increase in the number of AMs and non-centrality parameter ($n$ represents the number of AMs and $k$ is the scale of non-centrality) of $\Delta RSS$. However, we remind that this is for an ideal case, where benign frames are alternatively interleaved by malicious frames. Also the higher level coefficients are ignored. Therefore, the above computation provides an estimate of improvement of $\lambda$ rather than a deterministic value.

According to (3.18) and (3.19), Figure (3.6) compares the effect of increase in the number of AMs and non-centrality on detection performance. It can be seen in the figure that when the non-centrality parameter is scaled by 4, the detection performance of single AM outperforms the performance of a system with 12 AMs which has a fixed non-centrality. Other parameters of Figure 3.6 are: $\mu_g - \mu_a = 10$ and $\sigma = 5$. 
3.7.2 Performance of the Spoofing Prevention Algorithms

Static Threshold

For the static method, FPR and network overhead have almost fixed values depending on the value of the threshold. Therefore, the threshold is defined so that there is a balance between network overhead and PR. As the threshold becomes larger, the system loses its ability to prevent attacks in which the attacker is in close proximity of the genuine node or have close RSS values. False positives introduce network overhead, since when a genuine frame is dropped due to exceeding the RSS threshold, several retransmissions might be required until the RSS lies in the legitimate range. The expected number of tries until a successful transmission is formulated as:

\[ E = \frac{1}{P(|RSS - \mu_g| < \tau |genuine)} = \frac{1}{1 - 2\Phi (\mu_g - \tau; \mu_g, \sigma_g)} \] (3.22)

Figure 3.8 show the relationship between the expected number of retries, \( E \), and the threshold, while the relationship between \( E \) and FPR is presented in Figure 3.9.

Dynamic Threshold

In the dynamic approach, since the threshold is assigned dynamically, based on the difference between mean values, not only attacks with close distance to the genuine node are preventable, but also unnecessary network over head is avoided. These, however, come with the penalty of additional computation and processing delay.
3.8 Experiments

3.8.1 Testbed

In order to evaluate the performance of our spoofing detection approach, we conducted two experiments in an IEEE 802.15.4 network testbed. The network was established in a...
real office environment located in communication lab, in the Department of Electrical and Computer at the University of British Columbia. The office size is 9903 ft$^2$. The network contained four landmarks (AMs), an attacker and a genuine node. We used six telosB motes as network nodes; four were programmed to act as AMs to monitor the RSS of received frames. The AM motes were connected to four personal computer (PC) systems, and the RSS readings were directly transferred to the PCs. Two other motes which had the role of the attacker and the genuine node, were programmed to send constant bit rate (CBR) traffic with 5 frames per second. The position of the AMs and the genuine node are depicted in Figure 6 by arrows and a star sign respectively. During the course of experiments the attacker was placed in different locations depicted in Figure 3.10 by bold dots.

### 3.8.2 Performance of the Spoofing Detection Algorithm

#### Experiment 1

The goal of the first experiment was to study the changes in SDC under normal and attack conditions. We wanted to confirm that in practice SDC provides a more separable representation of the data compared to RSS.
In this experiment, first the RSS logs of genuine node frames were collected by AM3 for the duration of four hours. Then, the $SDCs$ of RSS streams were extracted, and their mean values and variances were calculated. In computing the $SDC$, we used the absolute value of $DCs$ to avoid cancelation of $DCs$ with different polarities. The window size was set to 64. In the next step, the attacker node was placed in 50 cm, 1 m, 3 m, 4 m and 5 m distance from the genuine node. For each position RSS log of genuine and attacker nodes were captured for 1 hour, and the mean value and variance of $SDC$ were calculated. The ratio of malicious and benign traffic was 0.5. The experiment results are presented in Table 3.1. As it can be seen from the table, $\Delta SDC$ provides a larger $\lambda$ in several orders of magnitude. In addition, when the distance between two nodes expands, $\lambda$ increases. Table 1 also includes the result of single AM spoofing detection using the proposed algorithm for each dataset. Even when the distance is as low as 50 cm, using SDC, attack detection is to some extent possible, while RSS based approach is completely incompetent due to the similar mean values. As the distance increases to 3 m, the detection performance improves to a satisfactory level. In previous works, on the other hand, even with multiple AMs, the
minimum detectable distance was reported to be about 6 m.

**Experiment 2**

In the next experiment we evaluated the performance of the proposed spoofing detection mechanism. In the testbed, the attacker node was placed in each position marked by a bold dot in Figure 3.10, for 5 minutes, and transmitted CBR traffic. During the whole experiment the benign node was located at the position depicted by a star in the figure, and sent CBR traffic with the same rate as the attacker. Overall, 90 different placements of attacker and benign nodes were tested. The 4 AMs, monitored the stream of RSS values from both the attacker and the genuine nodes, and stored the values in a log file. At the end of the experiment the log files were collected and the detection algorithm was applied for each AM separately. We used 4 AMs for a single AM detector and analyzed the results separately to study the effect of the position of AM on detection performance. The mean and variance of the RSS of the genuine node was calculated using all samples received during the experiment, which includes 135000 samples. We observed a variance of 4.68 dB for these samples. Therefore, the standard error mean of the genuine node calculated using the experiment samples is 0.0058dB. Considering that the granularity of the measured RSS values is 1dB which is much larger than this value, this error does not affect the experiment
Chapter 3. Spooﬁng Detection and Prevention System for ZigBee-Based Home Area Networks

results. The detection performance of each AM is depicted in Figure 3.11. As the ﬁgure suggests, AM1 has the best detection performance. The reason is the closer distance of AM1 to the genuine node. We remind that according to (3.5), attack detection depends on the ratio of distance between attacker and genuine nodes to the AM, rather than the distance between 2 nodes. Therefore, for a ﬁxed distance between the nodes, a closer AM has a better chance of attack detection.

While the experiment was conducted in an ofﬁce building with a usual amount of people movement, to study the effect of moving objects, we deliberately introduced more movements in close proximity of AM3. As Figure 3.11 shows, we observed worst but still acceptable performance for this AM compared to the other 3 AMs.

To further study the effectiveness of the magnitude and frequency features on detection process, Figure 3.11 also includes the ROC curve of the detection processes purely based on magnitude \((R)\) and frequency \((SDC)\) features. As it can be seen in Figure 3.11, SDC provides a better detection performance. Also we can see that combining both features signiﬁcantly improves the performance. The average ROC of the 4 AMs is shown in Figure 3.12. In addition, we studied the effect of the ratio of malicious and benign trafﬁc. The results are shown in Figure 8. When the rate of malicious and benign trafﬁc is close, the ﬂuctuations in RSS stream is high; therefore, \(SDC\) can effectively distinguish the malicious trafﬁc. However, when the ratio of malicious trafﬁc is very high, the RSS stream will have less ﬂuctuations, since the trafﬁc will mostly belong to one node (the attacker in this case). Yet step 1 of the algorithm will be very effective in this scenario. Therefore, by applying both features the spoofing detection mechanism can successfully detect a vast rang of malicious trafﬁc ratio. However, when the trafﬁc rate of the benign node is much higher than that of the malicious node, the algorithm will not be effective. Yet, such scenario can be less hazardous. At least most DoS attacks require high trafﬁc rate. Figure
Figure 3.11: ROC curve of the spoofing detection algorithms for a) AM1, b) AM2, c) AM3, d) AM4

3.13 represents the ROC curve of the spoofing detection in AM4 using $R$, $SDC$ and both features under various attack ratios.

3.8.3 Performance of the Spoofing Prevention Algorithms

Figure 3.14 shows experiment results for static threshold prevention method. In this experiment AM1-AM4 play the role of victims (V1-V4). We used 4 victims in our experiment in order to study the effect of position of the victim in prevention performance. As it can be observed from Figure 3.14, the agent on V1 has the best performance. The reason is the closer distance between V1 and the genuine node. We also studied the effect of mov-
ing objects on the performance. While the testbed was located in an office with a usual amount of people movement we deliberately introduced more movements around V3 which shows the worst, still acceptable performance compared for this victim node compared to other victim nodes. Figure 3.15 shows the average ROC of the static spoofing prevention algorithm for four victims. To test the dynamic method, its prevention performance was measured when the threshold was 1/2, 1/3, 1/4, and 1/5 of the distance between the mean values. The window size, meaning the number of frames that were clustered each time, was 64. The results for four victims are shown in Table 3.2. Comparing Figure 3.14 and Table 3.2 shows a significantly higher performance for dynamic threshold method. We emphasize that these results show the performance of an automatic prevention system. Unlike an IDS, here false positives do not cause operational cost or administrator negligence, but the cost is some network overhead and delay. Even for the worst case in Table 3.2 which is the 94.42% PR and 64.81% FPR, the most significant cost is 184% network overhead which only happens when an attack is detected.

Figure 3.12: Average ROC of the spoofing detection algorithm.
Table 3.2: Prevention performance for dynamic threshold spoofing prevention

<table>
<thead>
<tr>
<th>Threshold</th>
<th>1/2</th>
<th>1/3</th>
<th>1/4</th>
<th>1/5</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>PR(%)</td>
<td>FPR(%)</td>
<td>95.33, 1.52</td>
<td>97.53, 6.00</td>
</tr>
<tr>
<td>V2</td>
<td>PR(%)</td>
<td>FPR(%)</td>
<td>92.11, 12.05</td>
<td>95.11, 18.75</td>
</tr>
<tr>
<td>V3</td>
<td>PR(%)</td>
<td>FPR(%)</td>
<td>82.73, 23.57</td>
<td>89.45, 39.21</td>
</tr>
<tr>
<td>V4</td>
<td>PR(%)</td>
<td>FPR(%)</td>
<td>92.94, 5.02</td>
<td>95.57, 8.76</td>
</tr>
</tbody>
</table>

Table 3.3: Comparison of different RSS-based spoofing detection techniques (NA: not applicable, - : was not provided in the paper, *: average of the results of 4 AMs)

<table>
<thead>
<tr>
<th>Approach</th>
<th>Test area ($\text{ft}^2$)</th>
<th>Network type</th>
<th>DR 1AM (%)</th>
<th>FPR 1AM (%)</th>
<th>No. of AMs</th>
<th>DR (%)</th>
<th>FPR (%)</th>
<th>Min. distance ($\text{ft}^2$)</th>
<th>DR (%)</th>
<th>FPR (%)</th>
<th>Resistant to env. changes</th>
<th>Detects high rate attacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;SDC</td>
<td>9903</td>
<td>802.15.4</td>
<td>* 92.63</td>
<td>* 0.00</td>
<td>4</td>
<td>99</td>
<td>0.0</td>
<td>9.84</td>
<td>98.08</td>
<td>2.56</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>K-Means</td>
<td>1600</td>
<td>802.15.4</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>95.7</td>
<td>0.0</td>
<td>9.5</td>
<td>20.00</td>
<td>90.00</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>PAM [20]</td>
<td>1600</td>
<td>802.15.4</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>96.5</td>
<td>0.0</td>
<td>10.0</td>
<td>20.00</td>
<td>90.00</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>Fourier</td>
<td>-</td>
<td>802.11</td>
<td>0.05</td>
<td>NA</td>
<td>6</td>
<td>95.6</td>
<td>-</td>
<td>16.40</td>
<td>72.2</td>
<td>-</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Signalprint [19]</td>
<td>11625</td>
<td>802.11</td>
<td>NA</td>
<td>NA</td>
<td>6</td>
<td>95.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Multi-Gaussian [21]</td>
<td>16000</td>
<td>802.11</td>
<td>64.1</td>
<td>1.00</td>
<td>20</td>
<td>94.4</td>
<td>1.0</td>
<td>3.0</td>
<td>9.84</td>
<td>84.3</td>
<td>1</td>
<td>No</td>
</tr>
</tbody>
</table>

3.9 Discussions and Comparison

A summary of various RSS-based spoofing detection methods is provided in Table 3.3. The datasets used for the performance evaluation in the experimental section of the papers are not the same. Yet, similar approaches have been taken in setting up the testbeds. In some papers both IEEE 802.11 and IEEE 802.15.4 networks were evaluated. In this case we only included the IEEE 802.15.4 results since it is the focus of this work. If the experiments in the previous works were merely based on WiFi, we included the IEEE 802.11 results in Table 3.3. According to [19] and [21], the detection performance for IEEE 802.11 and IEEE 802.15.4 is close. Though, the minimum distance for detectable attacks in IEEE 802.15.4 is a little higher (about 5 feet) than IEEE 802.11.
3.10 Summary

In this chapter we have studied the existing RSS-based spoofing detection methods for static IEEE 802.15.4 networks and explained their limitations. In addition to long detection delay, ineffectiveness in mitigating high rate attacks and lack of robustness against environmental changes, most existing approaches rely on multiple AMs, which is not cost effective. Further, we have presented a novel spoofing detection technique which employs both magnitude and frequency features of RSS streams to provide a high detection performance even with a single AM. Evaluations of the proposed method through experiment and analysis have proved its high performance both for single and multi-AMs. In addition to introducing an efficient approach for spoofing detection, we have introduced two algorithms for automatic spoofing prevention using RSS values and investigated their performances through analysis and experiments which proved the effectiveness of our approach.
Figure 3.13: ROC of spoofing detection in AM4 based on a) R, b) SDC, c) proposed algorithm.)
Figure 3.14: Experimental ROC for static threshold spoofing prevention.

Figure 3.15: Average experimental ROC of V1, V2, V3 and V4 for static threshold spoofing prevention.
Chapter 4

HANIDPS, An Intrusion Detection and Prevention System for ZigBee-Based Home Area Networks

4.1 Introduction

The dominant HAN technology in North America and many other countries is ZigBee. Being located in insecure environment and use of wireless technology make HANs vulnerable to cyber attacks \[3\],\[4\]; this necessitates application of appropriate IDSs. At the same time, since HANs are located in areas far from the utility, receiving IDS alarms by the utility and acting upon them introduces a large operational cost and delay in stopping the attacks. Considering the large scale of smart grids, when human response is expected, a small percentage of false alarms results in a high operational cost. Therefore, in the context of HAN, intrusion prevention mechanisms which not only detect but also stop the attacks are highly preferable.

In this chapter we present a novel model-based intrusion detection and prevention system (IDPS) for ZigBee-based HANs, HANIDPS. We use SEP 2.0 application protocol specification as well as IEEE 802.15.4 standard to define a thorough feature space for HANIDPS. HANIDPS employs a model-based detection module and a machine learning-
Chapter 4. HANIDPS, An Intrusion Detection and Prevention System for ZigBee HANs

Based prevention module which dynamically learns the best strategy against an attacker. Through analysis and experiments we show that HANIDPS is able to detect various attack types with a high performance. The main contributions of this chapter are:

- To the best of our knowledge we are the first who address the problem of automatic intrusion prevention in ZigBee-based HANs. Considering that in HANIDPS the prevention operation is performed automatically, the costs of false positives are low and limited to some network overhead. Also the delay in stopping the attacks is significantly shortened compared to when human intervention is required. This reduces the damages caused by possible attacks.

- HANIDPS is a novel algorithm which utilizes a model-based IDS along with a dynamic machine learning-based prevention technique to detect and prevent intrusions with low FPR and without prior knowledge of attacks.

The rest of this chapter is organized as follows. In Section 4.2 we compare our method with existing approaches. An overview of HAN security threats is provided in Section 4.3. Architecture and algorithm of HANIDPS is explained in Section 4.4. Section 4.5 and 4.6 present performance analysis and experimental evaluations of HANIDPS. Section 4.7 summarizes the chapter.

4.2 Related Work and Comparison

Designing IDSs tailored for smart grid subsystems has attracted the attention of researchers over the last few years. Mitchell et al. [26] introduced a behavior-rule based IDS (BIDS) for securing head ends, distribution access points and smart meters. For each section a set of high-level behavior rules were defined. An intrusion was detected when the behavior rules were violated. This method provides a high accuracy; however, since the behavior
rules are high level, BIDS is subject to detection delay. Besides, it does not provide an insight into the cause of misbehaviors, therefore presents little information for stopping the attack. A hierarchical distributed IDS for AMI was proposed in [27]. The distributed IDS components were connected through a wireless mesh network. Each component employed support vector machine (SVM) and immune system for detecting intrusions. Applying the same solution for different AMI networks including HANs, NANs and head ends which use different protocols and have different traffic features makes this method inefficient. Unlike [26] and [27] which use the same mechanism for intrusion detection for various AMI networks, we focus on HANs. This enables us to provide a high performance mechanism with the ability of both detecting and preventing the attacks. Authors of [28] and [29] targeted the detection of false data injection (FDI) attacks in smart grids. Lo et. al [28], proposed a hybrid IDS framework for AMI which uses power information and sensor placement to detect FDIs. They introduced algorithms for placing the sensors on lines or feeders to enhance the detection performance. Chen et. al [29], exploited spatial-temporal correlations between grid components for real-time detection of FDIs. A distributed IDS tailored for wireless mesh networks employed in NANs, was proposed in [30]. This work specifically targeted the network layer attacks. In [31] a specification-based IDS for communications between smart meters and data aggregators was presented. Using C12.12 standard protocol a set of constrains on data transmissions was made and attacks were detected by monitoring the violations of the security policy. Authors of [90] introduced a two-tier IDS for automatic generation control (AGC) in smart grids. The first tier was a short-term adaptive predictor for system variables, and the second tier performed state inspection to investigate the presence of anomalies. Combination of the two tiers provided a balance between accuracy and real-time requirements of IDS for AGC. We on the other hand focus on HAN protocols and provide a solution for not only detecting but also preventing the
attacks. Model-based IDS for intrusion detection in wireless sensor networks has attracted the attention of researchers in the past [32],[33],[91]. When the number of applications and protocols are limited, model-based IDSs are very effective, since they provide a low FPR and are capable of detecting new attacks. In [32] and [33] model-based IDSs for modbus networks in SCADA were presented. Ioannis et al. [91] introduced a specification-based IDS for detecting network layer attacks in wireless sensor networks including blackholes and grayholes. HANIDPS also uses a model-based approach. However, it distinguishes itself from [32],[33],[91] by covering the unique requirements of the area. First, HAN is a stationary network which eliminates the need for monitoring moving objects. Second, HAN coverage area is comparably small and most communications are single hop. Therefore, unlike most IDS solutions for wireless sensor networks which are tailored for network layer protocols our focus is on PHY and MAC layers. Third, none of [32],[33],[91] addressed the problem of intrusion prevention for a wide range of attack types. A game theory-based approach for detecting and preventing distributed DoS using transport layer flooding attacks in wireless sensor networks was presented in [92]. Defense mechanism in this work was dropping the packets once an attack is detected. HANIDPS employs a wider and more effective set of features and actions and therefore is capable of detecting and preventing a larger variety of attacks.

4.3 Home Area Network Security Threats

Threats against AMI can be viewed in three different ways: by type of attacker, motivation and attack technique. In [18] a threat model for AMI was provided which lists the types of attackers and their motivations as follows:

- Curious eavesdroppers, who are motivated to learn about the activity of their neighbors by listening to the traffic of the surrounding meters or HAN.
• Motivated eavesdroppers, who desire to gather information about potential victims as part of an organized theft.

• Overly intrusive meter data management agencies, which are motivated to gain high-resolution energy and behavior profiles about their users, which can damage customer privacy. This type of attacker also includes employees who could attempt to spy illegitimately on customers.

• Unethical customers, who are motivated to steal electricity by tampering with the metering system installed inside their homes or to gain control of the devices which should be under control of the utility.

• Active attackers, who are motivated by financial gain or terrorist goals. The objective of a terrorist would be to create large-scale disruption of the grid, either by remotely cutting off many customers or by creating instability in the distribution or transmission networks. Active attackers attracted by financial gain could also use disruptive actions, such as DoS attacks.

• Publicity seekers, who use techniques similar to those of other types of attackers, but in a potentially less harmful way, because they are more interested in fame and usually have limited financial resources.

From the above list unethical customers, active attackers and publicity seekers require to perform active attacks to achieve their goals. Curious and motivated eavesdroppers might perform active or passive eavesdropping. IDSs are effective in protecting the network against active attacks. Overly intrusive meter data management agencies use the data available in head ends rather than targeting the HAN; therefore, are outside the scope of this work.
Smart grid is a new concept. A thorough threat model which provides details on possible attack scenarios and techniques, for different AMI networks and devices, is not available yet; providing such models is an open research topic. Few works have been done in this area over the past few years. In [93] a model for security analysis of smart meters was provided. The authors proposed a systematic method for modeling functionalities of smart meters and deriving attacks that can be mounted on them. McLaughlin et al. [94] conducted a multi-vendor penetration testing on AMI devices within NANs. They developed archetypal attack trees for three classes of attacks: energy fraud, denial of service and targeted disconnect. Grochcki et al. [95] surveyed various threats facing AMI and common attack techniques used to realize them; however, authors of [95] mentioned that methods of compromising HANs are beyond the scope of their work. While a thorough threat model for HANs does not exist in literature yet, and defining one is a stand alone research topic which is beyond the scope of this thesis, in the following we provide some examples of possible attacks against ZigBee HANs. We also explore a number of representative case studies to connect attacker objectives with individual attack steps. Table 4.1 shows a summary of attacks against AMI, based on the threat model provided in [95]. These attacks are applicable for HANs. For detail explanation of each attack we refer the readers to [95].

4.3.1 Case Studies

Illegitimate remote turn-on/off commands

This attack can be used by unethical customers, publicity seekers and active attackers with different motivations. An unethical customer might aim to gain control of a specific device which according to the customer utility agreement is under control of the utility. An antisocial publicity seeker might use this attack to achieve fame or unsettle a customer.
Table 4.1: Example of attacks against ZigBee HANs

<table>
<thead>
<tr>
<th>Category</th>
<th>Attack technique</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoS</td>
<td>Collision in packet transmission</td>
<td>HAN link layer</td>
</tr>
<tr>
<td>DoS</td>
<td>Jamming</td>
<td>HAN physical layer</td>
</tr>
<tr>
<td>DoS</td>
<td>Resource exhaustion (battery, bandwidth, CPU)</td>
<td>Node in HAN</td>
</tr>
<tr>
<td>DoS</td>
<td>Destroy node</td>
<td>Node in HAN</td>
</tr>
<tr>
<td>Spoofing</td>
<td>Impersonate regular node</td>
<td>Node in HAN</td>
</tr>
<tr>
<td>Spoofing</td>
<td>Impersonate master node</td>
<td>Node in HAN</td>
</tr>
<tr>
<td>Spoofing</td>
<td>Man-in-the-middle</td>
<td>HAN traffic</td>
</tr>
<tr>
<td>Spoofing</td>
<td>Brute-force</td>
<td>Node in HAN</td>
</tr>
<tr>
<td>Eavesdropping</td>
<td>Passively listen to traffic</td>
<td>HAN traffic</td>
</tr>
<tr>
<td>Eavesdropping</td>
<td>Active cryptanalysis</td>
<td>HAN traffic</td>
</tr>
<tr>
<td>Physical</td>
<td>Compromise meter</td>
<td>Node in HAN</td>
</tr>
</tbody>
</table>

An active attacker with malicious intentions such as committing theft, kidnapping, etc. can use this attack to distract the customers. Terrorists might use this attack in large scale to cause horror and chaos, or to affect the load curve in order to damage the power system equipment. As a case in point, the attacker sends turn-off messages to all controllable customer equipment. After passing a long enough time which guarantees that the equipment would turn on when allowed, the attacker sends turn-on permission messages. When applied in large scale, the attack results in a sudden increase in the load, which can affect the bulk electric grid. Considering that for performing this attack against HANs the attacker must be within the ZigBee communication range, conducting it in large scale is expensive. But terrorists can be very motivated and have high funding. This attack can be done using the following steps.

1. The attacker passively eavesdrops the network traffic or perform active network scanning to learn the link layer address of the authoritative node like EMS and the victim node.

2. The attacker needs to learn the authentication credentials of the authoritative node (if authentication is supported); this information can be obtained using brute-force
3. The attacker needs to know the encryption key to encrypt the control commands (if encryption is supported). Cryptanalysis techniques can be used for this purpose. Man-in-the-middle attacks might also be helpful in bypassing the encryption system.

4. The attacker conducts DoS against the authoritative node to stop it from sending legitimate control commands.

5. The attacker impersonates ID of the authoritative node and sends turn-on/off commands on its behalf to the victim node.

**Stealing customer information**

The motivation of this attack is to collect customer information and learn about customer behavior. For instance, in an organized theft an adversary can benefit from knowing the total electricity usage of the household to infer whether the customers are at home or not. The EMS is allowed to request this information from the smart meter, and the smart meter sends the information to EMS through encrypted messages. Considering that HAN traffic may be encrypted and authentication might be required for a node to access the network, this attack may involve the following steps.

1. The attacker passively eavesdrops the network traffic or perform active network scanning to learn the link layer address of the EMS.

2. The attacker needs to learn the authentication credentials of EMS (if authentication is supported); this information can be obtained using brute-force attack.

3. The attacker needs to know the encryption key to encrypt/decrypt the massages (if encryption is supported). Cryptanalysis techniques can be used for this purpose.
4. The attacker impersonates ID of the EMS and requests for the usage information.

5. The attacker decrypts the messages and collects the message contents.

**Denial of service against network nodes**

An unethical customer may conduct DoS against HAN nodes with the purpose of gaining control of a specific device. By conducting DoS against authoritative nodes or a sensor node on a specific device such as thermostat, customer intervenes with the control commands by the utility and does not allow the utility to control the device. In Section 4.6.1 we introduce several inexpensive DoS techniques against IEEE 802.15.4 networks.

### 4.4 Home Area Network Intrusion Detection and Prevention System

#### 4.4.1 Architecture

HANIDPS is designed for physical (PHY) and medium access control (MAC) layers of ZigBee HANs and has two modules, detection and prevention.

1) Detection module: The detection module monitors the network traffic between the smart meter and sensor nodes, as well as nodes’ behavior and extracts the network features. These features are analyzed and compared with the expected normal behavior based on the system specification. If one or more of the features are not normal an intrusion is detected and the prevention module is triggered.

2) Prevention module: Upon receiving an intrusion alarm an action or a set of actions are automatically performed to stop or mitigate a possible attack. HANIDPS preventive actions include spoofing prevention, interference avoidance and dropping malicious packets.
An adversary might use various attack types to disturb the network operation. Therefore, appropriate actions are required to countermeasure different attack scenarios. HANIDPS uses reinforcement learning to find the best strategy against an attacker. In reinforcement learning, the process of learning happens via trial and error. Through feedbacks received from the environment, HANIDPS learns what the best and most effective actions are.

HANIDPS has two components, monitoring agents and central IDPS (C-IDPS). Agents are installed on sensor nodes and are responsible for monitoring the behavior of the corresponding nodes. They count the packet error rate (PER) and keep a record of RSS values of received frames from each communicating node. Agents send measured PERs to the C-IDPS periodically through health messages. Once an attack is detected by C-IDPS, agents might also take part in prevention operations based on the recorded RSS values. C-IDPS is installed on a full function device super-node, which is a tamper resistant device with higher capacity and computational power compared to normal nodes. Network traffic between sensor nodes flow through C-IDPS where traffic features are extracted. C-IDPS is responsible for analyzing the network features and health messages received from monitoring agents to infer the state of the network nodes. Once an abnormal state is detected C-IDPS performs a dynamic defense mechanism in which through Q-learning a set of preventive actions are selected and performed. In the following subsections, we explain the feature space, preventive actions and decision making process to choose the best sequence of actions.

### 4.4.2 Features and State Space

The ZigBee alliance, HomeGrid Forum, HomePlug alliance and WiFi alliance created a consortium for interoperability of energy management devices in HANs. The goal was to define the interfaces and messages among smart appliances, smart meters and utilities. The
alliances published a new draft for smart grid application communication, SEP 2.0 in July 2012. NIST selected SEP 2.0 [96] as a standard profile for smart energy management in home devices. We used this document as well as the IEEE 802.15.4 protocol specifications and common features of wireless networks to extract network specifications used in defining the feature space for HANIDPS. Components of the feature space are as follows:

$f_1$) Datagram (D): PHY and MAC layer frame structures according to IEEE 802.15.4 specification are defined for C-IDPS. C-IDPS compares some features of the transmitting frames like frame size and reserved bits with the standard structure and decides whether the frames are normal. Also according to SEP 2.0, application layer protocol for home devices is Hyper Text Transfer Protocol with Secure Socket (HTTPS) over Transmission Control Protocol (TCP). The minimum and maximum length for each message, as well as the mandatory and optional fields were defined in the protocol specification. Therefore, the theoretical minimum and maximum length of packets can be calculated. The total length of a legitimate SEP 2.0 message is between 508 and 1524 bytes.

$f_2$) Traffic Rate (TR): SEP 2.0 states that ”to prevent overwhelming network resources, notifications should be sent to a given client for a given resource no more than once every 30 seconds. Notifications for conditional subscriptions should only be sent once within this time period for a given client for a given resource and any additional notifications should not be queued. All devices need to be considerate of network resources”. While in SEP 2.0 end devices are responsible for pulling network time from the network controller, it has been mentioned that the granularity of the devices must be one second. Therefore, a genuine node that complies with the protocol specifications does not surpass a limit for traffic rate.

$f_3$) RSS: According to the laws of physics, the signal strength at a receiver antenna is proportional to the spatial distance between the receiver and the sender. Beside distance,
Chapter 4. HANIDPS, An Intrusion Detection and Prevention System for ZigBee HANs

Table 4.2: Requirements of HAN according to the U.S. department of energy guideline.

<table>
<thead>
<tr>
<th>Smart Grid Functionality</th>
<th>Bandwidth, Latency, Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Metering Interface</td>
<td>100 kbps/node, 2-15 sec, 99-99.99%</td>
</tr>
<tr>
<td>Demand Response</td>
<td>100 kbps/node, 0.5-2 sec, 99-99.99%</td>
</tr>
<tr>
<td>Distributed Energy Resources</td>
<td>100 kbps/node, 0.02-15 sec, 99-99.99%</td>
</tr>
<tr>
<td>Electric Vehicles</td>
<td>100 kbps/vehicle, 2 sec-5 min, 99-99.99%</td>
</tr>
</tbody>
</table>

RSS depends on wireless environment features, such as absorption and multipath effect, which makes it hard to predict the power level of frames collected by a receiver. Thus, an attacker cannot simply adjust his power level to match the RSS of a legitimate node. This feature is useful in detecting spoofing attacks in which an adversary masquerades identity of a legitimate node and send traffic on its behalf. In Chapter 3, we have introduced a high performance RSS-based method for detecting spoofing attacks in static IEEE 802.15.4 networks. We use the same algorithm in HANIDPS to decide whether RSS values are normal or suspicious.

$f_4$) Sequence Number (SN): The regular ordering of sequence numbers according to the standard is defined for C-IDPS. Unusual sequence numbers can be suspicious.

$f_5$) Packet Error Rate (PER): A major cause of high PER is the use of a busy channel. IEEE 802.15.4 employs carrier sense multiple access/collision avoidance (CSMA-CA) to evaluate availability of the channel. Under normal condition PER must be low. However, illegitimate (such as jamming sources) or legal (such as WiFi interference) coexistence of other signals can increase the PER. This will affect the network throughput and cause latency in packet delivery. We use a mathematical model for ZigBee performance to find the acceptable PER which allows the required bandwidth and latency for HAN operations.

A summary of network requirements for HAN, as indicated by the U.S. department of energy [97], is provided in Table 4.2. In [98], a model for computing maximum delay and upper bound on data rate in beacon-enabled guaranteed time slot (GTS) Zigbee networks
was provided as in (4.1) and (4.2)

\[ D_{\text{max}} = \frac{b}{C} + (k - 1) \times BI - (T_{GTS} + T_R) - k \times T_{GTS} \] (4.1)

\[ B_{C,T}(t) = C \times T_{GTS} + C \times (t - (2 \times BI - T_R)) \] (4.2)

where \( D_{\text{max}} \) is the maximum delay, \( b \) is the minimum burst size, \( C \) is the service rate which for ZigBee is 250 kbps, \( k \) is the number of GTS slots, \( BI \) is the beacon interval, \( T_{GTS} \) is the duration of data transmission within the GTS slot, \( T_R \) is the portion of the GTS during which there is no data transmission, and \( B_{C,T}(t) \) represents the upper bound on data rate.

To account for network interference, and considering that the number of re-transmissions before declaring channel access failure in ZigBee specification is 5, [98] modified (4.1) and (4.2) by replacing \( C \) with \( C_{adj} \) as in (4.3):

\[ C_{adj} = P \times C \times \sum_{i=0}^{4} (1 - P)^i \] (4.3)

where \( P \) is the probability of a clear channel. As (4.1), (4.2) and (4.3) suggest the maximum delay and guaranteed data rate of a client is a function of the number of assigned GTSs and the probability of clear channel for the client. We calculate the minimum probability of clear channel for different numbers of assigned GTSs, which allows the required performance according to Table 4.2. PER for an IEEE 802.15.4 network in presence of interference is formulated as follows [99]:

\[ PER = 1 - (1 - P_b)^{N_z^{-\lceil \frac{P_R}{P_z} \rceil}} \times (1 - P_b^{\lceil \frac{P_R}{P_z} \rceil}) \] (4.4)

\( P_b \) is bit error rate (BER) without interference, \( P_b^{\lceil \frac{P_R}{P_z} \rceil} \) is BER with interference, \( N_z \) is the
Chapter 4. HANIDPS, An Intrusion Detection and Prevention System for ZigBee HANs

number of bits in a packet, \( b \) is the duration of bit transmission, and \( T_c \) is the collision time. Without interference \( P = 1 - P_b \) and in presence of interference \( P = 1 - P^I_b \).

Therefore, we can calculate the threshold for acceptable PER. For instance, using (4.1), (4.2) and (4.3) and according to Table 4.2 the theoretical ranges that satisfy the demand response requirements in terms of number of GTS slots and probability of clear channel are bounded by \((4,0.99), (5,0.7), (6,0.5), (7,0.3)\). In calculating these points the following parameters were used based on the ZigBee and SEP 2.0 specifications: \( b=508 \) bytes, \( C=250 \) kbps, \( BI=960 \) symbols.

\( f_6 \) Node Availability: Represents whether or not health messages are received from the agents.

The detection module evaluates each of the above features and if one of them is abnormal, an intrusion is detected and the prevention module is triggered. The results of the evaluation of these features are also used to define the state space for the prevention module. The components of the state space are defined as \( \{f_1,...,f_6\} \) where \( f_i, i=1,...,6 \) are binary values, assigned by evaluating features 1-6. For each of D, TR, RSS, SN, PER and NA, C-IDPS checks whether or not the feature is normal, as described above, and accordingly assigns a binary value to \( f_i, 0 \) if the feature is normal and 1 if it is abnormal.

### 4.4.3 Actions

\( a_1 \) Spoofing Prevention: In Chapter 3, we proposed an RSS based method for distinguishing and filtering illegitimate packets in static IEEE 802.15.4 networks. We use the dynamic threshold method introduced in Chapter 3, Section 3.6.2 as a defensive mechanism in HANIDPS.

\( a_2 \) Interference Avoidance: In [99] an algorithm for detecting and avoiding WiFi interference in ZigBee networks was proposed. The method is also effective in combating other
sources of interference such as jamming signals. We adopt the interference avoidance scheme in [99] as one of the actions that HANIDPS may perform when encountering with an abnormal state. The summary of the algorithm is as follows. C-IDPS checks its link quality indicator (LQI). LQI is a MAC layer parameter which indicates the current quality of received signals, and provides estimation on how easily received signals can be demodulated. LQI value ranges from 0 to 255 and is inversely related to PER. When LQI is small, it can be inferred that a high PER is due to poor link quality rather than problems with end device. If the LQI is low the coordinator makes all the routers within the personal area network (PAN) to perform interference assessment through energy detection (ED) scans defined in ZigBee protocol. During an ED test, transceiver scans all the IEEE 802.15.4 complaint channels in the frequency band supported by the transceiver. If ED is beyond the threshold of 35 (which corresponds to the noise level between -65 dBm to -51 dBm) interference is detected. Based on the results of ED scans the coordinator selects a channel with an acceptable quality and all PAN devices migrate to this new channel.

*a3*) Packet Drop: C-IDPS discards the received packets, which datagram or sequence numbers deviate from those predicted by the protocol, without forwarding them to the intended destinations.

*a4*) Packet Forward: C-IDPS forwards the received packets without further processing. This is helpful when an attacker targets the IDPS by sending high rate traffic and by exhausting the IDPS causes DoS. Another example is when none of the actions is effective in mitigating the attack and only impose overhead to the system.

### 4.4.4 Learning Algorithm

After evaluating the network features, C-IDPS infers the network state $s$ and if an intrusion is detected by the detection module, it performs an action $a$ which results in transition to
a new state \( s' \). The transition probability only depends on the current state and action, and is independent from all previous states and actions; therefore, satisfies the Markov property. Considering that attackers use different strategies to target the network, the interaction between HANIDPS and the attacker creates a dynamic environment, meaning that by taking an action in a given state, the next state is unpredictable. Therefore, among various existing reinforcement learning algorithms we use Q-learning which is suitable for dynamic environments. In Q-learning a utility function is defined as a map between the state-action pairs and their Q values. Q-values predict the cumulative reward that will be received following the state-action. When the state space is not too large a look up table (LUT) can effectively be used as the utility function. The LUT can be initialized by zero or based on previous knowledge of the environment. During the learning process the time is divided into decision epochs. In each epoch the agent chooses action \( a \) in state \( s \) which results in transition to state \( s' \) and receiving a reward or penalty based on how appropriate the transition is. After each epoch the LUT is updated using (4.5):

\[
Q(s, a) = Q(s, a) + \alpha (R(s, s', a) + \max_{a'} \gamma Q(s', a') - Q(s, a))
\] (4.5)

where \( \alpha \) is the learning rate and \( \gamma \) is the discount factor. We use polynomial learning rate in which \( \alpha = 1/(1 + t)^\omega \) since it has a faster convergence rate compared to linear learning rate [100]. As explained in Section 4.4.2, in our problem the state elements are binary values assigned based on whether or not the network features are normal. At state \( s \), state elements are \( \{f_{s1}, ..., f_{s6}\} \). At each state four actions described in Section 4.4.3 are possible: spoofing prevention, interference avoidance, packet drop and packet forward. By performing action \( a \) in state \( s \) the process moves to state \( s' \) with state elements \( \{f'_{s1}, ..., f'_{s6}\} \). To assess the new state, C-IDPS extracts and reevaluates the features of network traffic after performing the action. Following each action a reward is received based on the state
transition as formulated in (4.6). We define the reward as a function of the changes in state elements and the cost of performing an action as follows:

\[ R(s, s', a) = \sum_{i=1}^{6} \beta_i (f^{s}_i - f^{s'}_i) - \text{cost}(a) \]  

(4.6)

where \( \beta_i \) is the weight of feature \( i \) which is assigned according to the importance of each feature. For instance, given that node availability might have more priority than having a normal traffic rate, by choosing \( \beta_6 > \beta_2 \), higher reward will be assigned to the action which keeps the node available. The cost function, \( \text{cost}(.) \), reflects the costs associated with performing an action. It accounts the network overhead, imposed delay and resource usage such as battery consumption following performing an action. As a case in point, the cost of dropping packets is lower than changing the channel which results in some network overhead, delay and resource usage. While the LUT can be initialized randomly, there exist some relationship between the features and actions. For instance when the datagram does not comply with the specifications, dropping the packets might be a better choice compared to other actions. This knowledge can be employed to initialize the LUT to reduce the learning time. If actions \( a_2 \) or \( a_4 \) is selected, \( \{f^{s'}_1, ..., f^{s'}_6\} \) are equal to the elements of the new state \( s' \). Otherwise, if actions \( a_1 \) or \( a_3 \) is performed, C-IDPS reevaluates features \( f_1 \) to \( f_4 \) of the forwarded packets by the C-IDPS and accordingly defines \( \{f^{s''}_1, ..., f^{s''}_4\} \). \( f^{s''}_5 \) and \( f^{s''}_6 \) are equal to \( f^{s'}_5 \) and \( f^{s'}_6 \), respectively. We used this method because when actions \( a_1 \) or \( a_3 \) are performed C-IDPS drops some of the network packets and in order to decide how suitable these actions have been we need to evaluate some features of the forwarded packets rather than all of the network traffic.
4.5 Performance Analysis

4.5.1 Performance of the Detection Module

Among 6 features used in HANIDPS to detect network abnormalities, 4 (D, TR, SN, PER) are directly extracted from system specifications and are not considered major sources of false positives. For instance, a healthy packet never has a D different from what the protocol defines. SN of legitimate packets also does not circumvent the specification. The probability of having a TR and PER beyond the threshold under normal condition is very low, since SEP 2.0 specified strict rules for traffic rates and network requirements. While ambient noise, temporary system faults or wireless communication faults can produce false positives in evaluating NA and some other features, the major cause of false positives in HANIDPS is the RSS evaluation results, since RSS is a statistical parameter and a trade off between FPR and DR is required. FPR of the detection module is formulated as follows:

\[
FPR = P \left( \bigcup_{i=1}^{6} (f_i = 1) \mid normal \right) \leq \sum_{i=1}^{6} P^i_{normal} (f_i = 1) \approx P^3_{normal} (f_3 = 1) \quad (4.7)
\]

where \( P^i_{normal} \) is the probability distribution of \( i^{th} \) feature under normal condition. As we have shown in Chapter 3:

\[
P^3_{normal} (f_3 = 1) = 1 - F_{\chi^2(m)} \left( \frac{\tau}{2\sigma^2} \right) \quad (4.8)
\]

\( \chi^2(m) \) is Chi-square distribution with \( m \) degrees of freedom which is equal to the number of AMs used in spoofing detection. HANIDPS uses one AM which is in the C-IDPS. \( F_{\chi}(.) \) represents the CDF of \( \chi \), \( \sigma \) is the variance of RSS values and \( \tau \) is a threshold. When \( \tau \) is
Chapter 4. HANIDPS, An Intrusion Detection and Prevention System for ZigBee HANs

exceeded, a spoofing attack is detected. The DR of the spoofing detection module is:

\[ DR = P_{\text{attack}}^3 (f_3 = 1) = 1 - F_{\chi^2(m, \frac{\lambda}{2\sigma^2})} \left( \frac{\tau}{2\sigma^2} \right) \]  

(4.9)

where \( P_{\text{attack}}^3 \) is the probability distribution of RSS feature under attack and \( \lambda \) is the distance between SDC distribution of the attacker and the genuine nodes. For further explanation on performance of the spoofing detection module we refer the readers to Chapter 3.

For other features attacks are only detectable if they create a network traffic that does not comply with the SEP 2.0 specifications. This, however, enables detection and prevention of a variety of attack types.

4.5.2 Performance of the Prevention Module

The intrusion prevention module performs two types of actions, training and operation.

1) Training actions: These actions are taken for the purpose of training the Q-learning. They are selected before the algorithm is converged and are not effective in mitigating attacks. Therefore, they are only considered sources of network and system overhead. Assuming a simple case when one action, \( a_j \), is effective in mitigating the attack, and \( T \) iteration is required before the algorithm converges, the overhead of the learning algorithm is:

\[ \text{Overhead} = \frac{1}{3} \sum_{t=1}^{T} \sum_{i=1, i \neq j}^{4} P_i(t) O(a_i) \]  

(4.10)

where \( P_i(t) \) is the probability of choosing action \( i \) at iteration \( t \), and depends on the initial values, reward function and the order of states-actions in the LUT. \( O(a_i) \) represents the overhead of performing action \( i \). \( T \) is the convergence time of the algorithm. In [100] the
convergence rate of Q-learning was studied. The authors showed that in asynchronous Q-learning with polynomial learning rate, aside from parameters of Q-learning, convergence time is related to the covering time, $L$. The covering time indicates the number of state-action pairs starting from any pair, until all state-actions appear in the sequence with probability of at least 0.5. The order of this dependency is $\Omega(L^{2+\frac{1}{\omega}})$ which is optimized for $\omega = 0.77$. Having 6 binary features, the state space size of HANIDPS is 64; for each state there exist 4 possible actions. However, not all states are experienced during a specific attack. The number of states that are crossed during an attack, and therefore the covering time depend on the attack complexity. The number of states can range from 2 when the attacker uses a specific attack type which follows a same routine over time, to a few when the attacker reacts and adjust his/her strategy according to the actions taken by HANIDPS. Our experiments in Section 4.6 shows that the algorithm converges with the required precision in few iterations which keeps the learning overhead in an acceptable range.

2) Operation actions: These actions are taken after the algorithm is trained and are chosen as the best defense mechanism against the attack.

2.1) Performance of Spoofing Prevention: In Chapter 3, Section 3.7.2 we evaluated the performance of the spoofing prevention algorithm in detail.

2.2) Performance of Interference Avoidance: Experiment results in [99] showed that with the ED scan duration of 135 ms, the proposed algorithm provides the best balance between scan duration and accuracy, where the LQI readings for 4600 packet transmission for each channel is analyzed. During this time network nodes cannot transmit normal traffic. In return, the authors showed that by choosing a less interfered channel, the sensors battery life can be prolonged by up to 2-3 years; while if the network operates under high interference, a high PER will result in a large retransmission rate which wastes
Chapter 4. HANIDPS, An Intrusion Detection and Prevention System for ZigBee HANs

the energy of sensors. The other two actions including packet forward and packet drop does not impose significant delay or network overhead.

4.6 Evaluations

To evaluate the performance of HANIDPS, we study existing attacks against IEEE 802.15.4 networks and analyze the detection and prevention capability of HANIDPS against them. Further, we conduct two experiments to show how HANIDPS dynamically learns the most efficient strategy against an attack.

4.6.1 Home Area Network Intrusion Detection and Prevention System against IEEE 802.15.4 attacks

1) Radio Jamming: Radio jamming is intentional or unintentional emission of radio signals which by decreasing the signal to noise ratio disturb data flow of a wireless network. When the network is under jamming, PER and NA are not in the expected range. High interference increases the PER, and since nodes are not able to communicate properly, health messages will not be received by the C-IDPS. When these two features of the feature space are abnormal, HANIDPS detects an attack and triggers the prevention module. From the actions defined for the HANIDPS, interference avoidance is effective in stopping unintentional jamming. C-IDPS changes the network channel to a new high quality one. Coexistence of WiFi networks is one of the significant concerns in ZigBee HANs [99], since high rate WiFi traffic can cause unintentional jamming. When jamming is unintentional the interfering device will not change its channel and therefore by migrating the HAN to the new channel the problem is resolved. This is also true for simple cases of intentional jamming. But in more complicated attacks, for instance when the attacker also changes its
channel or conduct a wide band jamming which covers the whole ZigBee frequency range, interference avoidance scheme will not be effective. Yet these attacks are more expensive and energy consuming.

2) *Replay-protection:* IEEE 802.15.4 uses a replay-protection mechanism in which the sequence number of a received frame is compared with the sequence number of the previous frame. If the former is equal or smaller than the latter this frame will be dropped. An attacker can send frames with large sequence numbers to a receiver, causing it to drop legitimate frames. The detection module checks the sequence number of the frames. If they do not comply with the protocol standard, SN feature will be abnormal and the attack will be detected. By dropping the illegitimate packets, which is one of the actions defined for the HANIDPS, this attack can effectively be stopped. The prevention module learns the proper action after a few iterations.

3) *Steganography:* An attacker uses the reserved fields of packets to create a hidden channel and transfer hidden data. A detailed investigation of steganography attacks in IEEE 802.15.4 was reported in [101]. HANIDPS checks datagram of packets and if it is abnormal (for instance when reserved bits are not 0 as it happens in steganography) detects an attack. The effective action against this attack is dropping malicious packets which is learned through reinforcement learning and performed by the prevention module in HANIDPS.

4) *Back-off manipulation:* A malicious node steals channel access of legitimate nodes by using an instantly short back-off period. The malicious node is either one of the network nodes that has been penetrated, or an outside node which forges the ID of one or more legitimate nodes and conducts spoofing to access the network. In both cases the attacker has a very high traffic rate and TR for that node will be 1. Also since nodes are not able to transmit their packets properly, health massages might not be received by C-IDPS and
NA will be 1. Besides, not being able to access the channel increases the PER. In the case of spoofing, RSS values are not normal either. Hence, by causing abnormal TR, NA, PER and RSS the attack is easily detectable. The prevention module can mitigate the attack by filtering high rate traffic from the attacker and not allowing the coordinator to assign the channel to that node. Another action that can be helpful, in case of abnormal RSS values, is spoofing prevention. Through interaction with the environment and receiving rewards and penalties HANIDPS converges to the best action.

5) DoS against data transmission during contention free period (CFP): A malicious node extracts ID and GTS number of legitimate nodes through eavesdropping; then forges their IDs and by sending GTS deallocation requests terminates their traffic [58]. Since the malicious and legitimate nodes are not located at the same place, their RSS values will not be similar and the attack is detectable by evaluating RSS feature. Also through spoofing prevention action, packets with abnormal RSS values are distinguished and filtered. Therefore, once the prevention module learns what the appropriate action is, the attack is stopped.

6) DoS against GTS requests: An adversary keeps track of the GTS list and fills up all available GTSs by sending several GTS allocation requests. As a result, legitimate nodes will not have the chance to transmit their data during the CFP [58]. Again, since RSS of the attacker does not match that of the genuine node, the attack is detectable and through spoofing prevention malicious packets can be distinguished and dropped.

4.6.2 Experiment 1

We study the performance of HANIDPS under a simple spoofing attack where other than RSS values traffic features of malicious node is consistent with the protocol. DoS against GTS requests is an example of this attack. We used the same method and testbed as it
Chapter 4. HANIDPS, An Intrusion Detection and Prevention System for ZigBee HANs

has been explained in Chapter 3, Section 3.8 to implement and evaluate the performance of the HANIDPS against a spoofing attack.

2) Performance of the detection module

By applying the spoofing detection mechanism on RSS logs, in average we observed 92.5% DR for 0% FPR. The average ROC curve based on the RSS logs of 4 AMs is depicted in Fig. 4.1. When the attack is detected the state of the victim node is \{0,0,1,0,0,0\} which triggers the prevention module.

3) Performance of the prevention module

Training Phase: We implemented Q-Learning in MATLAB. The LUT was initialized by 0.5. We considered the same weight for all features, and assigned the cost of 0.25, 0.5, 0.1, 0 to actions 1-4, respectively. The highest cost was defined for interference avoidance since it is an energy consuming process. The first four actions were exploratory [100]. Then we reduced the exploration rate by a factor of \(\frac{4}{i}\) as the algorithm proceeded (\(i\) is the number of iteration). Use of exploratory actions allows the algorithm to explore all actions and converge to the most effective ones. The learning rate was polynomial with \(\omega = 0.77\) and the discount factor was \(\gamma = 0.1\). Following the spoofing prevention action the state of the node was changed to \{0,0,0,0,0\} while for other actions the state remained unchanged. We observed that for this simple case where only two states are experienced, the algorithm learned the best action after 4 iterations.

Operation phase: We evaluated the performance of the spoofing prevention mechanism
Table 4.3: Performance for dynamic threshold spoofing prevention.

<table>
<thead>
<tr>
<th>Threshold coefficient</th>
<th>1/2</th>
<th>1/3</th>
<th>1/4</th>
<th>1/5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR(%)</td>
<td>93.46</td>
<td>96.07</td>
<td>97.14</td>
<td>97.62</td>
</tr>
<tr>
<td>FPR(%)</td>
<td>6.16</td>
<td>11.17</td>
<td>16.21</td>
<td>21.68</td>
</tr>
</tbody>
</table>

as the effective action against the attack. We applied the spoofing prevention algorithm to RSS logs. This time the AM nodes played the role of a victim. Performance was measured when the threshold was 1/2, 1/3, 1/4 and 1/5 of the distance between the mean values. The window size, i.e. the number of frames which are clustered each time was 64. The results are shown in Table 4.3.

4.6.3 Experiment 2

In the second experiment we consider a scenario in which an attacker uses ID of a legitimate node and sends packets that do not comply with the standard datagram. For instance, the value of some of the reserved fields in the packets are different from what have been specified in the standard, this is an example of Steganography attack. In this scenario the state of the victim node is \(\{1,0,1,0,0,0\}\). Both \(a_2\) and \(a_3\) can stop the attack and change the state to \(\{0,0,0,0,0,0\}\). However action 3 has a lower overhead and in this case has a higher accuracy. By implementing this scenario with the same parameters as in experiment 1 we observed that the algorithm was converged to choose action 3 after 4 iterations.

4.7 Summary

In this chapter we have introduced HANIDPS, a novel IDPS for ZigBee-based HANs. Considering the insecure environment, use of wireless technology and limited resources of HAN devices, HAN is vulnerable to cyber attacks which necessitates application of appropriate IDSs. Also due to the large scale and high cost of false positives, in the context of HAN,
Chapter 4. HANIDPS, An Intrusion Detection and Prevention System for ZigBee HANs

IDPSs which not only detect but also automatically stop the attacks are highly required. HANIDPS combines a model-based intrusion detection method tailored for HAN specifications and a machine learning-based prevention technique which enables dynamic defense against adversaries without prior knowledge of the attacks. Using novel techniques for spoofing prevention, and through utilization of effective mechanism for countermeasuring intentional and unintentional interference, HANIDPS secures the network against a variety of attack types. Extensive analysis and simulations have proved the effectiveness of our approach.
Chapter 5

Detection of Malicious Activities in AMI Using Customers Consumption Patterns

5.1 Introduction

One important property of AMI is that unlike most traditional IT systems, a good portion of the data transferred through AMI is to a high degree predictable. Smart meters send their usage reports to the utilities in predefined time intervals. Not only are the time intervals between messages constant, but also it is possible to find a statistical model for customer’s consumption pattern. Irregularities in usage pattern can be a sign of some malicious activities. In this section, we leverage the predictability of AMI data to detect adversarial activities such as energy theft and attacks against direct and indirect load control.

Non-technical losses is one of the major concerns in smart grids, since application of digital smart meters and addition of a cyber layer to the metering system introduce numerous new vectors for energy theft. Current AMI ETDSs are mainly categorized into three groups [102]: state-based, game theory-based and classification-based. State-based detection schemes [28],[36],[37] employ specific devices, like wireless sensors and radio-
frequency identification tags, to provide a high detection accuracy, which however, come with the price of extra investment required for the monitoring system including device cost, system implementation cost, software cost and operating/training cost. In game theory-based methods [38],[39], the problem of electricity theft detection is formulated as a game between the electricity thief and the electric utility. These methods may present a low cost and reasonable, though not optimal, solution for reducing energy theft. Yet, how to formulate the utility function of all players, including thieves, regulators and distributors, as well as potential strategies is still a challenging issue. classification-based approaches [40],[41],[42],[43],[44],[45],[46],[47] take advantage of the detailed energy consumption measurements collected from the AMI. Under normal condition customers’ consumption follow certain statistical pattern; irregularities in usage pattern can be a sign of some malicious activities. Data mining and machine learning techniques are used to train a classifier based on a sample database, which is then utilized to find abnormal patterns. Since these techniques take advantage of the readily available smart meter data, their costs are moderate. However, several shortcomings in existing classification-based schemes limit their DR and cause a high FPR.

One problem with utilization of machine learning classifiers in ETDS is data imbalance; i.e., the numbers of normal and abnormal samples are not in the same range. Benign samples are easily available using historic data. Theft samples, on the other hand, rarely or do not exist for a given customer. Besides, due to zero-day attacks in many cases examples of attack class cannot be obtained from the historic data. Lack of a thorough dataset of attack samples limits the DR. Furthermore, classification-based methods are vulnerable to contamination attacks, where by granular changes in data and polluting the dataset an adversary deceives the learning machine to accept a malicious pattern as a normal one.
Another challenging issue that affects the performance of classification-based methods is the fact that several non-malicious factors can alter the consumption pattern, e.g., change of residents, change of appliances, seasonality, etc. If such factors are not dealt with properly, they will result in a high FPR. As argued in [103], due to the base-rate fallacy phenomenon, the FPR is the limiting factor for the performance of an intrusion detection system. This is particularly true for ETDSs where false positives are very costly. Once an energy theft attack is detected, on-site inspection is required for final verification, which is an expensive procedure.

Finally, most existing classification-based methods require a high sampling rate to achieve an acceptable accuracy. However, an effective ETDS must not jeopardize customers’ privacy. As discussed in [104], while there are several techniques to mitigate privacy concerns such as power mixing and data aggregation, the privacy-by-design principle of data minimization is the most effective one. The required sampling rate of smart meters, which enables achievement of smart grid goals like load management and demand response, is still a controversial issue. The higher the sampling rate, the more is the risk of revealing customers’ private information. Thus, ETDSs that do not rely on a high data collection frequency are preferable.

Beside energy theft, increase in use of information technology for demand side management introduces new types of cyber-intrusions that aim to change the load curve in order to damage the power system equipment by causing circuit overflow or other malfunctions. These attacks work by tampering the pricing information or direct load control commands. We propose a detection mechanism for such attacks that work based on monitoring of the usage pattern of customers.

The main contributions of this chapter are:

- We design a novel algorithm for detecting energy theft attacks against AMI, CP-
BETD. CPBETD employs transformer meters along with monitoring of abnormalities in customers’ consumption patterns to provide a cost-effective and high performance solution for energy theft detection. Through application of appropriate clustering techniques and transformer meters, unlike existing classification based methods, CPBETD is robust against contamination attacks and non-malicious changes in consumption patterns, and therefore achieves a higher DR and a lower FPR.

- We address the problem of imbalanced data and zero-day attacks by generating a synthetic attack dataset, benefiting from the fact that theft patterns are predictable. Through extensive experiments we show that this significantly improves the DR and enables the detection of a wide range of attack types.

- We study the effect of sampling rate on detection performance, and show that compared to existing methods, CPBETD provides a higher performance with a lower sampling rate. Hence, it has a smaller effect on customers’ privacy.

- We argue that the metrics commonly used for ETDS performance evaluation in the literature are not adequate. While DR and in some cases FPR are used for performance evaluation, we suggest that Bayesian detection rate (BDR) and associated costs of the ETDS should also be considered to justify its application.

- We test the performance of CPBETD with real data from over 5000 customers. The dataset is publicly available and can be used as a benchmark for comparison among different energy theft detection methods. The best and most recent energy theft detection solutions existing in the literature are also tested for comparisons. The results prove the effectiveness of our approach.

- We introduce instantaneous anomaly detection IAD which is an algorithm for detecting attacks against direct and indirect load control by monitoring abnormalities
in consumption patterns in a neighborhood. Through simulations we show the effectiveness of the proposed algorithm.

The rest of this chapter is organized as follows. In Section 5.2, we survey the literature related to electricity theft detection in AMI. Threat models for electricity theft attacks are provided in Section 5.3. Section 5.4 describes the CPBETD algorithm. A theoretical analysis of CPBETD is provided in Section 5.5 and experimental result are provided in Section 5.6. In Section 5.7 we introduce IAD and we evaluate its performance in Section 5.8.

5.2 Related Work

In this section we review existing ETDSs in the literature, which employ consumption data of smart meters to find fraudulent customers. Monitoring of customers’ load profiles for signs of energy theft in traditional power systems had attracted the attention of researchers in the past. In [105], using historical consumption data, a data mining method along with SVM classifier were used to detect abnormal behaviors. The average daily consumptions of customers over a two year period were calculated and the long term trend in energy consumption was used to detect fraudulent customers. This method is only capable of detecting abrupt changes in load profile. Besides the detection delay is about two years. In [41], using six months usage reports, five attributes including average consumption, maximum consumption, standard deviation, sum of the inspection remarks and the average consumption of the neighborhood were chosen to create a general pattern of power consumption for each customer. K-means based fuzzy clustering was performed to group customers with similar profiles. A fuzzy classification was then performed and Euclidean distances to the cluster centers were measured. Customers with large distances to the cluster centers were considered potential fraudsters. Clustering the customers and relying
on long-term measurements limits the accuracy of this ETDS and causes long detection delay. Having more detailed metering information in AMI, CPBETD provides a much better performance with a much shorter delay.

A series of works [42–44] by Depuru et al. have studied the use of AMI data in classification-based ETDSs. In [42] SVM was used to approximate the consumption pattern of customers based on 96 readings of smart meters in AMI for each day. The classifier was trained using historical data of normal and theft samples. New samples were classified based on some rules as well as the SVM results. In [43] a neural network model was incorporated to estimate SVM parameters in order to reduce the training time of the classifier, and a data encoding method was proposed to improve the efficiency and speed of the classifier. Their method, however, is only effective in detecting energy theft attacks that result in zero usage reports, since in one step of the encoding procedure the metering data are converted into binary values. Therefore, the proposed classification technique cannot detect a wide range of attack types. In [44], an improved encoding technique was proposed and both SVM and a rule engine based algorithm were applied on encoded data to improve the classification accuracy. Portions of the algorithms were parallelized to reduce the detection time. Still the algorithm suffers from the common shortcomings of classification based methods as we have described in detail in the introduction Section. The database used for performance evaluation is not publicly available and the authors did not explain how they obtained theft samples for all customers or what percentages of the training and testing data were theft patterns. On the other hand, in this chapter we focus on addressing problems associated with classification-based techniques such as imbalanced data, contamination attacks and dynamic property of consumption patterns to improve the detection accuracy. We also argue that detection time might not be as crucial as other factors like FPR or implementation cost of the ETDS.
In [45], the problem of class imbalance for energy theft detection in traditional power systems was addressed. Considering that the number of people who commit fraud is much less than the number of honest customers, standard classifiers are overwhelmed by benign samples and tend to ignore the minority class. To achieve a higher performance, different classification techniques, including one-class SVM, optimum path forest and C4.5 decision tree, were combined in [45]. The combined classifier showed 2-10% improvements over individual classifiers. The shortcoming of this method is that while it imposes a high computational load, the performance improvement is not substantial. We address the problem of class imbalance by generating a dataset of malicious samples using the benign ones. As our experiments show, this method significantly improves the classification performance.

In [36], a multi-sensor energy theft detection framework for AMI (AMIDS) was presented. AMIDS collects evidences of malicious behavior from three types of information sources: 1) cyber-side network and host-based IDSs, 2) on-meter anti-tampering sensors, and 3) power measurement-based anomalous consumption detectors through nonintrusive load monitoring (NILM). These three types of information were combined to minimize FPR. While combining the information from different sources is effective in reducing FPR, the algorithms chosen for detecting anomalies have some drawbacks. Use of NILM, which requires a high sampling rate, reveals information about the types and time of use of appliances in customers’ premises. In this work we focus on detecting energy theft attempts only based on customers’ consumption patterns that can also be used as a part of a multi-sensor framework like AMIDS. Compared to the NILM-based technique, CPBETD provides a high performance with a much lower sampling rate.

Salinas et al. [46] introduced a privacy preserving ETDS that uses peer-to-peer (P2P) computing. A central meter is deployed in each neighborhood to measure the total elec-
electricity consumption at each time instance, which is assumed to be a linear combination of the energy consumption of all customers in the area. By solving a linear system of equations suspicious users are found. While the proposed method is valuable in that it is the only privacy preserving scheme so far, there are several limitations in this approach. Although the method is vulnerable to inaccuracies in technical loss (TL) calculation, this effect has not been studied. Besides, this method is only effective for detecting energy theft attacks with constant reduction rates, where the real consumption values are multiplied by a constant less than one for several consecutive readings. However, there are many other energy theft scenarios, such as sending random numbers as data. Our approach is capable of detecting more diverse attack types. CPBETD also employs central meters but in a completely different approach. Here we use distribution transformer meters to short list areas with high probability of theft, and to overcome the limitations of classification-based techniques.

Authors of [47] suggested modeling the probability distributions of the normal and malicious consumption patterns, and application of the generalized likelihood ratio (GLR) test to detect energy theft attacks. They used auto regressive moving average (ARMA) to model customers’ normal and malicious consumption distributions. They assumed that an attacker would choose a probability distribution that decreases the mean value of the real consumption. This, however, is not necessarily true with AMI. Considering the dynamic pricing in smart grids, by only changing the order of meter readings without altering the average, electricity theft is possible. Another major issue with ARMA-GLR detector is that it is only effective if the normal behavior and attack patterns can accurately be modeled by an ARMA process. Mashima et al. [47] also studied non-parametric algorithms such as exponentially weighted moving average, which provide a better precision when the model assumptions are inaccurate. They observed a better performance for ARMA compared to
other models. Our tests show that CPBETD significantly outperforms the ARMA-GLR detector.

5.3 Threat Model

The main objective of an energy theft attack is to pay less than the real value for the consumed energy. Energy theft attacks against AMI are categorized into three groups:

- Physical attacks: where illegal customers physically tamper with their meters to report a lower usage, for instance through application of a strong magnet to cause interference, or by reversing or disconnecting the meters. Customers might also directly wire high consuming appliances to an external feeder, bypassing the meters.

- Cyber attacks: can be used within the smart meters or over the communication link to the utility company. Examples include gaining privileged access to the meter firmware, tampering with the meter storage, interrupting measurements, and intercepting the communication link.

- Data attacks: target the metering values and are enabled through physical and cyber attacks.

Detailed explanation of attacks in each category are available in [36]. CPBETD is designed to detect abnormalities in the consumption patterns. Considering that the outcome of each attack is manipulation of the meter readings, CPBETD is effective in detecting all types of attacks presented above.
5.4 Consumption Pattern Based Electricity Theft Detection Algorithm

Among the three key elements of information security, i.e., confidentiality, integrity and availability, energy theft attacks target data integrity. The major cost of energy theft is financial loss to the utility company caused by unpaid usage. Therefore, while performance is important, the monetary cost introduced by the detection mechanism must be minimized. For the task of theft detection, false positives can be very expensive, since once a suspected fraudulent customer is detected, on-site inspection is required. On the other hand, factors such as detection delay might not be as critical, since the attacks barely introduce immediate damage and once an attack is detected the associated financial loss can be compensated through appropriate fines. In Section 5.5 we argue that the acceptable FPR depends on factors such as theft rate, which diverges over different regions; hence ETDSs with adjustable FPRs are preferable. CPBETD is designed based on the mentioned properties, and has two phases, i.e., training and application.

5.4.1 Training Phase

1) The algorithm is trained to estimate the TL in transmission lines within the NAN, $E_{TL}$. Several mechanisms exist for this purpose. As a case in point in [106] a method for precise calculation of TL in the branches of a distribution system was proposed. For each branch a specific circuit is assumed. Using the data of consumed energy and currents collected by smart or traditional power meters and by employing least square regression, the likely resistances at the lines connecting the consumption points to the distribution transformers as well as the non-ohmic losses are calculated. These parameters are then used to predict TLs in future time intervals. By comparing the total power loss with the estimated TL,
non-technical loss (NTL) in distribution transformer level is calculated. This method does not require frequent measurements. Besides, it is robust against falsified data caused by energy theft attempts during the application phase. Since if the metering data shows a lower usage compared to the true amount of consumption, the algorithm will calculate a lower TL. This increases the gap between the total amount of loss and TL which is the NTL. The only limiting factor in the proposed method is that the data used during the training phase must be genuine, otherwise the model will not be accurate. There are methods [107–109] that mostly rely on the physical characteristics of the distribution lines rather than customers data to calculate the TL in each segment; yet they are less accurate due to the dependency of such features on environmental conditions and inaccurate knowledge of the circuit elements. While in most works a high precision was reported, a thorough quantitative evaluation was missing. However, since the uncertainties in TL are usually less than the alterations in customers consumptions, a higher precision for calculating TL compared to NTL is expected. The average calculation error in [109] was reported to be around 0.5%. Design of the TL estimator is outside the scope of this work. We assume that $DR_{TM}$ and $FPR_{TM}$ are the DR and FPR, respectively, of the NTL detector.

2) The next step is data preprocessing, including operations like dimension reduction (the process of reducing the number of random variables) and normalization. Each data vector in the dataset includes the meter readings of a customer over a 24 hour period; for instance for $n$ measurements per hour the data vector has $24 \times n$ elements. While there are several methods for dimension reduction that minimize the information loss by extracting the important features of data, we sum up the in-between samples to make the algorithm compatible with different metering rates. That is, rather than applying a feature extraction technique that saves the key information of a higher dimension data vector while reducing the data size, we only add up the samples. This is because we assume that the
time interval of the data received from smart meters is not known.

3) Once the data are converted into the proper format, $k$-means algorithm [110] is performed on the benign dataset. Several non-malicious factors can alter the consumption pattern, such as seasonality, change of appliances, and different usage during weekdays and weekends. In order to have a better DR, $k$-means clustering with different values of $k$ is performed on the data, and each time the silhouette value of the clusters is calculated. A peak in the silhouette plot for $k=l$ shows that the data are originated from $l$ different distributions. Clusters that have few members are eliminated and will not be used for training the classifier. This can help to prevent pollution of the benign dataset by undetected attacks. We use $l'$ to denote the final number of clusters after eliminating the small groups.

4) The next step is preparing a dataset for training the classifier. While a dataset of benign samples for each customer is easily obtainable using historic data, malicious samples might not be available, since energy theft might never or rarely happen for a given customer. In order to address the problem of imbalanced data, one solution is application of single-class classification techniques where the classifier is trained only using normal samples. However, as shown in Section 5.6 for the present application the performance of one-class classifier is poor. Another solution is to utilize density function approximation methods as in [111]. However, such approaches are only effective if the data can accurately be modeled by the approximation function, which in practice might be difficult to attain. Instead, we propose to create a dataset of malicious samples using the benign samples. The goal of energy theft is to report a lower consumption than the true amount used by consumer, or to shift the high usage to low tariff periods. Thus, it is possible to generate malicious samples using the benign ones.

Assuming that $\mathbf{x} = \{x_1, \ldots, x_n\}$ is a vector of true consumption values for a 24 hour period
with \( n \) samples, the utility will receive \( y = \{y_1, ..., y_n\} \) as the meter readings. For honest customers \( y = x \) while for fraudulent users \( y = h(x) \), so that \( \sum_{i=1}^{n} y_i \leq \sum_{i=1}^{n} x_i \). Through studying different scenarios for electricity theft and their effects on measured values, \( h(.) \) is extractable. For instance \( h(x) = \alpha x \), \( 0 \leq \alpha < 1 \) is one possibility. Therefore, it is practical to generate malicious samples using the benign dataset. Although defining all possible functions that meet this condition is not practical, through considering a variety of scenarios and taking advantage of generalization property of SVM, a thorough dataset of attack samples can be generated.

5) The next step is training the classifier. Among several existing techniques for regression and classification, we chose SVM due to its superior performance in many applications compared to traditional methods like likelihood ratio test and neural networks. Furthermore, successful utilization of SVM for problems that deal with the similar data type [105] motivates us to use this classifier. Here, the classifier has \( l' + 1 \) classes, \( l' \) for benign and 1 for malicious samples.

6) SVM parameters can be adjusted to attain different performance in terms of DR/FPR. However, as shown in Section 5.6, the ROC of an SVM by itself does not provide a wide range for DR/FPR. In order to achieve the appropriate performance, first the required FPR, \( FPR_{req} \), is calculated as described in Section 5.5. Let \( DR_{SVM} \) and \( FPR_{SVM} \) denote the best achievable performance by SVM, parameter \( m \) is calculated as follows:

\[
m = -\log\left(\frac{FPR_{req}}{FPR_{TM} \times FPR_{SVM}}\right)
\]

(5.1)

where \( m \) represents the number of times an abnormal pattern must be detected before an energy theft alarm is raised. This mechanism, however, will also reduce the DR:

\[
DR = (DR_{TM} \times DR_{SVM})^m \leq DR_{TM} \times DR_{SVM}
\]

(5.2)
The larger is the value of $m$, the smaller is the FPR, yet it also means a larger detection delay.

### 5.4.2 Application Phase

1) For each neighborhood one or more transformer meters measure the total electricity provided to the customers in the area, $E_{TM}(t)$. This value is compared with the total amount of consumption reported by the smart meters of the corresponding distribution transformer, $\sum E_{SMi}(t)$. NTL is reported if for any time, $t$, during a day, $E_{TM}(t) > \sum E_{SMi}(t) + E_{TL}(t) + \varepsilon$, where $\varepsilon$ is the calculation error for TL. This parameter adjusts the trade off between $DR_{TM}$ and $FPR_{TM}$, and therefore the DR and FPR of the whole algorithm. In CPBETD attacks are detectable if the amount of stolen electricity in the area is greater than $\varepsilon$. The above test is performed each time new samples are collected.

2) Each new sample is preprocessed and converted into the proper format consistent with the training set.

3) SVM is applied to a new sample to determine whether it belongs to the benign or attack class.

4) If Step 1 did not detect an anomaly and the new sample was classified as benign by the SVM, the new sample is added to the benign dataset and the corresponding attack patterns will be generated and added to the attack dataset.

5) If NTL was detected in Step 1 and the classifier recognized an attack, a suspicious behavior of the smart meter is reported. An energy theft is detected when the suspicious behavior of the smart meter is repeated $m$ times over a certain period. During this time new samples are stored in a temporary database. Once an energy theft is detected, an appropriate action such as on-site inspection is exercised. Based on the amount of NTL calculated in Step 1, higher priority for inspection is assigned to smart meters located in...
areas with larger NTL. If a theft is verified, samples in the temporary database are added to the attack dataset. Otherwise, they will be added to the benign dataset and their corresponding attack patterns to the attack dataset.

6) Another possibility is when Step 1 does not detect an NTL, but SVM recognizes an anomaly. This condition might have three causes. It might be due to SVM misclassification, or because of error in NTL calculation in Step 1. These cases are not expected to happen frequently in consecutive days. On the other hand, the condition might be because of some alterations in consumption habits, for instance due to changes of residents or appliances, etc., which result in major changes of usage pattern. In this case the condition will persist. Therefore, when SVM detects an anomaly while there is no sign of NTL in Step 1, the new sample is stored in a temporary database. If in the following days this condition happens frequently, the old dataset will be discarded and a new dataset based on the samples in the temporary database will be generated. Once the dataset is large enough, the classifier is retrained. A credibility factor, \( cf_i \), is assigned to each smart meter, which is a binary variable initialized to one. When a non-malicious anomaly is detected, as described above, \( cf_i \) is set to zero and when the situation is resolved it will be set back to one. When the algorithm detects an energy theft, smart meters with \( cf_i = 1 \) have a higher priority for further action. This step of the algorithm makes CPBETD robust against non-malicious changes in consumption pattern.

7) If NTL was detected in Step 1, but SVM did not recognize any anomaly and the condition persists, it shows that an attack might be happening, but SVM cannot identify it. In this case the benign dataset of the customers is analysed for sign of data contamination attack, in which by gradual changes in data and polluting the dataset an adversary deceives the learning machine to accept a malicious pattern as a normal one. The long-term trend in daily usage of the customer is studied. A descending slope in long-term consumption
Chapter 5. Detection of Malicious Activities in AMI Using Customers Consumption Patterns

curve can be a sign of contamination attack. If analysis of historic data does not show a contamination attack, CPBETD raises an alarm to indicate that an attack might be happening but the algorithm cannot detect it, and the algorithm continues its normal operations for new samples. This happens in rare scenarios, e.g., when a new load with high consumption is directly connected to a feeder. This step of the algorithm makes CPBETD robust against contamination attacks.

Pseudo-codes of the algorithm for application phase are provided in Fig. 5.1.

5.5 Performance Analysis

5.5.1 Required Performance

As explained in [40], the revenue of a distribution utility is formulated as follows:

\[
R(e, D) = \sum_i T(q^i - q^i_u) + \sum_i \rho(e, q^i_u, D)F(q^i_u)
\]  

(5.3)

where \(q^i\) is the total consumption of user \(i\), \(q^i_u\) is the unbilled part of the usage, \(T\) is the price of electricity, \(\rho\) denotes the probability of detecting an electricity theft, \(e\) is the effort invested in anti-fraud technologies, \(D\) is the anomaly detection test, and \(F(\cdot)\) represents the recovered fines from the detected theft. On the other hand, the associated costs include the investments for protecting against theft, \(\Psi(e)\), and the costs of producing enough electricity to meet the demand of all customers \(C(\cdot)\). Therefore, the total profit of the utility is:

\[
R(e, D) - C(q^i) - \Psi(e).
\]

(5.4)

The operational cost of handling \(D\) is proportional to the effort, \(e\), of the utility to manage false alarms. At the same time, increasing \(e\) improves the probability of detecting a theft.
Chapter 5. Detection of Malicious Activities in AMI Using Customers Consumption Patterns

Input: $NS$ (new sample), $threshold_1$, $threshold_2$, $threshold_3 ( > threshold_2)$
Output: $attack$ (boolean)
Variables: $counter_1$ (number of times an attack is detected), $counter_2$ (number of non-malicious anomalies), $NTL$ (boolean), $CF_i$ (binary), $SVM_{out}$ (boolean), $TDB_1$ (temporary database), $TDB_2$ (temporary database)

\[
\text{if } E_{TM} > \sum_i E_{SM_i} + E_{TL} + \epsilon \\
\text{NTL}=true;
\text{else } \\
\text{NTL}=false;
\text{end if;}
\text{classify } NS \text{ by SVM, if it is classified as attack, then } SVM_{out}=true, \text{ otherwise } SVM_{out}=false; \text{if } NTL=true \\
\text{if } SVM_{out}=true \\
\text{if } counter_1 \geq threshold_1 \\
\text{attack}=true; \\
\text{add } TDB_1 \text{ to the attack dataset; } \\
\text{go to end; } \\
\text{else } \\
\text{counter}_1++; \\
\text{add } NS \text{ to } TDB_1; \\
\text{end if} \\
\text{else if } SVM_{out}=false \\
\text{Raise an alarm informing about the risk of an undetectable attack; } \\
\text{end if} \\
\text{else if } NTL = false \\
\text{if } SVM_{out}=true \\
\text{if } counter_2 < threshold_2 \\
\text{counter}_2++; \\
\text{add } NS \text{ to } TDB_2; \\
\text{else if } threshold_2 \leq counter_2 \leq threshold_3 \\
\text{counter}_2++; \\
\text{CF}_i = 0; \\
\text{add } NS \text{ to } TDB_2 \\
\text{else if } counter_2 > threshold_3 \\
\text{discard the old dataset; } \\
\text{generate a new dataset using } TDB_2; \\
\text{retrain the SVM; } \\
\text{CF}_i = 1; \\
\text{end if}; \\
\text{else if } SVM_{out}=false \\
\text{attack}=false; \\
\text{end if}; \\
\text{end if;}
\]

Figure 5.1: Pseudo-codes of application phase

Therefore, the optimal $D$ is the one with maximum $\rho$ subject to an upper bound on FPR. The required FPR might vary over different regions due to the base-rate fallacy phenomenon. The Bayesian detection rate (BDR) is defined as:

\[
P(I|A) = \frac{P(I) \times DR}{P(I) \times DR + P(\bar{I}) \times FPR} \quad (5.5)
\]
where \( I \) stands for intrusion and \( A \) for alarm. For IDS alarms to be reliable, this probability must be high. As (5.5) suggests, BDR not only depends on DR and FPR, but also is affected by the probability of occurrence of an intrusion. When this probability is low, no matter how big the DR is, the denominator will be subjugated by the FPR term. Therefore, in order to achieve a reasonable value for BDR, very low FPR is required. In the case of energy theft \( P(I) \) is usually a small value and might significantly vary in different areas.

### 5.5.2 Classification Method

In SVM, a model is generated based on the training data and is used to predict the target values of the test data. First, using a kernel function, the data vector is mapped to a higher dimensional space where data classes are more distinguishable. Then a separating hyper-plane with maximum margin to the closest data points in each class is found. This concept translates into a convex quadratic optimization problem formulated in (5.6).

\[
\begin{align*}
\min & \quad \frac{1}{2} \|w\| + C \sum_i e_i \\
\text{such that} & \quad y_i (w^T \phi(x_i) + b) \geq 1 - e_i \\
\phi(x_i) &= e^{-\gamma \|x\|^2} 
\end{align*}
\]  

(5.6)

where \( x_i \) is the \( i^{th} \) input data, \( y_i \) is the class label for \( x_i \), \( w \) is the normal vector to the hyper-plane, \( C \) is the penalty for the error term, \( e_i \) measures the constraint violation, \( \phi \) is the kernel function, and \( b \) determines the offset of the hyper-plane from the origin along the normal vector [112].
5.5.3 Clustering Method

Silhouette plots are applied to determine the number of clusters within a dataset. Assume that the data have been clustered into $k$ clusters and for each sample $i$, $a(i)$ is the average dissimilarity of $i$ with other samples within the same cluster. Also $b(i)$ is the least average dissimilarity of $i$ to any other clusters. The Silhouette value, $s(i)$ is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

(5.7)

The average of $s(i)$ over all samples within a cluster shows how close the samples in the cluster are, and averaging over the entire dataset shows how properly the data have been clustered. We use this method to determine the number of clusters in the dataset of each customer. Defining separate classes for distinct clusters can help to achieve a higher classification accuracy.

5.6 Performance Evaluations

We used the smart energy data from the Irish Smart Energy Trial [113] in our tests. The dataset was released by Electric Ireland and Sustainable Energy Authority of Ireland (SEAI) in Jan 2012; it includes half hourly electricity usage reports of over 5000 Irish homes and businesses during 2009 and 2010. Customers who participated in the trial had a smart meter installed in their homes and agreed to take part in the research. Therefore, it is a reasonable assumption that all samples belong to honest users. The large number and variety of customers, long period of measurements and availability to the public make this dataset an excellent source for research in the area of analysis of smart meters data. For each customer there is a file containing half hourly metering reports for a 535 day period. We reduced the sampling rate to one per hour and for each customer divided the
file into a dataset of 535 vectors, each with 24 components. Then based on the dataset of benign samples, for each sample $x = \{x_1, ..., x_{24}\}$, we generated 6 types of malicious samples as follows:

for $t=1,...,24$

1) $h_1(x_t) = \alpha x_t$, $\alpha = \text{random}(0.1,0.8)$

2) $h_2(x_t) = \gamma_t x_t$, $\gamma_t = \text{random}(0.1,0.8)$

3) $h_3(x_t) = \beta_t x_t$

$$\beta_t = \begin{cases} 0 & \text{start\_time} < t < \text{end\_time} \\ 1 & \text{else} \end{cases}$$

$\text{start\_time} = \text{random}(0,23 - \text{minimum\_off\_time})$

$\text{duration} = \text{random}(\text{minimum\_off\_time},24)$

$\text{end\_time} = \text{start\_time} + \text{duration}$

here $\text{minimum\_off\_time}=4$;

4) $h_4(x_t) = \gamma_t \text{mean}(x)$, $\gamma_t = \text{random}(0.1,0.8)$

5) $h_5(x_t) = \text{mean}(x)$

6) $h_6(x_t) = x_{24-t}$

We have defined functions $h_1(\cdot),...,h_6(\cdot)$ based on existing and possible electricity theft scenarios. Functions $h_1(\cdot)$ and $h_2(\cdot)$ represent scenarios where a fraction of the customer usage is reported. $h_1(\cdot)$ multiplies all the samples by the same randomly chosen value, while $h_2(\cdot)$ multiplies each meter reading by a different random number. We have used a uniform random number generator in our simulations. $h_3(\cdot)$ formulates a common electricity theft method in which the smart meter does not send its measurements or sends zero for some duration defined by $\text{minimum\_off\_time}$ in the formula. $h_4(\cdot)$ and $h_5(\cdot)$ orderly report a factor and the exact value of the average of readings during the day. $h_6(\cdot)$ reverses the
order of readings. \(h_5(\cdot)\) and \(h_6(\cdot)\) represent attacks against indirect load control mechanisms in which the price of electricity varies during different hours of the day; while the total amount of electricity usage stays the same, the usage is reported to happen during the low tariff periods. \(h_1(\cdot), ..., h_6(\cdot)\) formulate some examples of electricity theft attacks that we have chosen based on the existing reported attacks and possible attack scenarios as explained above to evaluate the performance of our method. In practice the relationship between theft and benign samples can be found and formulated by analyzing the existing attack samples in historic database of fraudulent customers, and then be used to generate the attack dataset for all customers. Also benefiting from the generalization property of SVM, the algorithm is able to detect attacks for which the exact function in not defined in generating the training dataset. We study this effect in Section 5.6.3. An example of the daily consumption of a customer is shown in Fig. 5.2. Fig. 5.3 shows the corresponding attack patterns. Experiment results in the rest of this section are for the anomaly detector part of CPBETD. The total DR and FPR of the algorithm are scaled by the \(DR_{TM}\) and \(FPR_{TM}\).
5.6.1 Experiment 1: One-class Support Vector Machine

In the first test we examine the performance of one-class SVM, in which the classifier is trained only using normal samples. Among 535 samples of the benign dataset we use 458 samples for training and 77 samples for testing. In each 7 consecutive days one sample is randomly chosen for the testing set and the other 6 for the training set. Also for each attack type, 535 samples are generated. In total, testing set includes 3287 samples, scaled in the range [-1,1]. We choose radial basis function (RBF) as the SVM kernel, and use grid search [112] to find the best values for the kernel parameter, $\gamma$, and the upper bound in the fraction of training points, $\nu$. Other SVM parameters are $C=50$, and $e=0.1$. The above procedure is repeated for 5000 customers. For each customer the best performance is considered the one that provides the highest difference (HD) between DR and FPR for different combinations of $\gamma$ and $\nu$. The average value of HD for 5000 customers is found to be 47%, with the average DR and FPR being 76% and 29%, respectively. Fig. 5.4 shows the ROC curve for three customers with best, average and worst performances. As the
Chapter 5. Detection of Malicious Activities in AMI Using Customers Consumption Patterns

5.6.2 Experiment 2: Multi-class Support Vector Machine

In the second experiment, we train the classifier using both benign and malicious samples. After preprocessing we employ $k$-means clustering on the benign dataset and study the silhouette plots to find the best value for $k$. For most customers we observe that $k=1$ or $2$ provides the best result. Then we train a multi-class SVM with $k+1$ classes. The silhouette plots for 2 customers with $k=1$ and $k=2$ are shown in Fig. 5.5. Originally, the training set includes 458 benign and 2748 malicious samples. We use over-sampling, in which the members of the minority class are replicated, to make the number of benign and attack samples equal. This helps to avoid the classifier from being biased toward the larger class. We use RBF kernel and apply grid search to find the appropriate values for $\gamma$ and $C$; the value of $e$ is 0.1. For this experiment we observe the average HD of 83% with the average DR and FPR being 94% and 11%, respectively. Fig. 5.6 shows the ROC

Figure 5.4: ROC curves for three customers with the worst, average and the best detection performance using one-class SVM.

results suggest without abnormal samples in training phase, classification performance is not promising.
Chapter 5. Detection of Malicious Activities in AMI Using Customers Consumption Patterns

Figure 5.5: Silhouette plots for customers with $k=1$ and $k=2$.

Figure 5.6: ROC curves for three customers with the worst, average, and the best detection performance using multi-class SVM.

curve of three customers with best, intermediate and worst performances. As we can see, multi-class SVM provides a significantly better performance compared to single class SVM.
Chapter 5. Detection of Malicious Activities in AMI Using Customers Consumption Patterns

Figure 5.7: ROC curves for three customers with the worst, average and the best detection performance using multi-class SVM with undefined attacks.

<table>
<thead>
<tr>
<th>Performance</th>
<th>Exp. 1</th>
<th>Exp. 2</th>
<th>Exp. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD(%)</td>
<td>47</td>
<td>83</td>
<td>70</td>
</tr>
<tr>
<td>DR(%)</td>
<td>76</td>
<td>94</td>
<td>86</td>
</tr>
<tr>
<td>FPR(%)</td>
<td>29</td>
<td>11</td>
<td>16</td>
</tr>
</tbody>
</table>

5.6.3 Experiment 3: Multi-class Support Vector Machine with New Attacks

An attacker might use a different $h(\cdot)$ than those used to train the classifier. To study this effect, in this experiment we only train the SVM using $h_3(\cdot)$, yet we include all $h_i(\cdot)$s in the test set. The SVM parameters are similar to experiment 2. We observe the average performance of HD = 70% with DR = 86% and FPR = 16%. Three examples of ROC curves are shown in Fig. 5.7. Though in this experiment we observe a lower DR compared to Experiment 2, the performance is still significantly better than one class SVM. A summary of the experiment results is provided in Table 5.1.
5.6.4 Overall Performance of Consumption Pattern Based Electricity Theft Detection algorithm

While the exact value for the probability of an energy theft is unknown and varies across different areas, we use the result of an experiment in [114] to evaluate the performance of CPBETD. In 2001, the Arizona public service company conducted a study to provide an accurate estimate of energy theft in its coverage area. The goal of the study was to find the extent of meter tampering and the resulting financial loss with 95% accuracy. Among 868,000 customers, 550 were randomly selected including rural (65%) and urban, residential (88%) and industrial users. The results showed the definite meter tampering rate of 0.72%.

Considering the tampering rate of 0.72%, Figure 5.8 shows the relationship between BDR and FPR. As the figure suggests, in order to achieve a 90% BDR even for 100% DR, FPR must be around 0.1%. Table 5.2 shows the DR, FPR and BDR for different values of $m$.

By increasing $m$, appropriate values for FPR and therefore BDR are achievable, while DR stays in an acceptable range. Considering the ROC curves of the SVM, obtaining such a low FPR for $m=1$ is either impossible or happens for a very low DR.
Table 5.2: BDR for different values of \( m \)

<table>
<thead>
<tr>
<th>( m )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR(%)</td>
<td>94</td>
<td>88</td>
<td>83</td>
<td>78</td>
</tr>
<tr>
<td>FPR(%)</td>
<td>11</td>
<td>1</td>
<td>0.1</td>
<td>0.01</td>
</tr>
<tr>
<td>BDR(%)</td>
<td>6.3</td>
<td>38</td>
<td>85</td>
<td>98</td>
</tr>
</tbody>
</table>

5.6.5 Effect of Sampling Rate on Performance

To study the effect of sampling rate on detection performance, we tested the average performance of 100 customers for different metering frequencies. Results are shown in Table 5.3. We observe a peak at 12 samples/day, which indicates that higher sampling rates do not necessarily result in a better performance. This can be due to the fact that for lower sampling rates, the effect of theft on each sample might be more apparent. For instance, when a fraudulent customer aims to report 5KW less usage than his actual consumption by uniformly dividing this amount over all samples, for 24 samples the changes in usage pattern is less apparent than when 5KW is spread over 12 samples. Therefore, the attack might be more apparent for lower sampling rates. On the other hand, for very low sampling rates the consumption pattern cannot be modeled accurately. Using the lowest sampling rate that provides an acceptable performance is important, not only to minimize the required resources, but also to preserve customers privacy. In [115] the authors showed that for 0.5 samples/second, one can find as much detailed information as the title of the movie or the TV channel displayed in the household. In [36], using 0.5 samples/minute the authors where able to find the type and time of usage of appliances inside the household. These information cannot be obtained precisely using 1 sample/hour. For 1 sample/day, one cannot find detailed information about consumption habits of a user, still it is possible to conclude whether or not the customers are at home. This information can be used for example for the purpose of committing burglary. On the other hand, the utility requires metering data for demand response management. In [40] a game theory
Table 5.3: Effect of sampling rate on detection performance

<table>
<thead>
<tr>
<th>Sampling Rate(samples/day)</th>
<th>DR(%)</th>
<th>FPR(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62.1</td>
<td>28.0</td>
</tr>
<tr>
<td>2</td>
<td>76.5</td>
<td>16.0</td>
</tr>
<tr>
<td>4</td>
<td>88.7</td>
<td>9.0</td>
</tr>
<tr>
<td>6</td>
<td>90.2</td>
<td>9.7</td>
</tr>
<tr>
<td>12</td>
<td>93.1</td>
<td>11.0</td>
</tr>
<tr>
<td>24</td>
<td>89.8</td>
<td>12.3</td>
</tr>
</tbody>
</table>

approach was proposed to formulate the privacy-preserving demand-response problem as the task of finding the maximum allowable sampling rate that keeps the demand lower than a predefined maximum value. We cannot say there is a specific point at which the users’ measurement is or is not private anymore. However, the lower the sampling rate the less information can be extracted from the measured data. So the sampling rate that allows the utility to achieve its goals regarding demand response etc. can be found and the utilities can be forced by law to not collect metering data with higher than the required frequency. One advantage of CPBETD is that it can provide a good performance even for the sampling rate as low as 4 samples/day.

5.6.6 Discussion and Comparison

Customers’ normal consumption pattern may vary due to several non-malicious factors. These changes can be temporary, periodic or permanent and potentially can cause false positives. CPBETS uses mechanisms to handle these situations. Short-term changes are the result of some unusual behaviors, for instance having a big party, which do not last for more than one or few days. Application of distribution transformer meters help to reduce false alarms due to short-term changes since a suspicious behavior is reported only when both distribution transformer meters and SVM detect an anomaly. Therefore, if a customer has a benign unusual behavior while NTL is not detected in distribution transformer level,
false positive will not be generated, unless another customer within the neighborhood area is committing fraud at the same time which results in detection of NTL in the distribution transformer. In this case the customer is considered suspicious. To handle this situation we suggested calculation of the required FPR and adjustment of parameter $m$ so that theft will only be reported when a suspicious behavior is repeated several times. Moreover, customers might have dissimilar usage habits over different days of week (weekday - weekend) or seasons. Using k-means clustering enables the algorithm to separate distinct distributions in data, and separate classifiers can be trained accordingly. If a time dependency is observed for clusters, the corresponding classifiers are labeled, and based on the time of new samples they are classified using the appropriate classifier. Finally, permanent changes can happen due to change of residents, appliances, etc. As explained in Step 6 of the application phase, CPBETD can identify these changes and avoid generating false positives.

It is also possible that the dataset that is initially used to train the algorithm contains theft samples. One solution is harder supervision during the data collection period. Using the NAN NTL detector, areas suspicious to energy theft can be short listed and customers in the area can be considered for on-site inspection. During this phase application of TL calculation methods that do not rely on customers data is recommended.

One limitation of machine learning methods for anomaly detection in the security domain is vulnerability to contamination attacks. Concurrent use of NAN NTL detector along with the customers’ anomaly detector in CPBETD helps to mitigate this concern. When an NTL is detected in NAN, while there is no sign of anomaly in customers’ usage, the trend in daily usage of customers can be monitored. A descending slope in long-term consumption curve can be an indication of a contamination attack. CPBETD is capable of detecting a variety of attack types. The only energy theft scenario that cannot be detected is when a customer employs a new appliance and directly connects it to an outside feeder.
Table 5.4: Comparison among energy theft detection methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DR(%)</td>
<td>67</td>
<td>94</td>
<td>NA</td>
<td>attack1: 96 others: NA</td>
</tr>
<tr>
<td>FPR%</td>
<td>28</td>
<td>11</td>
<td>NA</td>
<td>attack1: 9 others: NA</td>
</tr>
<tr>
<td>Required sampling rate</td>
<td>low</td>
<td>low</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Robustness against non-malicious changes</td>
<td>vulnerable</td>
<td>robust</td>
<td>vulnerable</td>
<td>robust</td>
</tr>
<tr>
<td>Privacy preservation</td>
<td>medium</td>
<td>medium</td>
<td>week</td>
<td>strong</td>
</tr>
</tbody>
</table>

Since there is no sign of this appliance in the historic dataset, the anomaly will not be detected.

Table 5.4 shows a comparison among the proposed algorithm and the most recent and best ETDSs existing in the literature. We implemented the ARMA-GLR detector proposed in [47], and tested its performance using the SEAI dataset. The results are provided in Table 5.4. The NILM based technique in AMIDS [36] requires around 0.5 samples/minute to function properly, and thus the 2 samples/hour sampling rate in the SEAI dataset is inadequate to measure its performance. We also implemented the LUD algorithm in [46], which is based on P2P computing. While the algorithm is effective in detecting attack1 in which all usage measurements are multiplies by the same coefficient \( h_1(t) \), it performed poorly for other attack types.

5.7 Instantaneous Anomaly Detection Algorithm

Although monitoring the usage pattern of each customer over a long period is useful in detecting malicious activities with long-term effects on a specific customer, such as en-
ergy theft attempts, it is not suitable for attacks that cause immediate damage. Besides, some attacks target several customers simultaneously with minimum influence on each separately; such attacks are not detectable by monitoring the usage pattern of customers independently. To address these issues, we introduce instantaneous anomaly detection (IAD). IAD leverages the fact that in a given time, the consumption of all customers in a NAN follow a certain pattern. In other words, the approximation function for IAD in a given time is a function of price and meter reading of all or a group of HANs within the NAN.

A database of normal and malicious patterns is generated using the most recent samples. Each vector in the database consists of the meter reading of all smart meters in a NAN for a given time, along with the electricity price and time of measurement. Using the database, an SVM classifier is trained to distinguish between normal and malicious patterns. Every time new metering data are received from the smart meters, a new sample vector is created and classified. Since a NAN might include hundreds to thousands of HANs, either the HANs should be sub-grouped, or before classification, the sample vector should be preprocessed and its dimension should be reduced using an appropriate feature extraction method. The classifier then decides whether the sample belongs to normal or malicious classes. Normal patterns are added to the dataset of normal samples. Anomalous patterns trigger an alarm and after final decision, are added to the database of normal or malicious patterns.

The detection delay of the proposed method is close to the time interval between the smart meter report transmissions. Using IAD, attacks which are distributed over a group of customers are detectable. Furthermore, IAD is more robust against false positives resulting from benign changes in load profile. While it is possible that the consumption pattern of a specific customer changes due to some unusual condition such as having a big party, it
is less likely that all or a large group of HANs within a NAN experience none-malicious changes simultaneously.

5.8 Performance Evaluation of Instantaneous Anomaly Detection

5.8.1 Dataset

In order to evaluate the effectiveness of IAD against adversarial activities such as attacks against direct and indirect load control we provided a synthetic dataset as follows.

Dataset of Benign Patterns

We generated the probability distribution functions (PDF) of electricity usage for typical households as a function of time and electricity price. According to [116] among 33 appliances reported in 2001 Residential Energy Consumption Survey (RECS), done by U.S. Department of Energy (DoE), only 17 appliances use 85% of the total household consumption. Therefore, as [116] we assumed that these 17 appliances constitute all of the electric devices in a household. We used the same energy profile for appliances as in [116] which is shown in Table 5.5. In our study, the dataset is based on the consumption pattern of ten residential customers. Through careful study and interview with one resident from each dwelling, we calculated the consumption distribution of each customer separately. The interviewees were PhD students with major of electrical engineering; they had a deep knowledge in statistics and probability science, and were well aware of the purpose of the study. For each appliance we created a PDF as a function of time and price. For each device we asked the interviewee to determine the probability of using the device during different hours in a working day (Monday-Friday) \((P_i(t), i=1,\ldots,17)\). We also asked the
Interviewee about the probability of using a specific device knowing the price of the electricity ($P_i(pr), i=1,\ldots,17$). We defined 3 levels for electricity price: low, medium, and high. Then, for each device we calculated its PDF as in (5.8).

$$P_i(t, pr) = P_i(t)P_i(pr), i = 1,\ldots,17$$  \hspace{1cm} (5.8)

Where $P_i(t, pr)$ is the consumption probability of device $i$ at time $t$ with price $pr$. Based on the distribution functions, we produced a dataset of normal consumption patterns. For each appliance per hour a random number, $rand$, was generated. We employed ”Random” class in Java as the pseudo-random number generator. Using the random number, the
appliance status, $S_i(t, pr)$, was set to a binary value of ON or OFF according to (5.9).

$$S_i(t, pr) = \begin{cases} 
on \text{rand} \leq P_i(t, pr) 
, i = 1, ..., 17 
\text{off} \text{rand} > P_i(t, pr) 
\end{cases}$$ (5.9)

In the next step, the consumption function, $F_i(t, pr)$, of each appliance was found based on the appliance status, and its profile.

$$F_i(t, pr) = \begin{cases} 
\text{Power}_i \quad S_i(t, pr) = \text{on} 
0 \quad S_i(t, pr) = \text{off} 
, i = 1, ..., 17 
\end{cases}$$ (5.10)

In (5.10), $\text{Power}_i$ is the power consumption of device $i$ according to its energy profile. Finally we calculated the total power function, $E(t, pr)$, as summation of the consumption function of all appliances.

$$E(t, pr) = \sum_{i=1}^{17} F_i(t, pr)$$ (5.11)

Using the above method we produced a dataset of normal consumption patterns, including 800 samples for each customer.

**Dataset of Malicious Patterns**

Attacks against direct and indirect load control (DLC/IDLC) constitute some of the major cyber security threats against AMI.

Attacks against indirect load control (ILC): Load control mechanisms aim to modify the consumption pattern of customers. One common method of load controlling is ILC in which customers are motivated to change their load curves by dynamic pricing. Price information is transmitted to the smart meters through AMI. Either customers use the pricing information to adjust their usage manually, or automatic energy consumption scheduling
(ECS) units within the HANs receive the pricing signal. Based on the customer priorities and pricing information, ECS controls the operation of electrical devices. By injecting false prices, an attacker can affect the usage profile of the customers. We simulated this condition through type 1 attacks as explained below.

- **Type 1**: We exchanged the probability of consumption over low and high price tariffs, which has the same effect as sending a low price signal when the real price is high and vice versa.

Attacks against direct load control (DLC): DLC is another load control mechanism in which some of the customers’ loads are directly controlled by the utility. Automated DLC systems, send control signals such as turn on/off, through AMI. By comprising the DLC, an attacker can affect the load curve. For instance, by injecting false ”turn on” signal to a large number of appliances, the adversary can cause a large sudden increase in total load, which can affect the power quality and damage the customer and utility’s equipment. DLC attacks are simulated through type 1 and type 2 attacks.

- **Type 2**: We added six mega Watt extra power to the normal consumption during random hours and for random durations between 15 minutes to 3 hours.

### 5.8.2 Test Results

We considered a NAN including 20 HANs. The dataset contained 400 samples for each of the benign, Type 1 and Type 2 attack classes. For Type 1 we assumed that all 20 HANs were under attack; since in order to cause immediate damage, this attack needs to work in large scale. 400 samples of Type 2 attacks equally contained samples where 25%, 50%, 75% and 100% of the HANs were under attack. Components of the sample vectors were the power consumption of 20 HANs at 6pm. We also assumed that at 6pm,
Table 5.6: Cross classification and detection performance of IAD

<table>
<thead>
<tr>
<th>Class</th>
<th>Nor. (%)</th>
<th>T. 3 (%)</th>
<th>T. 4 (%)</th>
<th>Acc. (%)</th>
<th>FP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nor.</td>
<td>85.00</td>
<td>14.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T. 1</td>
<td>14.50</td>
<td>85.50</td>
<td>0.00</td>
<td>85.75</td>
<td>14.00</td>
</tr>
<tr>
<td>T. 2</td>
<td>3.00</td>
<td>0.50</td>
<td>96.50</td>
<td>98.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

the price of electricity is high. We used 4-fold cross validation for training and testing to avoid overfitting of the classifier. The datasets were divided into 4 equal size subsets. Every time one subset was tested while the 3 remaining subsets were used to train the classifier. Parameters of the SVM were: $C=5$, $\epsilon=0.05$ and $\gamma=0.3$. Classification results are shown in Table 5.6. We observed satisfactory detection performance. However, as it can be seen from the table, IAD provides a lower detection performance for Type 1 attacks. The reason is, in simulating the effect of false prices on consumption pattern, we used a probabilistic approach in which a higher consumption probability was assigned to high tariff periods and vice versa. Therefore, there might be some samples in dataset of Type 1 attacks that are very similar to the benign class. In future smart grid, where energy management modules are installed to automatically control the household appliances based on the energy price, and even some devices are under direct control of the utility, the pattern of normal consumption regarding the pricing information will be more predictable and distinguishable.

To support the idea that under direct or automatic load control the performance of the detection method will improve, we studied the effect of the increase in the difference between probability of usage for low and high prices. Fig. 5.9 shows the accuracy of IAD in detection of Type 1 attacks vs. the probability difference. We observed that when the difference is increased the detection accuracy is improved.
5.9 Conclusion

In this chapter we have introduced CPBETD, a new algorithm for detecting energy theft in AMI. CPBETD relies on the predictability of customers’ normal and malicious usage patterns. Along with application of SVM anomaly detector, the algorithm uses silhouette plots to identify the different distributions in the dataset, and relies on distribution transformer meters to detect NTL at the transformer level. We have shown that these features provide a high performance and make the algorithm robust against non-malicious changes in consumption pattern as well as data contamination attacks. In practice the required performance for an ETDS may vary across different regions. We have shown that by introducing some delay to the detection algorithm, an adjustable performance to match different objectives is achievable. Using extensive experiments on a real dataset of 5000 customers, we have shown that the proposed algorithm provides a high performance even with a low sampling rate, which helps to preserve customers’ privacy. We have also introduced IAD, an algorithm to detect attacks against DLC and ILC in AMI with short delay, by monitoring abnormalities in consumption patterns. Through simulations and analysis we have proved the effectiveness of the proposed method.
Chapter 6

Conclusions and Future Work

In this chapter, we highlight the contributions of each chapter and summarize the results. We also present a number of possible directions for further research.

6.1 Summary of Accomplished Work

In this thesis we have studied security and privacy issues in smart grids, and proposed several novel algorithms for detecting malicious activities against AMI. In particular, in Chapter 3 and 4 we have addressed the problem of intrusion detection and prevention in ZigBee-based HANs as one of the most vulnerable AMI subsystems. We have proposed HANIDPS which is an IDPS tailored for unique requirements and challenges of intrusion detection in HANs. We have also designed methods for detecting major malicious activities against AMI including energy theft and attacks against direct and indirect load control in Chapter 5. Such attacks can be originated from any of the AMI networks comprising HANs, NANs and WANs. A summary of the work accomplished in each chapter is as follows:

- In Chapter 2, we have investigated cyber security and privacy issues in smart grid. We have surveyed the existing solutions and provided direction for further research to develop methods tailored for smart grid requirements.

- In Chapter 3, we have proposed algorithms for detecting and preventing spoofing
attacks in static IEEE 802.15.4 networks. The proposed algorithms use the spatial dependency of RSS values to detect and distinguish malicious frames. We have argued that this parameter is hard to forge by attackers, therefore is effective in detecting and preventing the attacks. We have evaluated the performance of the proposed methods through extensive analysis and experiments, and tested their functionality under different attack scenarios. We have also studied the resource usage and network overhead of the proposed methods. Evaluation results proved the effectiveness of our approach. Further, we have compared the performance of the proposed algorithms with existing RSS based techniques and showed that our approach provides a higher performance with lower resource usage. Considering that the proposed algorithms in this chapter are RSS based, which has the same properties in IEEE 802.11 standard as in IEEE 802.15.4 standard, these algorithms are also applicable for single antenna WiFi local area networks.

- In Chapter 4, we have argued that due to the large scale and high cost of false positives, in the context of HAN, IDPSs which not only detect but also automatically stop the attacks are highly required. We have introduced HANIDPS, a novel IDPS for ZigBee-based HANs. HANIDPS combines a model-based intrusion detection method tailored for HAN specifications, and a machine learning-based prevention technique which enables dynamic defense against adversaries without prior knowledge of the attacks. Using novel techniques for spoofing prevention, and through utilization of effective mechanism for countermeasuring intentional and unintentional interference, as well as detecting and dropping frames which are not compliant with HAN standards, HANIDPS secures the network against a variety of attack types. Moreover, we have studied the existing attacks against ZigBee networks and showed the effectiveness of our approach against them. Although HANIDPS has been designed
for ZigBee HANs it is also applicable for other wireless sensor networks with similar properties, i.e., for networks that employ limited numbers of protocols and application and it is possible to define a thorough and accurate feature space for them using the corresponding standards and specifications.

• In Chapter 5, we have introduced novel techniques for detecting malicious activities against AMI by leveraging the predictability property of AMI data. We have proposed CPBETD, a consumption pattern-based energy theft detector which through application of appropriate classification and clustering techniques along with transformer meters provides a high performance and cost-effective solution for detecting energy theft attempts. We have shown that unlike existing classification based techniques, the presented method is robust against non-malicious changes in consumption patterns as well as contamination attacks. In addition, we have argued that existing metrics for evaluating the performance of ETDSs presented in the literature are inadequate and other parameters such as Bayesian detection rate and implementation and operational costs must also be considered. We have proved the effectiveness of CPBETD by evaluating its performance on real data of 5000 customers. We have, further, introduced IAD for detecting attacks against direct and indirect load control. IAD monitors the consumption pattern of a group of customers within a NAN, and detects abnormal behaviors with a short delay. We have provided a synthetic dataset to evaluate the performance of IAD and observed satisfactory results.

6.2 Suggestions for Future Work

In the following, some interesting directions for extending the work presented in this dissertation are introduced.
1. **Application of physical layer features for spoofing detection and prevention:** In Chapter 3, we introduced algorithms for detecting spoofing attacks in static IEEE 802.15.4 networks based on RSS values of received frames. We argued that since physical layer parameters are hard to forge, they can effectively be used for distinguishing genuine and malicious nodes. One possible direction for future research is application of other physical layer features, such as round trip time (RTT) of the request to send/clear to send (RTS/CTS) handshakes. Using multiple parameters can result in a more accurate attack detection solution. Besides, it can help to develop algorithms that are adaptable for non-static and mobile networks.

2. **Finding outliers to detect energy theft attempts:** In Chapter 5, we proposed an ETDS which distinguishes unethical customers by monitoring abnormalities in consumption patterns of each customer. One possible direction to extend the proposed work is to group customers with similar usage pattern in a neighborhood and apply cross-correlation and outlier detection algorithms, to identify customers exhibiting different trend in their usage compared to similar customers. A customer with different pattern compared to the neighbors is more suspicious for energy theft. For instance, this can be helpful in detecting artificial-looking patterns.

3. **Prioritizing theft detection alerts based on monetary loss:** Another direction to extend our work on energy theft detection in Chapter 5 is to prioritize theft detection alerts based on the resulted financial loss. By comparing the malicious pattern with expected normal pattern monetary loss due to theft attempts can be calculated. Onsite inspection is an expensive procedure; responding to all theft alerts equally is not cost effective for the utilities. Instead, theft attempts that result in higher loss to the utilities must have more priority for on-site inspection.
Bibliography


[87] ZigBee smart energy standard 1, 2014.


<table>
<thead>
<tr>
<th>Reference</th>
<th>Description</th>
</tr>
</thead>
</table>