2019

Crypto-aided MAP detection and mitigation of false data in wireless relay networks

Xudong Liu
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Communication Commons

Recommended Citation
https://lib.dr.iastate.edu/etd/17730

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Crypto-aided MAP detection and mitigation of false data in wireless relay networks

by

Xudong Liu

A dissertation submitted to the graduate faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering (Communications and Signal Processing)

Program of Study Committee:
Sang Wu Kim, Major Professor
  Yong Guan
  Chinmay Hegde
  Thomas Daniels
  Long Que

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University
Ames, Iowa
2019
Copyright © Xudong Liu, 2019. All rights reserved.
DEDICATION

I would like to dedicate this thesis to my beloved wife Yu Wang, without her emotional support I would not be able to complete this work.

Also I want to express my sincere thanks to my parents for their fully support and help during the time of my doctoral program.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>x</td>
</tr>
<tr>
<td>CHAPTER 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Research Motivation and Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Literature Review</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Research Objectives</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Outline of Thesis</td>
<td>5</td>
</tr>
<tr>
<td>1.5 References</td>
<td>5</td>
</tr>
<tr>
<td>CHAPTER 2. BAYESIAN TEST FOR DETECTING FALSE DATA INJECTED PACKET IN WIRELESS RELAY NETWORKS</td>
<td>9</td>
</tr>
<tr>
<td>2.1 Overview</td>
<td>9</td>
</tr>
<tr>
<td>2.2 System Model</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Bayesian Test</td>
<td>11</td>
</tr>
<tr>
<td>2.3.1 Sufficient Statistic</td>
<td>11</td>
</tr>
<tr>
<td>2.3.2 Bayesian Test</td>
<td>12</td>
</tr>
<tr>
<td>2.4 Probability of Detection Error</td>
<td>14</td>
</tr>
<tr>
<td>2.4.1 Probability of False Alarm</td>
<td>14</td>
</tr>
<tr>
<td>2.4.2 Probability of Missed Detection</td>
<td>15</td>
</tr>
<tr>
<td>2.4.3 Asymptotical Analysis</td>
<td>15</td>
</tr>
<tr>
<td>2.5 Numerical Results</td>
<td>18</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>20</td>
</tr>
<tr>
<td>2.7 Appendix: Proof of Proposition 1</td>
<td>21</td>
</tr>
<tr>
<td>2.8 References</td>
<td>25</td>
</tr>
<tr>
<td>CHAPTER 3. ENHENCE BAYESIAN TEST DETECTION PERFORMANCE WITH CRYPTOGRAPHY FOR SHORT LENGTH PACKET</td>
<td>27</td>
</tr>
<tr>
<td>3.1 Overview</td>
<td>27</td>
</tr>
<tr>
<td>3.2 Message Authentication Code</td>
<td>27</td>
</tr>
<tr>
<td>3.2.1 Probability of False Alarmed and Missed Detection for MAC</td>
<td>29</td>
</tr>
</tbody>
</table>


LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Symbols.</td>
<td>11</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>MAP combining rule.</td>
<td>32</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>List of notations and terms.</td>
<td>39</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>MAP combining rule in crypto-aided MAP test.</td>
<td>86</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Two-hop wireless relay transmission.</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Probability mass function of $W(Z)$ given $H_0$ and $H_1$; $e = 0.1$.</td>
<td>16</td>
</tr>
<tr>
<td>2.3</td>
<td>Probability of missed detection, $P_M$, versus probability of false alarm, $P_F$, for different values of $W(F)/n$; (127,64) BCH code, $d_{\text{min}} = 21$, $n = 127$, $e = 0.1$.</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>Probability of missed detection, $P_M$, and probability of false alarm, $P_F$, versus the fraction of bits modified, $W(F)/n$; (127,64) BCH code, $e = 0.1$, $\beta = 0.9$.</td>
<td>20</td>
</tr>
<tr>
<td>2.5</td>
<td>Probability of detection error versus BER between S and D, $e$, for different $(n,k)$ BCH codes; $\beta = 0.9$.</td>
<td>21</td>
</tr>
<tr>
<td>3.1</td>
<td>The message authentication code example.</td>
<td>28</td>
</tr>
<tr>
<td>3.2</td>
<td>Frame structure: (a) cryptographic integrity check, (b) crypto-physical integrity check. $h(X_i)$ is the MAC tag for $X_i$ and $h(X_1,X_2,X_3,\ldots,X_N)$ is the MAC tag for $X_1,X_2,X_3,\ldots,X_N$.</td>
<td>29</td>
</tr>
<tr>
<td>3.3</td>
<td>Crypto-Physical Integrity Check.</td>
<td>30</td>
</tr>
<tr>
<td>3.4</td>
<td>Crypto-Physical Integrity Check.</td>
<td>33</td>
</tr>
<tr>
<td>3.5</td>
<td>Probability of detection error versus latency (bits); $N = 1$, $\beta = 0.9$, $e = 0.05$, (127,64) BCH.</td>
<td>34</td>
</tr>
<tr>
<td>4.1</td>
<td>(a) Wireless relay network model and (b) information flow of $X_i$. $Y_{s,i}$ is the overhead packet from the $i$th source $S_i$ and $Y_{r,i}$ is the received packet from the relay $R$.</td>
<td>37</td>
</tr>
<tr>
<td>4.2</td>
<td>The example of $C_{ij}$.</td>
<td>42</td>
</tr>
<tr>
<td>4.3</td>
<td>The structure of $F_i$.</td>
<td>45</td>
</tr>
<tr>
<td>4.4</td>
<td>Selective MAP test.</td>
<td>52</td>
</tr>
<tr>
<td>4.5</td>
<td>Average probability of detection error versus the number of source nodes, $N$, for different combination of BERs; BCH(63,24), $W(F_i) = d_{\text{min}}$.</td>
<td>53</td>
</tr>
<tr>
<td>4.6</td>
<td>Probability of detection error versus the number of source nodes for different values of $\alpha$, $N$; BCH(31,21), $e_i = 0.1$ for $i = 1,2,\cdots,N/2$ and $e_i = 0.05$ for $i = N/2 + 1,\cdots,N$, $W(F_i) = d_{\text{min}}$.</td>
<td>57</td>
</tr>
<tr>
<td>4.7</td>
<td>Probability of detection error versus the number of source nodes for different values of $\alpha$ with its upper bound, $N$; BCH(31,21), $e_i = 0.1$ for $i = 1,2,\cdots,N/2$ and $e_i = 0.05$ for $i = N/2 + 1,\cdots,N$, $W(F_i) = d_{\text{min}}$.</td>
<td>58</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.8</td>
<td>Probability of detection error versus the overhearing bit error rate for different value of $N$; BCH(63,24), $B = 10$, $\alpha = 0.2$.</td>
<td>59</td>
</tr>
<tr>
<td>4.9</td>
<td>Probability of detection error versus percentage of modified bits in a packet, $W(F_i)/n$, for different value of $N$; BCH(31,21), $\alpha = 0.3 e_i = 0.1$ for $i = 1,2,\cdots,N/2$ and $e_i = 0.05$ for $i = N/2 + 1,\cdots,N$.</td>
<td>60</td>
</tr>
<tr>
<td>4.10</td>
<td>Probability of detection error versus the number of source nodes $N$ for different values of $(n,k)$; BCH(31,21), $\alpha = 0.2 e_i = 0.1$ for $i = 1,2,\cdots,N/2$ and $e_i = 0.05$ for $i = N/2 + 1,\cdots,N$.</td>
<td>61</td>
</tr>
<tr>
<td>4.11</td>
<td>Average probability of detection error versus the number of source nodes, $N$; BCH(63,24), $e_i = 0.2$ for $i = 1,2,\cdots,N/3$ and $e_i = 0.05$ for $i = N/3 + 1,\cdots,N$, $W(F_i) = d_{\min}$.</td>
<td>62</td>
</tr>
<tr>
<td>5.1</td>
<td>Packet recovery flow.</td>
<td>69</td>
</tr>
<tr>
<td>5.2</td>
<td>Normalized throughput versus probability of mistakenly accepting a modified packet; BCH(63,24), $\alpha = 0.2$, $N = 20$, $e_i = 0.1$ for $i = 1,\cdots,N/2$ and $e_i = 0.05$ for $i = N/2 + 1,\cdots,N$.</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>Normalized throughput versus probability of mistakenly accepting a modified packet; BCH(63,24), $\alpha = 0.2$, $N = 20$, $e_i = 0.1$ for $i = 1,\cdots,N/2$ and $e_i = 0.05$ for $i = N/2 + 1,\cdots,N$.</td>
<td>77</td>
</tr>
<tr>
<td>6.1</td>
<td>The Cryptographic Check algorithm.</td>
<td>81</td>
</tr>
<tr>
<td>6.2</td>
<td>Crypto-aided MAP test.</td>
<td>82</td>
</tr>
<tr>
<td>6.3</td>
<td>Selective Crypto-Aided MAP Test.</td>
<td>88</td>
</tr>
<tr>
<td>6.4</td>
<td>Probability of detection error versus number of source nodes, $N$, for different detection schemes; BCH(63,24), $B = 10$, $\alpha = 0.5$, $e_i = 0.1$ for $i = 1,2,\cdots,N/2$ and $e_i = 0.05$ for $i = N/2 + 1,\cdots,N$.</td>
<td>92</td>
</tr>
<tr>
<td>6.5</td>
<td>The LHS and RHS of MAP combining rule values versus the overhearing bit error rate, given $(\hat{H}_C, \hat{H}_P) = (H_0, H_1)$; BCH(63,24), $B = 10$, $\alpha = 0.2$.</td>
<td>93</td>
</tr>
<tr>
<td>6.6</td>
<td>Probability of detection error versus MAC length, $B$, for different values of $\alpha$; BCH(63,24), $N = 20$, $e_i = 0.1$ for $i = 1,2,\cdots,N/2$ and $e_i = 0.05$ for $i = N/2 + 1,\cdots,N$.</td>
<td>94</td>
</tr>
<tr>
<td>6.7</td>
<td>Probability of detection error versus the overhearing bit error rate for different value of $N$; BCH(63,24), $B = 10$, $\alpha = 0.2$.</td>
<td>95</td>
</tr>
</tbody>
</table>
In here, I would like to officially thank to all the persons who have helped me during the time of my Ph.D. program. First and foremost, I must express my gratitude to my major professor Sang Wu Kim, who has patiently guide my research during these years. It is him let me know how to conduct a science research and how to write the thesis. His advices and insights often encouraged me in various aspect of my study, they can always help me to overcome obstacles when I had difficulty in the research time, so I can keep progress until today.

I would also like to thank Prof. Yong Guan, Prof. Chinmay Hegde, Prof. Toms Daniels and Prof. Long Que for being committee members in my POS. Thanks to their efforts and suggestions, the research in this thesis could become more robust and comprehensive.

In addition, I want to thank all my friends, with their companion, my daily life could become full of joy and delight.
Relays are widely used in various types of networks to extend the coverage and increase the communication reliability. The underlying assumption is that relay nodes will faithfully forward the received data to the next node. However, a relay node in practice may misbehave by acting uncooperatively: such as refuse signal relaying, garbling the transmitting signal or performing intended malicious attack. Among all of types of misbehaviors, previous studies has shown that relay malicious attack can cause significant degradation of communication, this intended malicious attack are usually known as false data injection or Byzantine attack. By altering the transmitting bits or injecting false data in transmission packets, relay nodes in the routing path could pretend to forward the correct message, but in fact, change the information so that he can deceive the destination to process his intended message, which may result in severe consequence.

In the research of this thesis, we first develop the Bayesian test based on sufficient statistic for detecting the relay misbehavior at the packet level in lossy one-way wireless relay networks. The Bayesian test is optimal in the sense of minimizing the probability of detection error. This novel approach works in the physical-layer and does not require any cryptographic keys, tracing symbols, or third party authentication. Instead, the overheard packet from the source is exploited, which is available for free due to the broadcast nature of wireless medium. The enabling factor in our approach is that the overheard packet, although erroneous due to channel impairments, contains the true information of source packet, hence can serve as a reference in verifying the correctness of relayed packet. By deriving the probability of false alarm and missed detection as a function of the overhearing bit error rate (BER), the number of bits modified in a packet with the packet length. We found that the probability of false alarm and missed detection decrease as the packet length increases and converge to zero, regardless of the overhearing BER, if the packet length is sufficiently long. The Bayesian test accuracy, essentially, depends on bit error rates (BERs) in overhearing
channel, sometimes BERs may not be good enough. To address this issue, a cross-layer detection mechanism where the cryptographic detection outcomes are used to estimate the false injection rate is proposed. To reduce the computational cost and bandwidth overhead associated with the cryptographic detection, the packets are attached with only a short length keyed hash. Our results will show that the detection performance that can be provided with the perfect knowledge of the false injection rate can be achieved by having only message authentication code (MAC) with about 15% of the its attaching packet bits length.

In order to make proposed Bayesian detection scheme adapted to the emerging high-speed wireless relay networks, where massive short packets are collected from multiple sources through a relay. A new detection scheme named as maximum a posteriori probability (MAP) test, is developed for detecting false data in wireless relay networks composed of multiple sources. The MAP test can minimize the probability of detection error, meanwhile does not incur any overhead. The fundamental challenge is that the reference information provided by the overheard packet is erroneous (noisy) due to the channel impairment between the source and the destination. We seek to mitigate the effects of unreliable reference information on the accuracy of the MAP test by leveraging the multiplicity of source nodes. It can be shown that the average probability of detection error decreases exponentially as the number of source nodes increases, regardless of the reliability of the reference information. Hence, a powerful integrity check mechanism can be constructed at the physical layer if the number of source nodes is large enough. Then, we discuss a packet recovery mechanism after the MAP test to repair false alarmed packets, i.e. unmodified packets that are mistakenly declared as modified. The proposed recovery mechanism relies on the likelihood ratio test (LRT) that conveys the confidence that a particular packet is modified or not. It allows the destination to determine, with no additional feedback or information from the source nodes, which packets are likely to be correct and to recover the remaining packets, thereby improving the throughput.

To enhance the performance of MAP test, another cross-layer detection mechanism, named as crypto-aided MAP test is proposed, in which the MAP test is able to extract prior information of relay attack probability by collecting the detection outcomes of cryptographic hash check results.
The proposed scheme seeks to apply a lightweight hash tag on each packet such that latency and overhead for integrity check can be made insignificant. Also to overcome the drawbacks caused by the short hash, the physical layer detection outcomes and cryptographic hash check results are combined using maximum a posteriori probability (MAP) detection rule to assure low missed detection rate.

Above all, the proposed scheme in this thesis can provide a high accuracy efficient detection efficiency to guarantee data integrity for those high-density wireless networks while operating under resource constraints. Also our results show that crypto-aided MAP test has significant improvement over conventional cryptographic and it doesn’t requires any training set or pre-embedded tracing bits to help make the decision, like other exist detection method does. In addition, the detection error of crypto-aided MAP test can be reduced to zero if the number of collected packets is sufficiently large, even if the bit error rate in overhearing channel is very high. This novel technique will benefit a wide variety of limited resources wireless applications where the traffic type are tend to be short and massive, meanwhile the information needs to be communicated in trustworthy and timely manner.
CHAPTER 1. INTRODUCTION

1.1 Research Motivation and Background

The development of wireless techniques has grown explosively during the past decades and it has been considered as a key technique in future communication networks. Due to the fading and vulnerability nature of wireless media, the radio transmission range of single wireless device is usually limited and itself is almost impossible to fulfill all the requirement of a intact networks. In order to extend the coverage and enhance the communication reliability of a wireless network Access (2010) Chandra et al. (2011), relaying techniques has drawn considerable attention in recent years. In reality, they have been widely applied in various types of networks applications Wu et al. (2018); Swamy et al. (2015), such as wireless sensors networks (WSNs) Cheng et al. (2008), remote control application Bulakci et al. (2011), notably the Internet of Things (IoT) Xu et al. (2016); Lu et al. (2017) and the emerging fifth-generation (5G) system Gupta and Jha (2015). Furthermore, in order to save the energy and avoid the redundant transmission of data, data from different sources are usually aggregated in relay node before forwarded to the base station in WSN Akila and Sheela (2017); Parameswari and Raseen (2013), by doing so, the communication efficiency throughout whole network can be significantly improved.

However, all of relaying systems above essentially have a underlying assumption in common: that is they all assume relay nodes will faithfully comply with cooperative protocols and forward their received data to the destination/next node. In practice, wireless networks are often deployed in a remote or hostile environment Poornimima and Amberker (2010), relay nodes has high chance to be compromised or controlled by the adversary Zou et al. (2016), such malicious relay in the routing path may pretend to forward the correct message Sun et al. (2016), but in actual, modify the data and send falsified information to the base station Vempaty et al. (2013). This relay misbehavior, is known as false data injection or Byzantine attack Awerbuch et al. (2004); Modares et al. (2011),
has be shown that it will reduce the communication reliability significantly Dehnie et al. (2007). Because the false data can follows validate data structure, any techniques that are used to correct bit changes by checking the data format, as error correction code (ECC) does, would be unable to against this type of attack. Therefore, the requirements for data integrity become crucial for proper system operation as the data move through the relay, an effective approach that can mitigate of false data attack in wireless relay networks, particularly in the high-speed future networks, is much more necessary than before.

1.2 Literature Review

In tradition, the cryptographic scheme using secret keys at upper layers is the most widespread approach to ensure data integrity, with pre-sharing the secret messages (named as keys) between source and destination in the initial phase, the receiver could be able to detect whether the messages from relays are legitimate or not. For example, in Papadimitratos and Haas (2006) the message authentication code at the network layer is applied to detect the relay misbehavior. Unfortunately, such approaches may not be desirable since encrypting all the message may impose larger overhead and additional bandwidth if it adopts a complicate cryptographic algorithm. To avoid this situation, those cryptographic integrity check techniques Paar et al. (2010a); Hu et al. (2005) usually span large messages so that the overhead for integrity check can be made insignificant. However, longer data packets incurs longer latency, as a result, a large memory space is required at destination end, because the integrity check can begin only after all messages are received and stored in a memory. Other cryptography involved techniques, as in Dhurandher et al. (2018); Sreevidya and Rajesh (2018), an cryptography RSA algorithm is implemented to detect relays misbehavior. In Miao et al. (2017); He and Yener (2013) sensors’ outputs are coded by special cryptographic scheme. Such approaches, are able to detect relay’s false data injection attacks, but at the expense of a higher computational cost.

Alternatively, the techniques that aims to detect relay misbehavior from the physical layer perspective are also developed in those years. For amplify-and-forward (AF) system Hou et al. (2013),
the source node can detect relay’s malicious misbehavior by listening to the relay transmission and exploiting the correlation property between the transmitted and received signals. Other works that based on the channel state information (CSI) are also investigated: for example, a likelihood ratio approach Kim (2015) that exploits information of wiretap channel was developed in detecting the relay misbehavior. In Graves and Wong (2012), relay misbehavior check were accomplished through the estimation of the attack channel, which is obtained from the received conditional distribution empirically observed by a source node. Moreover, the relay misbehaviors detection based on the probability distribution of observation have attained noteworthy consideration as well: in Yin et al. (2017); Lv et al. (2018), authors shows that some specific types of malicious attacks can be detected with high probability even without reliable channel state information (CSI). And the authors of Cao et al. (2016) proved that detecting Byzantine attacks without clean reference is feasible if the network satisfies a non-manipulability condition and a sufficiently large number of channel observations is available. More recently, the impact of false data injection attacks and its mitigation methods have been investigated Liu et al. (2016); Tan et al. (2017). By exploiting the broadcast nature of wireless media, Bayesian hypothesis test in detecting false data injection for single packet has been studied in Liu et al. (2018). And a false data injection defending schemes based on greedy algorithm with margin setting algorithm are derived in Yang et al. (2017); Wang et al. (2017). The authors of Rawat and Bajracharya (2015) developed a chi-square detector and cosine similarity matching approaches in detecting false data injection.

In particularly, schemes based on tracing-symbols have been developed Khalaf et al. (2014); Lo et al. (2012, 2015), by inserting some random tracing symbols into the transmission bits and comparing tracing symbols with known symbols, the destination is capable of detecting whether relay has modified packets or not. These tracing-symbols based techniques are highly dependent on the number of tracing symbols used: its accuracy can be unsatisfied if number of tracing-symbols is just a small portion within overall transmission symbols. On the other hand, its overhead may become significant if the portion of tracing-symbols are very high or the transmission packets is within short length.
All those above mentioned techniques may work efficiently either with certain circumstances or under the wireless network environment where long data packets are supported. However, in the future wireless communication system Durisi et al. (2016a); Bennis et al. (2018), such as 5G system Schaich et al. (2014); Osseiran et al. (2014), the traffic type tends to be short and massive packets, the requirements for minimal overhead, low latency and high reliability are much more demanding than before. The schemes, which incurs large overhead due to the packet length change, like tracing-symbols based; or imposes high latency due to necessity of collecting sufficient number of observations, like distribution of observation based. Those techniques may not be efficient under the condition of those new requirements.

1.3 Research Objectives

The objectives of this research is to develop a new detection technique to address the false data injection issues in wireless relay networks. Moreover, this detection technique should have several characteristics: First of all, the devices in system are usually small and resources-limited, for example, like distributed sensors networks Kuorilehto et al. (2005), so this scheme should mainly work on the physical layer and its computation complexity should be low; otherwise working in the up-layers or with a complicate algorithm usually demands much more computational cost, which may not be guaranteed in practice. Secondly, it should be robust against any types of false data injection attack, this is because the adversary’s attack models are often hard to acquire or predict. Assuming some particular attack model, like most of literature does, would only potentially increase risk that adversaries may degrade the mitigation schemes by simply switching its attack model. In final, this scheme should be able to be easily applied to any type of wireless communication system, especially for the emerging high-speed wireless networks. That is to say, its structure must be effective with data format where the transmission packets tend to be short and the number of packets tend to numerous.


1.4 Outline of Thesis

This thesis is organized as follows, chapter 2 describes the system model been investigated, where source transmit messages to destination target. But due to channel poor quality between them, source also broadcast the signal to a near relay node and Let node help relaying the message to destination target. However since the relay is untrustworthy, it may forward modified messages to the destination to undermine the information integrity. And we analyze the error probability of system model based on the different detection results. In chapter 3, we proposed an optimal detection method to detect the modified messages in the physical layer by exploiting the hamming distance [10] property between channel codewords. This is followed with the theoretical analysis for applying optimal detecting threshold. Since the calculation of optimal detecting threshold requires the pre-knowledge of the attack probability, a cryptography-aided estimation strategy of attack probability is developed. The numerical results and conclusions are in chapter 4 and chapter 5 respectively.

1.5 References


CHAPTER 2. BAYESIAN TEST FOR DETECTING FALSE DATA INJECTED PACKET IN WIRELESS RELAY NETWORKS

2.1 Overview

The emergence of resource-constrained, decentralized wireless relay networks imposes new challenges on data integrity measures Liang et al. (2008). Almost all the existing wireless relay communication systems assume that relay nodes will faithfully forward the received data to the next node. However, a relay node in the routing path may misbehave by acting selfishly Gupta et al. (2011), such as lower the transmission power, refuse signal relaying, garbling the original signal Yi et al. (2012) and performing intended malicious attack Zhang et al. (2014); Dehnie et al. (2007). Previous studies has shown that relay misbehaviors, especially the intended malicious attack which known as false data injection attack or Byzantine attack, may cause severe system malfunction. In this chapter, we develop the Bayesian test for detecting the relay false data injection attack at the packet level in lossy one-way wireless relay networks. The proposed approach exploits overheard erroneous packet from the source as a reference in verifying the correctness of relayed packet at the destination. Then the probability of false alarm and missed detection as a function of the overhearing error rate, the number of bits modified in a packet, and the packet length. It can be shown that the probability of false alarm and missed detection decrease as the packet length increases and converge to zero, regardless of the overhearing error rate, if the packet length is sufficiently long.

2.2 System Model

Consider a one-way two-hop wireless relay network in which the source, $S$, wishes to send its message to the destination, $D$, through an untrustworthy relay, $R$. The system model is shown in Fig. 2.1. In phase 1, $S$ sends its packet $X$, encoded by an $(n,k)$ linear code, to $R$. We assume that a hybrid ARQ type error control mechanism is employed to to detect a transmission (channel) error. If
there is an error (decoding failure), additional information is transmitted until the packet is correctly decoded by \( R \). The decoded word at \( R \) can be expressed as \( X_r = X \oplus E_{sr} \), where \( E_{sr} \) denotes the error vector between \( S \) and \( R \) and \( \oplus \) denotes the bit-wise modulo-2 addition. Due to the broadcast nature of the wireless medium, \( D \) can overhear the transmitted packet from \( S \) (dotted line in Fig. 2.1). However, the overhearing can be erroneous because of channel impairments between \( S \) and \( D \). The overheard packet can be expressed as \( Y_s = X \oplus E_{sd} \), where \( E_{sd} \) denotes the channel error vector between \( S \) and \( D \). We assume that bit errors occur randomly by interleaving and deinterleaving. In phase 2, \( R \) modifies the decoded word \( X_r \) into another packet \( X'_r \) with probability \( \alpha \). This mixed relay behavior of acting maliciously with probability \( \alpha \) and cooperatively with probability \( 1 - \alpha \) introduces an uncertainty about the relay misbehavior and makes the detection more challenging. Such a stochastic attack model was considered in various intrusion detection systems Gu et al. (2006). The modified packet \( X'_r \) can be expressed as \( X'_r = X_r \oplus F_r = X \oplus E_{sr} \oplus F_r \) for some non-zero vector \( F_r \) of length \( n \). The main difference between \( E_{sr} \) and \( F_r \) is that the former is an error vector introduced by nature (noise, fading), while the latter is an “intentional” error introduced by the adversary. Then, \( R \) sends \( X'_r \) to \( D \). The decoded word at \( D \) can be expressed as \( Y_r = X'_r \oplus E_{rd} = X \oplus F \), where \( E_{rd} \) denotes the error vector between \( R \) and \( D \), and \( F = F_r \oplus E_{sr} \oplus E_{rd} \) is the compound error vector between \( X \) and \( Y_r \). \( F \) captures both intentional error and natural error. \( D \) wants to verify the correctness of the decoded word \( Y_r \), i.e. whether \( Y_r = X \) or not (data integrity), by employing the overheard packet \( Y_s \) as a reference. It should be noted that since \( Y_r \) and \( X \) are codewords, \( F \) is a codeword by the property that a linear combination of any two linear
codewords is also a codeword Lin and Costello (2004). For the convenience, the symbols used in this chapter are summarized in Table 2.1.

Table 2.1  Symbols.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>message length</td>
</tr>
<tr>
<td>$X$</td>
<td>packet (codeword)</td>
</tr>
<tr>
<td>$n$</td>
<td>packet (codeword) length</td>
</tr>
<tr>
<td>$\oplus$</td>
<td>modulo-2 addition</td>
</tr>
<tr>
<td>$W(X)$</td>
<td>Hamming weight of $X$</td>
</tr>
<tr>
<td>$d$</td>
<td>Hamming distance of $Z$</td>
</tr>
<tr>
<td>$d_f$</td>
<td>Hamming distance of $F$</td>
</tr>
<tr>
<td>$E$</td>
<td>error vector on $S - D$ channel</td>
</tr>
<tr>
<td>$e$</td>
<td>bit error rate between $S$ and $D$</td>
</tr>
<tr>
<td>$F$</td>
<td>falsely injected data</td>
</tr>
<tr>
<td>$\delta$</td>
<td>detection threshold</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>false injection rate</td>
</tr>
</tbody>
</table>

### 2.3  Bayesian Test

In this section the Bayesian test for verifying the correctness of the decoded word $Y_r$ is described. The goal of Bayesian test is to determine whether $Y_r = X$ or not, i.e. $F = 0$ or not, from $Y_s$ and $Y_r$.

#### 2.3.1  Sufficient Statistic

Given $Y_s$ and $Y_r$, the mutual information (denoted by $I(x; y)$) between $F$ and $(Y_s, Y_r)$ is given by

$$I(F; Y_s, Y_r) = H(Y_s, Y_r) - H(Y_s, Y_r|F),$$  \hspace{1cm} (2.1)

where

$$H(Y_s, Y_r) = H(Y_s) + H(Y_r|Y_s).$$  \hspace{1cm} (2.2)
Then, it can be shown that

\[
H(Y_r|Y_s) = H(Y_s \oplus E_{sd} \oplus F_r \oplus E_{sr} \oplus E_{rd}|Y_s) = H(F \oplus E_{sd}),
\]

where

\[
F = F_r \oplus E_{sr} \oplus E_{rd}.
\]  
(2.4)

Similarly,

\[
H(Y_s, Y_r|F) = H(Y_s, X)
\]

\[
= H(Y_s) + H(X|Y_s)
\]

\[
= H(Y_s) + H(E_{sd}).
\]  
(2.7)

Therefore, we obtain

\[
I(F; Y_s, Y_r) = H(F \oplus E_{sd}) - H(E_{sd})
\]

\[
= H(Z) - H(Z|F)
\]

\[
= I(F; Z).
\]  
(2.10)

Overall, a sufficient statistic for \( F \) is \( Z = Y_s \oplus Y_r = F \oplus E_{sd} \). That is to say, \( Z \) contains all the information necessary to make a decision on \( F \) given \( Y_s \) and \( Y_r \).

2.3.2 Bayesian Test

Our approach to check the correctness of \( Y_r \), i.e., whether \( F = 0 \) or \( F \neq 0 \), builds on the theory of Bayesian test:

\[
\sum_{i=1}^{2^k-1} P(F = C_i|Z) \overset{F \neq 0}{\geq} P(F = 0|Z),
\]

where \( C_i \in \{C_1, \cdots, C_{2^k-1}\} \) is a non-zero codeword. Let

\[
\Lambda_i(Z) = \frac{P(Z|F = C_i)}{P(Z|F = 0)}
\]

\[
= \left( \frac{e}{1-e} \right)^{W(Z \oplus C_i) - W(Z)}
\]  
(2.13)
be the likelihood ratio between \( C_i \) and \( \mathbf{0} \), where \( e \) denotes the (overhearing) BER between \( S \) and \( D \), i.e. the probability that a bit in the error vector \( E_{sd} \) is 1, and \( W(Z) \) denotes the Hamming weight of \( Z \). If we let \( C_0 = \mathbf{0} \), then the Bayesian test (2.11) can be expressed as

\[
\sum_{i=1}^{2^k-1} \left( \frac{e}{1-e} \right)^{W(Z\oplus C_i)} P(F = C_i) \geq \beta \left( \frac{e}{1-e} \right)^{W(Z)}
\]  

(2.14)

where \( \beta = P(F = \mathbf{0}) \). If we let \( P_{sr} := P(E_{sr} \neq \mathbf{0}) \) and \( P_{rd} := P(E_{rd} \neq \mathbf{0}) \), then \( \beta = (1-\alpha)(1-P_{sr})(1-P_{rd}) \). For large number of codewords, i.e. \( 2^k \gg 1 \), the left hand side of (2.14) converges to a constant \( E[(e/(1-e))^{W(Z\oplus C_1)}] \) by the law of large numbers, that is to say

\[
E \left[ \left( \frac{e}{1-e} \right)^{W(Z\oplus C_1)} \right] \geq \frac{\tilde{H}_1}{\tilde{H}_0} \left( \frac{e}{1-e} \right)^{W(Z)}
\]  

(2.15)

where \( \tilde{H}_1 \) denotes the detection outcome being \( F \neq \mathbf{0} \) and \( \tilde{H}_0 \) denotes the detection outcome being \( F = \mathbf{0} \). From the MacWilliams and Sloane (1992), the probability that hamming weight of a BCH \((n,k)\) codeword equals to \( j \) can be approximated by \( \binom{n}{j}/2^n \), i.e. \( P[W(C_i) = j] = \binom{n}{j}/2^n \). Assume sufficient statistic \( W(Z) = d \), then we have

\[
P[W(Z \oplus C_i) = j] = P[W(C_i) = j - d + 2W(C_{i1})]
\]  

(2.16)

where \( C_{i1} \) a subvector of \( C_i \) that is composed of entries in the position where \( Z \) is 1 (since \( W(Z) = d \), the length of \( C_{i1} \) is also \( d \)). Let \( l \) denotes \( W(C_{i1}) \), so the (2.16) can be approximated as

\[
P[W(C_i) = j - d + 2W(C_{i1})] = \sum_{l=0}^{d} \frac{\binom{j-d+2l}{l}}{2^n} \frac{d}{2^n}.
\]  

(2.17)

Then the LHS in (2.15)

\[
E \left[ \left( \frac{e}{1-e} \right)^{W(Z\oplus C_1)} \right] = \sum_{j=0}^{n} \left( \frac{e}{1-e} \right)^j \sum_{l=0}^{d} \frac{\binom{n}{j-d+2l} \binom{j}{l}}{2^n} \frac{d}{2^n}
\]

\[
= \sum_{j=0}^{n} \sum_{l=0}^{d} \left( \frac{e}{1-e} \right)^j \frac{\binom{n}{j-d+2l} \binom{j}{l}}{2^{n+d}}
\]  

(2.18)

note that \( j \) actually has the limitation that \( 0 \leq j - d + 2l \leq d \). And \( j \) also has to be an integer, so limitation range of \( j \) should be from \( \max(0, d-2l) \) to \( \min(n, n+d-2l) \). Thus (2.15) yields to

\[
\sum_{l=0}^{d} \sum_{j=\max(0,d-2l)}^{\min(n,n+d-2l)} \left( \frac{e}{1-e} \right)^{j-d} \frac{\binom{n}{j-d+2l} \binom{j}{l}}{2^{n+d}} \frac{\tilde{H}_1}{\tilde{H}_0} \geq \beta \left( \frac{e}{1-e} \right)^{W(Z)}\frac{1-\alpha}{\alpha}
\]  

(2.19)
Proposition 1: The left hand side (LHS) of (2.19), $E_F[(e/(1-e))^{W(Z\oplus F) - W(Z)}]$, is a monotonically increasing (decreasing) function of $W(Z)$ for $e < 1/2$. The steps of proving Proposition 1 are provided in Appendix 2.7.

Due to Proposition 1, it can be known that LHS of (2.19) is an increasing function of $d$ ($W(Z)$) when $e < 1/2$, and the RHS of (2.19) is a function of $\alpha$. Since $\alpha$ can be consider as unchanged during one codeword transmission, the RHS of (2.19) can be viewed as a constant. Therefore, from (2.19), we can infer that there must be exist (if does) a value $\delta$ that for all the $d < \delta$, if LSH of (2.19) < RHS of (2.19), then LSH of (2.19) \geq RHS of (2.19) if $d \geq \delta$. Hence, the Baysian test for $e < 1/2$ reduces to

\[ W(Z) \begin{cases} \overset{\hat{H}_1}{\geq} \delta \\ \overset{\hat{H}_0}{<} \end{cases} \]  

(2.20)

where $\delta$ is detection threshold.

Intuitively, it should be aware that the detection metric $W(Z) = W(F \oplus E_{sd})$ actually excludes the message bits from $X$ and only contains the modified bits $F$ and noise bits $E_{sd}$. Thus it is proportional to $W(F)$ on the average sense. In other words, the metric $W(Z)$ represents the hamming weight false injected vector, larger $W(Z)$ means larger $W(F)$ on average, meanwhile the higher chance the packet is is decided as $\hat{H}_1$.

### 2.4 Probability of Detection Error

The effectiveness of the proposed physical-layer detection technique can be measured by two metrics: the probability of false alarm and the probability of missed detection.

#### 2.4.1 Probability of False Alarm

The event of false alarm occurs if $W(Z) \geq \delta$ for some detection threshold $\delta$ under $F = 0$. Since the bit errors occur independently within $Y_s$, $W(E_{sd})$ is binomially distributed with parameters $n$

\[ W(E_{sd}) \sim \text{Binomial}(n, p) \]

If $e > 1/2$, the decision should be reversed.
and $e$, hereafter denoted by $\text{Binom}(n, e)$. Hence, the probability of false alarm is given by

$$P_F = P(W(Z) \geq \delta|F = 0) \quad (2.21)$$

$$= P(W(E_{sd}) \geq \delta) \quad (2.22)$$

$$= \sum_{i=\delta}^{n} \binom{n}{i} e^i (1 - e)^{n-i}. \quad (2.23)$$

### 2.4.2 Probability of Missed Detection

The event of missed detection occurs if $W(Z) < \delta$ under $F \neq 0$. Let $F_1$ denote the all-one subvector of $F$ and $F_0$ denote the all-zero subvector of $F$. For example, if $F = 10110$ then $F_1 = 111$ and $F_0 = 00$. Similarly, let $E_1$ denote the subvector of $E$ which is composed of the entries in the positions where $F$ is 1 and $E_0$ denote the subvector of $E$ which is composed of the entries in the positions where $F$ is 0. For example, if $E = 01010$ and $F = 10110$ then $E_1 = 001$ and $E_0 = 10$.

Now, suppose that $W(F) = d_f$. Since the each entry of $E$ is 1 has probability $e$ and they are also independent, we obtain the $W(F_0 \oplus E_0) \sim \text{Binom}(n - d_f, e)$ and $W(F_1 \oplus E_1) \sim \text{Binom}(d_f, 1 - e)$.

Then, the probability mass function of $W(Z)$ under $H_1$ is given by

$$P(W(Z) = d|F \neq 0) = P(W(F_0 \oplus E_0) + W(F_1 \oplus E_1) = d)$$

$$= \sum_{i=\min(d, d_f)}^{\min(d, d_f)} \binom{d_f}{i} (1 - e)^i e^{d_f - i} \left[ \binom{n - d_f}{d - i} e^{d - i} (1 - e)^{n - d_f - (d - i)} \right]. \quad (2.24)$$

### 2.4.3 Asymptotical Analysis

In this subsection, we show the asymptotic analysis when the code (packet) length, $n$, is sufficiently large. Fig. 2.2 shows the probability mass function (PMF) of $W(Z)$ given $H_0$ and $H_1$ for different packet lengths $n$ with almost same code rate $k/n$. Where a false packet $F$ is chosen from the set of nonzero valid codewords, i.e. $W(F) = d_f$. It can be observed that PMF curves are shrinking to the mean of $W(Z)$ as the packet length, $n$, increases, where the mean values of $W(Z)$ given $H_0$ and $H_1$ are given by

$$\mathbb{E}[W(Z|H_0)] = ne \quad (2.26)$$
and

\[ E[W(Z|H_1)] = (n - d_f)e + d_f(1 - e) \]

\[ = ne + d_f(1 - 2e), \]  

respectively. For large \( n \), it follows from the Law of Large Numbers that \( W(Z|H_0)/n \) and \( W(Z|H_1)/n \) converge to their mean values, \( e \) and \( e + d_f(1 - 2e)/n \), respectively. Therefore, the presence of false packet \( F \) can be detected almost surely as long as \( e < 1/2 \) by choosing the detection threshold locates at the middle between \( E[W(Z|H_0)] \) and \( E[W(Z|H_1)] \), thus the threshold \( \delta \) in (2.20) can be approximated by

\[ \delta = \frac{E[W(Z|H_0)] + E[W(Z|H_1)]}{2} \]

\[ = ne + d_f(0.5 - e). \]  

Meanwhile the \( P_F \) and \( P_M \) both converge to zeros as \( n \to \infty \) when \( e < 1/2 \). let

\[ U_0 = W(E)/n \]  

Figure 2.2 Probability mass function of \( W(Z) \) given \( H_0 \) and \( H_1; e = 0.1 \). 

\( W(Z|H_1)/n \) converge to their mean values, \( e \) and \( e + d_f(1 - 2e)/n \), respectively. Therefore, the presence of false packet \( F \) can be detected almost surely as long as \( e < 1/2 \) by choosing the detection threshold locates at the middle between \( E[W(Z|H_0)] \) and \( E[W(Z|H_1)] \), thus the threshold \( \delta \) in (2.20) can be approximated by

\[ \delta = \frac{E[W(Z|H_0)] + E[W(Z|H_1)]}{2} \]

\[ = ne + d_f(0.5 - e). \]  

Meanwhile the \( P_F \) and \( P_M \) both converge to zeros as \( n \to \infty \) when \( e < 1/2 \). let

\[ U_0 = W(E)/n \]  

(2.29)
and
\[ U_1 = W(F \oplus E)/n. \quad (2.30) \]

Then, it follows from the (2.26) and (2.27) that
\[ \mathbb{E}(U_0) = e \quad (2.31) \]
and
\[ \mathbb{E}(U_1) = e + df(1 - 2e)/n. \quad (2.32) \]

From the Chebyshev’s inequality, we obtain
\[ \Pr(|U_0 - e| > \epsilon) \leq \frac{\text{Var}(U_0)}{\epsilon^2}, \quad (2.33) \]
for any \( \epsilon > 0 \), where \( \text{Var}(U_0) = e(1 - e)/n \) is the variance of \( U_0 \). Therefore, the probability of false alarm is given by
\[ P_F = \Pr(U_0 \geq \delta/n) = \Pr(U_0 \geq e + df(1 - 2e)/(2n)) \leq \frac{\text{Var}(U_0)}{[df(1 - 2e)/(2n)]^2} \sim O(n^{-1}) \quad (2.34) \]

Since \( 0 < df/n < 1 \), we obtain \( \lim_{n \to \infty} P_F = 0 \). Similarly, we obtain
\[ \Pr(|U_1 - (e + df(1 - 2e)/n)| > \epsilon) \leq \frac{\text{Var}(U_1)}{\epsilon^2} \quad (2.36) \]
for any \( \epsilon > 0 \), where \( \text{Var}(U_1) \) is the variance of \( U_1 \). Since \( W(F \oplus E) = W(F_0 \oplus E_0) + W(F_1 \oplus E_1) \) and \( W(F_0 \oplus E_0) \) and \( W(F_1 \oplus E_1) \) are independent, we obtain
\[ \text{Var}(U_1) = \frac{\text{Var}(W(F_0 \oplus E_0))}{n^2} + \frac{\text{Var}(W(F_1 \oplus E_1))}{n^2}, \quad (2.37) \]
where
\[ \text{Var}(W(F_0 \oplus E_0)) = (n - df)e(1 - e) \quad (2.38) \]
\[ \text{Var}(W(F_1 \oplus E_1)) = df(1 - e)e. \quad (2.39) \]
Therefore, we obtain \( Var(U_1) = e(1 - e)/n \). Hence, the probability of missed detection is given by

\[
P_M = Pr(U_1 < \delta/n) = Pr(U_1 < e + df(1 - 2e)/(2n)) \leq \frac{Var(U_1)}{[df(1 - 2e)/(2n)]^2} \sim O(n^{-1})
\] (2.40)

Therefore, we obtain \( \lim_{n \to \infty} P_M = 0 \).

Remark: It can be seen from (2.35) and (2.41) that \( P_M \) and \( P_F \) converge to zero as \( n \) increases, regardless of \( e \) (except \( e = 1/2 \)) that signifies the reliability of the overheard packet. This means the relay misbehavior (packet modification) can be detected without error, even if the reference information (overheard packet) is erroneous, if the packet length is sufficiently long. The perfect detection \( (P_M = P_F = 0) \) may also be achieved by a cryptographic technique such as hash-based message authentication code (HMAC) Paar et al. (2010b). However, it needs a significant bandwidth overhead and computational cost, and error-free channel \( (E_{sd} = 0) \) for sharing keys between \( S \) and \( D \).

### 2.5 Numerical Results

This section provides numerical results for the case that each packet is encoded by BCH code. Fig. 2.3 shows the receiver operating characteristic (ROC) curve that illustrates the tradeoff between the probability of missed detection, \( P_M \), and the probability of false alarm, \( P_F \), for different values of \( W(F)/n \) which captures the percentage of bits that are modified by the relay. This curve is created by plotting \( P_M \) versus \( P_F \) as the detection threshold, \( \delta \), is varied between 0 and \( n \). One can see that \( P_M \) and \( P_F \) decrease monotonically as \( W(F)/n \) increases. This follows from the fact that the mean of \( W(Z)/n \) given \( F = \mathbf{0} \) and \( F \neq \mathbf{0} \) are \( e \) and \( e + W(F)(1 - 2e)/n \), respectively, while their variances are the same (see (2.31) and (2.32)). Therefore, the events \( F = \mathbf{0} \) (\( H_0 \)) and \( F \neq \mathbf{0} \) (\( H_1 \)) can be distinguished more clearly, hence \( P_M \) and \( P_F \) decrease, as the number of modified bits increases. This enforces the adversary to limit the modification to the smallest value, i.e. \( W(F) \) is equal to the minimum distance \( (d_{\text{min}}) \), in order to minimize the probability of attack being detect-
ed, which constrains the effectiveness of attack.

Figure 2.3  Probability of missed detection, $P_M$, versus probability of false alarm, $P_F$, for different values of $W(F)/n$; (127,64) BCH code, $d_{\text{min}} = 21$, $n = 127$, $e = 0.1$.

Fig. 2.4 shows the probability of missed detection, $P_M$, and probability of false alarm, $P_F$, versus the fraction of bits modified, $W(F)/n$. The detection threshold $\delta$ is chosen such that the average probability of detection error, $P_F\beta + P_M(1 - \beta)$, is minimized. It is observed that $P_M$ and $P_F$ decrease monotonically with $W(F)/n$. This means the presence of false packet can be detected more accurately if the adversary changes more bits in a packet.

Fig. 2.5 shows the average probability of detection error, $P_M(1 - \beta) + P_F\beta$, versus the BER between S and D, $e$, for different length codes of rate close to 0.5. It is assumed that the false packet $F$ is chosen from the set of minimum weight codewords such that the average probability of detection error is maximized. The detection threshold, $\delta$, is chosen such that the average probability of detection error is minimized. One can observe that the average probability of detection error decreases as $e$ decreases and that for a given $e$ it decreases as $n$ increases. The latter follows
Figure 2.4 Probability of missed detection, \( P_M \), and probability of false alarm, \( P_F \), versus the fraction of bits modified, \( W(F)/n \); (127,64) BCH code, \( e = 0.1 \), \( \beta = 0.9 \).

from the fact that the distribution of \( W(Z)/n \) given \( F = 0 \) and \( F \neq 0 \) concentrates more to its mean values as \( n \) increases (law of large numbers), hence the events \( F = 0 \) and \( F \neq 0 \) can be distinguished more clearly.

2.6 Conclusion

In this chapter, a novel physical-layer detecting scheme, named as Bayesian test is developed, which is based on sufficient statistic for detecting the relay false data injection attack at the packet level in wireless relay networks. Our results show that the probability of missed detection and false alarm decrease as the packet length increases and converge to zero, regardless of the overhearing error rate, if the packet length is sufficiently long. This indicates that the proposed physical-layer
Figure 2.5  Probability of detection error versus BER between S and D, e, for different \((n,k)\) BCH codes; \(\beta = 0.9\).

2.7  Appendix: Proof of Proposition 1

In this section we show the proof that the LHS in the (2.19) is an increasing function of \(d\) when \(e < 1/2\).

Assume \(W(Z) = d\) is an odd integer, we define \((\frac{d}{2})^+ = \lceil \frac{d}{2} \rceil\) and \((\frac{d}{2})^- = d - (\frac{d}{2})^+\). Let \(f(d)\)
When \( d \) now let \( f \) for concision, let
represents the LHS in the (2.19), thus
\[
f(d) = \sum_{l=0}^{d} \sum_{j=\max(0,d-2l)}^{\min(n,n+d-2l)} \frac{(\frac{e}{1-e})^{j-d} \binom{n}{j-d+2l} \binom{d}{l}}{2^{n+d}} \tag{2.42}
\]
\[
= \sum_{l=d-l'} \sum_{j=d-2l'}^{n} \frac{(\frac{e}{1-e})^{j-d} \binom{n}{j-d+2l'} \binom{d}{l} \binom{d}{d-l'}}{2^{n+d}} \tag{2.43}
\]
\[
+ \sum_{l=(\frac{d}{2})^+}^{d} \sum_{j=0}^{n+d-2l-2l'} \frac{(\frac{e}{1-e})^{j-d} \binom{n}{j-d+2l'} \binom{d}{l} \binom{d}{d-l'}}{2^{n+d}} \tag{2.44}
\]
for concision, let \( f_1(d) \) denotes (2.43) and \( f_2(d) \) denotes (2.44), so \( f(d) = f_1(d) + f_2(d) \). Now define
\( l' = d - l \), in other word, \( l = d - l' \), then \( f_1(d) \)
\[
f_1(d) = \sum_{l=0}^{(\frac{d}{2})^-} \sum_{j=d-2l}^{n} \frac{(\frac{e}{1-e})^{j-d} \binom{n}{j-d+2l} \binom{d}{l}}{2^{n+d}} \tag{2.45}
\]
\[
= \sum_{l=d-l'} \sum_{j=d-2l'}^{n} \frac{(\frac{e}{1-e})^{j-d} \binom{n}{j-d+2l} \binom{d}{l} \binom{d}{d-l'}}{2^{n+d}} \tag{2.46}
\]
\[
= \sum_{l=(\frac{d}{2})^+}^{d} \sum_{j=0}^{n+d-2l-2l'} \frac{(\frac{e}{1-e})^{j-d} \binom{n}{j-d+2l} \binom{d}{l} \binom{d}{d-l'}}{2^{n+d}} \tag{2.47}
\]
now let \( j' = j + d - 2l' \), that is to say \( j' = d + 2l' = j \), thus
\[
f_1(d) = \sum_{l'=\frac{d}{2}^+}^{d} \sum_{j'=0}^{n+d-2l'} \frac{(\frac{e}{1-e})^{j'-2d+2l'} \binom{n}{j'} \binom{d}{l'} \binom{d}{d-l'}}{2^{n+d}} \tag{2.48}
\]
When \( d \to d + 1 \),
\[
f_1(d+1) = \sum_{l=\frac{(d+1)}{2}^+}^{d+1} \sum_{j=0}^{n+(d+1)-2l-2l} \frac{(\frac{e}{1-e})^{j-2d+2l} \binom{n}{j} \binom{d+1}{l}}{2^{n+(d+1)}} \tag{2.49}
\]
\[
= \sum_{l=(\frac{d+1}{2})^+}^{d+1} \sum_{j=0}^{n+d-2l+1} \frac{(\frac{e}{1-e})^{j-2d+2l} \binom{n}{j} \binom{d+1}{l}}{2^{n+d}} \tag{2.50}
\]
Since \( d \) is odd, \( \left( \frac{d+1}{2} \right)^+ = \left\lceil \frac{d+1}{2} \right\rceil = \frac{d}{2} + 1 \). And when \( e < \frac{1}{2} \), \( \frac{e}{1 - e} < 1 \), so \( (\frac{e}{1 - e})^{(j - 2d + 2l) - 2} > (\frac{e}{1 - e})^{j - 2d + 2l} \). Besides,

\[
\frac{(d+1)}{2} = \binom{d}{l} \frac{d+1}{d+2l} \binom{d}{l},
\]

as long as \( l \geq \left( \frac{d+1}{2} \right) \)

\[
\binom{d}{l} \frac{d+1}{d+2l} \geq \binom{d}{l}.
\]

Use characteristics above, then for \( f_1(d+1) \)

\[
f_1(d+1) = \sum_{l=(2d+1)^+}^{d+1} \sum_{j=0}^{n+d-2l+1} \left( \frac{e}{1 - e} \right)^{(j-2d+2l) - 2} \binom{n}{j} \binom{d+1}{l} \frac{d+1}{2}.
\]

Likewise,

\[
f_2(d+1) = \sum_{l=(2d+1)^+}^{d+1} \sum_{j=0}^{n+d-2l+1} \left( \frac{e}{1 - e} \right)^{(j-2d+2l) - 2} \binom{n}{j} \binom{d+1}{l} \frac{d+1}{2}.
\]

if define \( k = j - 1 \), then

\[
f_2(d+1) = \sum_{l=(2d+1)^+}^{d+1} \sum_{k=-1}^{n+d-2l} \left( \frac{e}{1 - e} \right)^{(k-d)} \binom{n}{k-d+2l} \binom{d+1}{l} \frac{d+1}{2}.
\]

now we have \( f(d+1) = f_1(d+1) + f_2(d+1) > f_1(d) + f_2(d) = f(d) \), which implies the LHS in the (2.19) is increasing function when \( e < \frac{1}{2} \) and \( W(Z) = d \) is an odd integer.

Now assume \( W(Z) = d \) is an even integer. Under this condition, let’s define \( \left\lceil \frac{d}{2} \right\rceil = \left[ \frac{d+1}{2} \right] \) and
\[(d_2) = d - (\frac{d}{2})^+\]. Let \(g(d)\) represents the LHS in the (2.19) this time, then

\[
g(d) = \sum_{l=0}^{d} \sum_{j=0}^{\min(n,n+d-2l)} \frac{e^{-j} \binom{n}{j-d+2l}}{2^{n+d}} \binom{d}{l}
\]

\[
= \sum_{l=0}^{\frac{d}{2}^-} \sum_{j=0}^{d} \frac{e^{-j} \binom{n}{j-d+2l}}{2^{n+d}} \binom{d}{l}
\]

\[
+ \sum_{l=\left(\frac{d}{2}\right)^+}^{d} \sum_{j=0}^{n+d-2l} \frac{e^{-j} \binom{n}{j-d+2l}}{2^{n+d}} \binom{d}{l}
\]

\[
+ \sum_{j=0}^{n} \frac{e^{-j} \binom{n}{j-d+2l}}{2^{n+d}} \binom{d}{l},
\]

let \(g_1(d)\) denotes (2.59), \(g_2(d)\) denotes (2.60) and \(g_3(d)\) denotes (2.61), so \(g(d) = g_1(d) + g_2(d) + g_3(d)\).

Similarly by applying the same transformation method to \(f_1(d)\) we can also have

\[
g_1(d) = \sum_{j=0}^{n+d-2l} \sum_{l=\left(\frac{d}{2}\right)^+}^{d} \frac{e^{-j} \binom{n}{j-d+2l}}{2^{n+d}} \binom{d}{l}.
\]

Note that since \(d\) is an even integer, the \((\frac{d+1}{2})^+ = \lfloor \frac{(d+1)+1}{2} \rfloor = \lfloor \frac{d+1}{2} \rfloor = (\frac{d}{2})^+\). Go through the same process, when \(d \to d + 1\) and \(e < 1/2\),

\[
g_1(d + 1) = \sum_{l=\left(\frac{d+1}{2}\right)^+}^{d+1} \sum_{j=0}^{n+(d+1)-2l} \frac{e^{-j} \binom{n}{j-d+2l+2} \binom{d}{l}}{2^{n+(d+1)}}.
\]

\[
= \sum_{l=\left(\frac{d}{2}\right)^+}^{d+1} \sum_{j=0}^{n+d-2l+1} \frac{e^{-j} \binom{n}{j-d+2l-2} \binom{d+1}{l}}{2^{n+d}}
\]

\[
> \sum_{l=\left(\frac{d}{2}\right)^+}^{d} \sum_{j=0}^{n+d-2l} \frac{e^{-j} \binom{n}{j-d+2l} \binom{d}{l}}{2^{n+d}} = g_1(d),
\]
Similarly, if we define $k = j - 1$,

$$g_2(d + 1) = \sum_{l=(d+1)^+/2}^{d+1} \sum_{j=0}^{n+(d+1)-2l} \frac{n \choose j} \left( \frac{e}{1-e} \right)^{j-(d+1)} \left( \frac{d+1}{2} \right) \left( \begin{array}{c} n \choose j \end{array} \right) \left( \begin{array}{c} d+1 \choose l \end{array} \right) \frac{n}{2n+(d+1)} \right.$$  \hspace{1cm} (2.66)

$$= \sum_{l=\lfloor d/2 \rfloor + 1}^{d+1} (n+d-2l+1) \frac{e}{1-e}^{j-1-d} \left( \frac{n}{j-d+2l} \right) \frac{d+1}{2} \frac{n}{2n+d} \right.$$  \hspace{1cm} (2.67)

$$= \sum_{l=\lfloor d/2 \rfloor + 1}^{d+1} \sum_{k=-1}^{n+d-2l} \frac{(e^{d-k-d})(e^{d-k-d})^{(d+1)/2}}{2n+d} \right.$$  \hspace{1cm} (2.68)

$$> \sum_{l=\lfloor d/2 \rfloor + 1}^{d+1} \sum_{k=0}^{d-n+d-2l} \frac{(e^{d-k-d})(e^{d-k-d})^{(d+1)/2}}{2n+d} = g_2(d).$$  \hspace{1cm} (2.69)

As for the $g_3(d + 1)$:

$$g_3(d + 1) = \sum_{j=0}^{d} \frac{e}{1-e}^{j-d} \left( \frac{n}{j} \right) \left( \frac{d+1}{2} \right) \left( \begin{array}{c} d+1 \choose j \end{array} \right) \frac{n}{2n+d} \right.$$  \hspace{1cm} (2.70)

$$= \sum_{j=0}^{d} \frac{e}{1-e}^{j-d-1} \left( \frac{n}{j} \right) \left( \frac{d+1}{2} \right) \left( \begin{array}{c} d+1 \choose j \end{array} \right) \frac{n}{2n+d} \right.$$  \hspace{1cm} (2.71)

$$> \sum_{j=0}^{d} \frac{e}{1-e}^{j-d} \left( \frac{n}{j} \right) \left( \frac{d+1}{2} \right) \left( \begin{array}{c} d+1 \choose j \end{array} \right) \frac{n}{2n+d} = g_3(d),$$  \hspace{1cm} (2.72)

so we obtain $g(d + 1) = g_1(d + 1) + g_2(d + 1) + g_3(d + 1) > g_1(d) + g_2(d) + g_3(d) = g(d)$, which implies the LHS in the (2.19) is increasing function when $e < 1/2$ and $W(Z) = d$ is an even integer.

In conclusion, the LHS in the (2.19) is an increasing function when $e < 1/2$.

### 2.8 References


CHAPTER 3. ENHANCE BAYESIAN TEST DETECTION PERFORMANCE WITH CRYPTOGRAPHY FOR SHORT LENGTH PACKET

3.1 Overview

In the emerging high-speed wireless networks such as the Internet of Things (IoT) and fifth-generation (5G) system, short messages of size 10 ∼ 20 bytes represent the most common form of traffic between automated sensors, actuators, and controllers Johansson et al. (2015). Since Bayesian test detection depends on bit error rates (BERs) in overhearing channel, sometimes BERs may be not good enough to make Bayesian test meet the system requirement. To enhance the performance Bayesian test performance, especially when BERs are high, in this chapter, we consider attaching a short MAC tag of length $B$ bits to each packet and performing a lightweight cryptographic integrity check on each packet. A drawback of using short MAC is that a false packet may be validated by the verifier: the adversary can randomly pick a hash sequence and get the modified packet validated with probability $2^{-B}$, which may be a high value especially when $B$ is a small number. To address this shortfall, the physical-layer integrity check (Bayesian test) is performed in parallel with the lightweight cryptographic check, followed by combining the two integrity check outcomes, to get the synergy of the two complementary solutions. The goal is to minimize the overhead while improve overall detection accuracy.

3.2 Message Authentication Code

In tradition, the cryptographic hash check is a common solution to ensure data integrity. It is a mathematical algorithm that maps data of arbitrary size to a bit string of a fixed size (called a hash) and is designed to be a one-way function. One popular cryptographic hash check approach in authenticating legality of the received packets is called message authentication code (MAC).
Message authentication code (MAC) algorithm (as Fig. 3.1 shows), sometimes called a keyed (cryptographic) hash function Paar et al. (2010a), accepts as input a secret key and a message \( X \) to be authenticated, and outputs a MAC, \( h(key, X) \). Then, the receiver validates the received packet \( Y \) (from relay) by recomputing the its MAC based on \( Y \) and the key, \( h(key, Y) \), and comparing it to \( h(key, X) \) which is sent to \( D \) over a separate secure channel: if \( h(key, Y) = h(key, X) \) then \( Y \) is considered correct, otherwise, it is considered incorrect.

![Figure 3.1 The message authentication code example.](image)

Let \( N \) denote the number of packets, each of length \( n \) bits, that are checked together using \( B \)-bit MAC in a frame, as illustrated in Fig. 3.2(a). In practice, the payload size \( nN \) is typically chosen much larger than \( B \) in order to make the transmission overhead insignificant. The probability that the adversary randomly picks a hash sequence and gets a modified frame validated by the receiver (verifier) is \( 2^{-B} \). However, the true frame is always validated by the cryptographic check.

A drawback of the cryptographic integrity checking on \( N \) packets is that the detection outcome of \( H_1 \) does not tell which of \( N \) packets in a frame is modified. Therefore, the entire \( N \) packets that are checked together need to be declared modified (\( H_1 \)), if the detection outcome is positive. This causes the true packets that are checked together with false packets to be declared positive (\( H_1 \)) and therefore increases the probability of false alarm.
29

Figure 3.2 Frame structure: (a) cryptographic integrity check, (b) crypto-physical integrity check. \( h(X_i) \) is the MAC tag for \( X_i \) and \( h(X_1, X_2, X_3, \cdots, X_N) \) is the MAC tag for \( X_1, X_2, X_3, \ldots, X_N \).

3.2.1 Probability of False Alarmed and Missed Detection for MAC

Let \( X'_i \) denote the \( i \)th received packet from the relay, \( i = 1, 2, \cdots, N \). Without loss of generality, assume \( X_i \) is the packet of interest. Let \( X' := \{ X'_1, X'_2, X'_3, \ldots, X'_N \} \) and \( X := \{ X_1, X_2, X_3, \ldots, X_N \} \). Then, the probability of false alarm and missed detection on packet \( X_i \) are given by

\[
P_{F,C} = P(h(X') \neq h(X)|X'_i = X_i)
\]

and

\[
P_{M,C} = P(h(X) = h(X)|X'_i \neq X_i),
\]

respectively. Let \( H_0 \) denote the event \( X' = X \), i.e. \( X_1, X_2, X_3, \ldots, X_N \) are not modified by the relay. And Let \( H_1 \) denote the event \( X' \neq X \), i.e. \( X_1, X_2, X_3, \ldots, X_N \) are modified by the relay. Then the probability that a true packet is declared positive (false alarmed) is given by

\[
P_{F,C} = P(h(X') \neq h(X)|X'_i = X_i, H_0)P(H_0|X'_i = X_i)
\]

\[
+ P(h(X') \neq h(X)|X'_i = X_i, H_1)P(H_1|X'_i = X_i)
\]

\[
= 0 \times P(H_0) + (1 - 2^B)P(H_1|X'_i = X_i)
\]

\[
= (1 - 2^{-B})(1 - \beta^{N-1}),
\]
Similarly, the probability that a false packet is declared negative (missed detection) is given by

\[ P_{M,C} = P(h(X) = h(X)|X_i \neq X_i) \]

\[ = P(h(X) = h(X)|X' \neq X) \]

\[ = 2^{-B}. \]  

One observes that the probability of false alarm increases with \( N \) and converges to \( 1 - 2^{-B} \) when \( N \) is large enough. This follows from declaring positive (\( H_1 \)) on all packets in a frame if at least one of them is modified and the this modified packet is detected with probability \( 1 - 2^{-B} \).

### 3.3 Crypto-Physical Integrity Check

With a goal of minimizing the overhead without sacrificing the integrity, a lightweight cryptographic is attached after each packet, as Fig. 3.2(b) illustrates. After packets received, system applies a physical-layer integrity check in parallel with a this lightweight cryptographic integrity check. The block diagram of proposed scheme, called Crypto-Physical Integrity Check, is shown in Fig. 3.3. To get the same amount of overhead per packet for both crypto-physical integrity check and cryptographic check, we will assume \( b = B/N \) in the sequel\(^1\).

---

1. A short MAC of length \( B/N \) can be obtained by truncating the original MAC to the first \( B/N \) bits

![Figure 3.3 Crypto-Physical Integrity Check.](image-url)
3.3.1 MAP Combining Rule

The cryptographic check and physical-layer check (Bayesian test) outcomes are combined under the maximum a posteriori probability (MAP) rule. Let $\hat{H}_C$ and $\hat{H}_P$ denote the cryptographic check outcome and the physical-layer check outcome, respectively. Then, there will be four possible outcomes of $(\hat{H}_C, \hat{H}_P)$: $(H_0, H_0)$, $(H_0, H_1)$, $(H_1, H_0)$, $(H_1, H_1)$. The MAP combining rule is to decide

$$\hat{H}_{CP} = \begin{cases} H_0, & e \geq e_0(b, \beta) \\ H_1, & e < e_0(b, \beta) \end{cases}$$

(3.11)

for some threshold $e_0(b, \beta)$.

For the case of $(\hat{H}_C, \hat{H}_P)$ being $(H_0, H_0)$, $(H_1, H_1)$, the MAP rule yields $\hat{H}_{CP} = H_0$ and $\hat{H}_{CP} = H_1$, respectively. For the case of $(\hat{H}_C, \hat{H}_P)$ being $(H_1, H_0)$, the MAP rule yields $\hat{H}_{CP} = H_1$ because cryptographic check of single won’t produce false alarm event, which makes $P(H_0|\hat{H}_C = H_1, \hat{H}_P = H_0) = 0$. Let $P_{F,P}$ and $P_{M,P}$ be the probability of false alarm (2.23) and missed detection (2.25) for physical-layer check (Bayesian test), respectively. For the case of $(\hat{H}_C, \hat{H}_P)$ being $(H_0, H_1)$ the MAP rule yields

$$\left( \frac{1 - P_{M,P}}{P_{F,P}} \right)^{\hat{H}_{CP} = H_1} \geq \frac{\beta}{1 - \beta},$$

(3.10)

where $\beta = P(F = 0)$. Since $P_{F,P}$ and $P_{M,P}$ are increasing function of the BER between $S$ and $D$, $e$, the MAP rule in (3.10) reduces to

$$\hat{H}_{CP} = \begin{cases} H_0, & e \geq e_0(b, \beta) \\ H_1, & e < e_0(b, \beta) \end{cases}$$

(3.11)

for some threshold $e_0(b, \beta)$ where $[(1 - P_{M,P})/P_{F,P}] = 2^b\beta/(1 - \beta)$. That is, if the overhearing BER is larger than $e_0(b, \beta)$, then the combiner discards the physical-layer check outcome and chooses the cryptographic check outcome, otherwise, the combiner discards the cryptographic check outcome and chooses the physical-layer check outcome. The MAP combining rule is summarized in Table 3.1.
Table 3.1 MAP combining rule.

<table>
<thead>
<tr>
<th>$\hat{H}_C$</th>
<th>$\hat{H}_P$</th>
<th>$\hat{H}_{CP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$</td>
<td>$H_1$</td>
<td>$H_1$</td>
</tr>
<tr>
<td>$H_1$</td>
<td>$H_0$</td>
<td>$H_1$</td>
</tr>
<tr>
<td>$H_0$</td>
<td>$H_1$</td>
<td></td>
</tr>
<tr>
<td>$H_0$</td>
<td>$H_0$</td>
<td>$H_0$</td>
</tr>
</tbody>
</table>

3.3.2 Probability of Detection Error

Table 3.1 shows that the crypto-physical check outcome, $\hat{H}_{CP}$ is equal to the cryptographic check outcome, $\hat{H}_C$ if $e \geq e_0(b, \beta)$. Hence, the probability of missed detection and false alarm of the crypto-physical check are identical to those of the cryptographic detection outcome. That is, for $e \geq e_0(b, \beta)$

$$P_{M,CP} = 2^{-b}$$ (3.12)
$$P_{F,CP} = 0.$$ (3.13)

However, for $e < e_0(b, \beta)$,

$$P_{M,CP} = P(\hat{H}_C = H_0, \hat{H}_P = H_0|H_1)$$ (3.14)
$$= 2^{-b}P_{M,P}$$ (3.15)

and

$$P_{F,CP} = 1 - P(\hat{H}_C = H_0, \hat{H}_P = H_0|H_0)$$ (3.16)
$$= 1 - (1 - 0) \times (1 - P_{F,P})$$ (3.17)
$$= P_{F,P}.$$ (3.18)

Therefore, the probability of detection error, $P_{F,CP}\beta + P_{F,CP}(1 - \beta)$, is given by

$$P_{E,CP} = \begin{cases} 
2^{-b}(1 - \beta), & e \geq e_0(b, \beta) \\
2^{-b}P_{M,P} + 2^{-b}P_{M,P}(1 - \beta), & e < e_0(b, \beta)
\end{cases}$$ (3.19)

Remark: It can be seen that the MAP combiner chooses the cryptographic check outcome if the overhearing channel quality is below a certain threshold (i.e. $e \geq e_0(b, \beta)$). But if the overhearing channel quality is above this threshold, the crypto-physical check is selected. This choice is made to balance the trade-offs between missed detection and false alarm rates.
quality is above the threshold (i.e. $e < e_0(b,\beta)$) then it performs the logical OR combining of the cryptographic check outcome and the physical-layer check outcome. For the latter case the probability of missed detection can be reduced by a factor of $2^{-B/N}$ over the physical-layer integrity check, while the probability of false alarm remains the same.

3.4 Numerical Results

![Graph showing the average probability of detection error versus the bit error rate (BER) between the source and the destination, $e$, for different integrity checking schemes. The optimum detection threshold that minimizes the average probability of detection error, which is found by numerical search, is used for the physical-layer check and the crypto-physical check. It can be observed that the physical-layer integrity check yields a lower probability of detection error than the cryptographic]

Fig. 3.4 shows the average probability of detection error versus the bit error rate (BER) between the source and the destination, $e$, for different integrity checking schemes. The optimum detection threshold that minimizes the average probability of detection error, which is found by numerical search, is used for the physical-layer check and the crypto-physical check. It can be observed that the physical-layer integrity check yields a lower probability of detection error than the cryptographic
integrity check if $e$ is smaller than a threshold. However, the physical integrity check performs worse as $e$ increases. This is because the physical integrity check relies on the overheard packet $Y_s$ whose reliability depends on $e$. The crypto-physical integrity check takes the best of the physical integrity check and the cryptographic check, and yields a smaller probability of detection error than the physical-layer check and cryptographic check for all $e$. Fig. 3.5 illustrates the average

Figure 3.5 Probability of detection error versus latency (bits); $N = 1$, $\beta = 0.9$, $e = 0.05$, (127,64) BCH.

probability of detection error versus the latency (bits) for different integrity checking schemes. The latency (x-axis) is varied by varying the MAC length $B$. The optimum detection threshold that minimizes the average probability of detection error is used for the crypto-physical check. It is observed that the crypto-physical integrity check yields a lower probability of detection error than the cryptographic integrity check for a given latency, and that the improvement is more significant
at lower latency. However, if the latency is allowed to be sufficiently large, they provide the same probability of detection error.

3.5 Conclusion

In this chapter, crypto-physical integrity check technique that provides low-latency integrity assurance in wireless relay networks is developed. The scheme employs a physical-layer integrity check in parallel with a lightweight cryptographic integrity check, followed by combining the two integrity check outcomes, to get a synergy of the two complementary solutions. The physical-layer integrity check exploits overheard packet as a baseline and applies the Bayesian test in detecting integrity violation. The maximum a posteriori probability principle is applied in combining the lightweight cryptographic check outcome with the physical-layer check outcome such that the probability of detection error is minimized. We found that the crypto-physical integrity check can provide a lower probability of detection error than the cryptographic integrity check and that the improvement is more significant at low latency regime. The proposed technique will benefit a wide variety of low-latency applications where information needs to be communicated in a timely and secure manner under resource constraints.

3.6 References


CHAPTER 4. PHYSICAL LAYER MAP DETECTION OF FALSE DATA INJECTION IN WIRELESS DATA AGGREGATION NETWORKS

4.1 Introduction

In upcoming wireless system Durisi et al. (2016a); Bennis et al. (2018), notably Internet of Things (IoT) Durisi et al. (2016b); Azari et al. (2017) and the fifth-generation (5G) system, the traffic type tend to be short and massive. Short data packets on the order of tens of bits or lower will represent the most common form of traffic in emerging wireless applications. A core requirement in such applications is that the data integrity needs to be maintained as the data move through the network. Data integrity is the property by which data have not been changed or modified in an unauthorized or accidental manner. A common approach is to append a cryptographic message authentication code (MAC) on the data packet at the higher layer, which provides a piece of information about whether the data has been changed in transit Paar et al. (2010b). Such an approach, however, may incur a significant overhead for short packets, as the MAC length may approach or even exceed the data length. Also devices in those system are usually small, low cost and resources-limited, that is to say, applying a complete cryptographic authentication code is quite difficult Trappe et al. (2015).

In this chapter, we study the MAP detection of false data in short packets which are collected from multiple sources through a relay. The MAP test minimizes the probability of detection error and does not incur any overhead. The adversary (relay) modifies the received packets from the source nodes and forwards modified packets to the destination. The destination exploits the information that it overhears from the source nodes as a reference in verifying the correctness of the received packets. The fundamental challenge is that the reference information provided by the heard packet is erroneous (noisy) due to the channel impairment between the source and the destination. We seek to mitigate the effects of unreliable reference information on the accuracy of
the MAP test by leveraging the *multiplicity* of source nodes. We show that the average probability of detection error decreases *exponentially* as the number of source nodes increases, *regardless* of the reliability of the reference information. Hence, a powerful integrity check mechanism can be constructed at the physical layer if the number of source nodes is large enough.

### 4.2 System Model

We consider a wireless relay network in which $N$ sources, $S_1, S_2, \cdots, S_N$, send their packets $X_1, X_2, \cdots, X_N$, each of length $n$ bits, to a common node (destination), $D$, through a relay, $R$. The network model is illustrated in Fig. 4.1(a). An example of such network is clustered Internet of Things (IoT) where devices are clustered into several groups with each group having its own gateway Abhishek et al. (2018). We assume that each packet $X_i$ is an $(n,k)$ binary linear codeword and that $R$ operates as “decode-and-forward” relay. As such, “packet” and “codeword” will be used interchangeably in this paper.

![Figure 4.1](image.png)

Figure 4.1 (a) Wireless relay network model and (b) information flow of $X_i$. $Y_{s,i}$ is the overheard packet from the $i$th source $S_i$ and $Y_{r,i}$ is the received packet from the relay $R$.

In phase 1, each source node sends its packet to $R$. The received (decoded) word at $R$ from $S_i$ can be expressed as $X_{r,i} = X_i \oplus E_{r,i}$, where $E_{r,i}$ denotes the error vector between $S_i$ and $R$ and $\oplus$ denotes the bit-wise modulo-2 addition. Due to the broadcast nature of the wireless medium,
$D$ can overhear the transmitted packets from the source nodes (dotted lines in Fig. 4.1). The overheard packet from $S_i$ can be expressed as

$$Y_{s,i} = X_i \oplus E_{s,i}, \tag{4.1}$$

where $E_{s,i}$ denotes the channel error vector between $S_i$ and $D$. The effect of channel fading, noise, and interference is captured in $E_{s,i}$. Since the overhearing channel errors are typically not correctable, $E_{s,i}$ denotes the raw channel error vector before decoding. We assume that bit errors occur randomly by interleaving and deinterleaving.

In phase 2, $R$ modifies a fraction $\alpha$ of the received packets. For example, if $N_1$ packet are modified among the $N$ received packets, then $\alpha = N_1/N$. The data modification strategy that minimizes the likelihood of the modification being detected by the integrity check (or maximizes the likelihood of bypassing the integrity check) will be discussed in Section 4.3. The modified packet can be expressed as $X'_{r,i} = X_{r,i} \oplus F_{r,i}$, where $F_{r,i}$ denotes the false data that is injected by the relay. Then, $X'_{r,i}$ is sent to $D$. The received (decoded) word at $D$ from $R$ can be expressed as

$$Y_{r,i} = X'_{r,i} \oplus E_{d,i} = X_i \oplus F_i, \tag{4.2}$$

where $E_{d,i}$ denotes the channel error vector between $R$ and $D$ and

$$F_i := Y_{r,i} \oplus X_i \tag{4.3}$$

$$= E_{r,i} \oplus F_{r,i} \oplus E_{d,i} \tag{4.4}$$

is the compound error vector between $X_i$ and $Y_{r,i}$. The main difference between $(E_{r,i}, E_{d,i})$ and $F_{r,i}$ is that, the former are error vectors introduced by the nature (fading, noise), while the latter is an ‘intentional’ well-designed false data injected by the adversary. Therefore, $F_i \neq 0$ means $Y_{r,i} \neq X_i$. The information flow from $S_i$ to $D$ is illustrated in Fig. 4.1(b). An important property of the linear code is that linear combination of any codewords is also a codeword Lin and Costello (2004). Hence, the compound error vector $F_i$ should be a codeword as $Y_{r,i}$ is the decoded word and $X_i$ is the transmitted codeword. It makes the traditional error detection codes hopeless for detecting $F_i$. We assume that $\{F_{r,i}\}$ are independent and that the packet errors $\{E_{r,i}, E_{d,i}, E_{s,i}\}$
occur independently. Therefore, $F_i$’s are independent. The destination $D$ verifies the correctness of the received packets $Y_r := (Y_{r,1}, Y_{r,2}, \cdots, Y_{r,N})$ by employing the noisy (erroneous) overheard packets $Y_s := (Y_{s,1}, Y_{s,2}, \cdots, Y_{s,N})$ as a reference. The adversary’s objective is to modify the packets at $R$, where the packets are merged from multiple source nodes, without being detected by $D$. For convenience, Table 4.1 lists the notations and terms used in this chapter.

Table 4.1  List of notations and terms.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_j$</td>
<td>$j$th codeword</td>
</tr>
<tr>
<td>$D$</td>
<td>Destination</td>
</tr>
<tr>
<td>$d_{min}$</td>
<td>Minimum distance between codewords</td>
</tr>
<tr>
<td>$e_i$</td>
<td>BER between $S_i$ and $D$</td>
</tr>
<tr>
<td>$E_{d,i}$</td>
<td>Channel error vector between $R$ and $D$</td>
</tr>
<tr>
<td>$E_{r,i}$</td>
<td>Channel error vector between $S_i$ and $R$</td>
</tr>
<tr>
<td>$E_{s,i}$</td>
<td>Channel error vector between $S_i$ and $D$</td>
</tr>
<tr>
<td>$F_{r,i}$</td>
<td>Falsely injected packet by relay</td>
</tr>
<tr>
<td>$F_i$</td>
<td>$E_{r,i} \oplus F_{r,i} \oplus E_{d,i}$</td>
</tr>
<tr>
<td>MAC</td>
<td>Message authentication code</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of source nodes</td>
</tr>
<tr>
<td>$n$</td>
<td>Packet length</td>
</tr>
<tr>
<td>$R$</td>
<td>Relay</td>
</tr>
<tr>
<td>$S_i$</td>
<td>$i$-th source</td>
</tr>
<tr>
<td>$W(F_i)$</td>
<td>Hamming weight of $F_i$</td>
</tr>
<tr>
<td>$X_i$</td>
<td>Transmitted packet of $S_i$</td>
</tr>
<tr>
<td>$\bar{X}_{r,i}$</td>
<td>Decoded word by relay</td>
</tr>
<tr>
<td>$X'_{r,i}$</td>
<td>Modified packet by relay</td>
</tr>
<tr>
<td>$Y_{r,i}$</td>
<td>Decoded word by destination</td>
</tr>
<tr>
<td>$Y_{s,i}$</td>
<td>Overheard packet by destination</td>
</tr>
<tr>
<td>$Z_i$</td>
<td>$F_i \oplus E_{s,i}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$P(F_i \neq 0)$</td>
</tr>
<tr>
<td>$\mathbf{F}$</td>
<td>$(F_1, F_2, \cdots, F_N)$</td>
</tr>
<tr>
<td>$\mathbf{X}$</td>
<td>$(X_1, X_2, \cdots, X_N)$</td>
</tr>
<tr>
<td>$\mathbf{Y}_r$</td>
<td>$(Y_{r,1}, Y_{r,2}, \cdots, Y_{r,N})$</td>
</tr>
<tr>
<td>$\mathbf{Y}_s$</td>
<td>$(Y_{s,1}, Y_{s,2}, \cdots, Y_{s,N})$</td>
</tr>
<tr>
<td>$\mathbf{Z}$</td>
<td>$(Z_1, Z_2, \cdots, Z_N)$</td>
</tr>
</tbody>
</table>
4.3 The Physical-Layer MAP Detection

In this section, we develop the maximum a posteriori probability (MAP) test for detecting false data in the received packets $Y_r$ by employing the overhead packet $Y_s$ as a reference. The goal is to determine whether $Y_s = X$ or not, i.e. whether $F = (F_1, F_2, \cdots, F_N)$ equals 0 or not.

4.3.1 Sufficient Statistic

Recall the (2.10) it is already been proved that

$$I(F_i; Y_{s,i}, Y_{r,i}) = I(F_i; Z_i)$$ (4.5)

where $I(x; y)$ denotes the mutual information and

$$Z_i = Y_{s,i} \oplus Y_{r,i}$$ (4.6)

$$= F_i \oplus E_{s,i}.$$ (4.7)

Usually, $F_{r,i}$ won’t provide any information for $F_{r,j}$ if $j \neq i$, so $\{F_{r,i}\}$ are independent. Also, since packets errors $\{E_{r,i}, E_{d,i}, E_{s,i}\}$ are caused by channel noise, they occur independently. Therefore, $F_i = F_{r,i} \oplus E_{r,i} \oplus E_{d,i}$ are independent for $i = 1, 2, \cdots, N$, which results in

$$H(F) = \sum_{i=1}^{N} H(F_i),$$ (4.8)

and

$$H(F|Y_r, Y_s) = \sum_{i=1}^{N} H(F_i|Y_r, Y_s).$$ (4.9)

Since $Y_{r,j}$ and $Y_{s,j}$ provide no information about $F_i$ if $j \neq i$, we obtain

$$H(F_i|Y_r, Y_s) = H(F_i|Y_{r,i}, Y_{s,i}).$$ (4.10)
From (4.5), (4.8), (4.9) and (4.10), we obtain

\[ I(F; Y_r, Y_s) = H(F) - H(F|Y_r, Y_s) \]  
\[ = \sum_{i=1}^{N} [H(F_i) - H(F_i|Y_{r,i}, Y_{s,i})] \]  
\[ = \sum_{i=1}^{N} I(F_i; Y_{r,i}, Y_{s,i}) \]  
\[ = \sum_{i=1}^{N} I(F_i; Z_i) \]  
\[ = I(F; Z). \]  

That is, \( Z = Y_s \oplus Y_r = (Z_1, Z_2, ..., Z_N) \) contains all the information necessary to make a decision on \( F \) given \( Y_r \) and \( Y_s \). Therefore, \( Z \) is called the sufficient statistic.

### 4.3.2 General MAP Test

Our approach to check the correctness of \( Y_r \), i.e. determine whether \( F = 0 \) (\( H_1 \)) or \( F \neq 0 \) (\( H_0 \)), builds on the maximum a posteriori probability (MAP) rule:

\[ P(F \neq 0|Z) \overset{H_1}{\underset{H_0}{\gtrless}} P(F = 0|Z), \]  
\[ (4.16) \]

which minimizes the average probability of detection error essentially. Since \( P(F \neq 0|Z) + P(F = 0|Z) = 1 \), (4.16) can be expressed as

\[ P(Z) \overset{H_1}{\underset{H_0}{\gtrless}} 2P(Z|F = 0)P(F = 0). \]  
\[ (4.17) \]

\( F_i \)'s and \( E_{s,i} \)'s are independent, so are \( Z_i \)'s. the (4.17) can also be expressed as

\[ \prod_{i=1}^{N} P(Z_i) \overset{H_1}{\underset{H_0}{\gtrless}} 2 \prod_{i=1}^{N} [P(Z_i|F_i = 0)P(F_i = 0)]. \]  
\[ (4.18) \]

We employ \( C_j \), where \( j = 0, 1, \cdots, 2^k - 1 \), to denote the \( j \)th codeword in codebook, and \( e_i \), where \( i = 1, 2, \cdots, N \), to denote the bit error rate (BER) of overhearing channel between \( S_i \) and \( D \). Since
the bit errors occur independently by interleaving and deinterleaving, we obtain

\[ P(Z_i) = \sum_{j=0}^{2^k-1} P(Z_i|F_i = C_j)P(F_i = C_j) \]
\[ = \sum_{j=0}^{2^k-1} e_i^{W(Z_i \oplus C_j)}(1 - e_i)^{n-W(Z_i \oplus C_j)}P(F_i = C_j), \tag{4.19} \]

and

\[ P(Z_i|F_i = 0) = e_i^{W(Z_i)}(1 - e_i)^{n-W(Z_i)}, \tag{4.20} \]

where \( W() \) is Hamming Weight calculator. Therefore, the MAP test in (4.18) is given by

\[ \sum_{i=1}^{N} \ln \left[ 1 + \frac{1}{1 - \alpha} \sum_{j=1}^{2^k-1} \left( \frac{e_i}{1 - e_i} \right)^{W(Z_i \oplus C_j) - W(Z_i)} P(F_i = C_j) \right] \overset{H_1}{\gtrless} \ln 2, \tag{4.22} \]

where \( \alpha = P(F_i \neq 0) \). If we let \( C_{ij} \) denote the subvector of \( C_j \) composed of bits where \( Z_i \) is 1, for instance in Fig. 4.2, \( C_j = 1010101 \) and \( Z_i = 1111000 \), so \( C_{ij} = 1010 \).

Then we have

\[ W(Z_i \oplus C_j) - W(Z_i) = W(C_j) - 2W(C_{ij}), \tag{4.23} \]

so the (4.22) is equal to

\[ \sum_{i=1}^{N} \ln \left[ 1 + \frac{1}{1 - \alpha} \sum_{j=1}^{2^k-1} \left( \frac{1 - e_i}{e_i} \right)^{2W(C_{ij}) - W(C_j)} P(F_i = C_j) \right] \overset{H_1}{\gtrless} \ln 2. \tag{4.24} \]
Noted from Fig. 4.2 that $W(C_{ij})$ depends on $W(C_j)$. If we let $d = W(C_j)$ and $x = W(C_{ij})$, since $C_j$’s are codewords, the minimum hamming weight is equal to the minimum distance, $d_{\text{min}}$, and the maximum hamming weight is equal to bit length, $n$, the (4.24) yields to

$$
\sum_{i=1}^{N} \ln \left[ 1 + \frac{1}{1 - \alpha} \sum_{d=d_{\text{min}}}^{n} \left( \frac{W(Z_i)}{x} \sum_{x=0}^{W(Z_i)} \left( \frac{1 - e_i}{e_i} \right)^{2x-d} P_{x|d}(x) P_d(d) \right) \right] \overset{H_1}{\gtrless} \ln 2, \quad (4.25)
$$

where the probability of $x$ given $d$ is given by

$$
P_{x|d}(x) = \left( \frac{W(Z_i)}{x} \right) \left( \frac{d}{n} \right)^x \left( \frac{1 - d}{n} \right)^{W(Z_i)-x}, \quad (4.26)
$$

this is because if the number of codewords $2^k \gg 1$, so the position of 'bit-1' in $C_j$ can be measured as random. And $P_d(d)$ is defined the probability that adversary chooses $F_i$ which has the Hamming weight distance equals to $d$, i.e.

$$
P_d(d) = \sum_{j=1}^{2^k-1} P(F_i = C_j) \cdot \mathbf{1}(W(C_j) = d), \quad (4.27)
$$

where $\mathbf{1}()$ is indicator function and $\sum_{d=d_{\text{min}}}^{n} P_d(d) = \alpha$. Because of

$$
\sum_{x=0}^{W(Z_i)} \left( \frac{1 - e_i}{e_i} \right)^{2x} P_{x|d}(x) = \left[ \left( \frac{1 - e_i}{e_i} \right)^2 - 1 \right] \frac{d}{n} + 1 \right]^{W(Z_i)}, \quad (4.28)
$$

by inserting (4.26) and (4.27) back to (4.25), the general physical-layer MAP detection take the form as

$$
\sum_{i=1}^{N} \ln \left[ 1 + \frac{1}{1 - \alpha} \sum_{d=d_{\text{min}}}^{n} \left[ \left( \frac{1 - e_i}{e_i} \right)^2 - 1 \right] \frac{d}{n} + 1 \right]^{W(Z_i)} \left( \frac{e_i}{1 - e_i} \right)^d P_d(d) \right] \overset{H_1}{\gtrless} \ln 2. \quad (4.29)
$$

One thing should be aware that, even though the exact $P_d(d)$ is almost impossible to acquire in practice due to the fact that $P(F_i = C_j)$ is a secret only known to adversary, the distribution, $P_d(d)$, still has to be pre-determined by detector in order to perform physical-layer MAP detection (4.29) in practice. The strategy of how to decide $P_d(d)$ will be explained in Section 4.3.3

4.3.3 The Best Strategy for Adversary’s in choosing $F_i$

In general, for any communication system which applies detection mechanism, if an adversary wants to send falsified message successfully, the smartest way should always try to make his falsified
message won’t be detected as being modified (decided as $H_1$ by detector) with highest chance. That is to say, in order to against our physical-layer MAP detection, the adversary should carefully choose the non-zero $F_i$ such that the probability of being decided as $H_1$ is as low as possible.

As discussed in previous subsection, $P_d(d)$ in (4.29) is a parameter pre-determined by detector, not by adversary. That indicates that, in the view of adversary, the detector’s observation, $W(Z_i)$, is the only input parameter in (4.29) he can have an influence. And we will show in this subsection that, no matter which type of distribution $P_d(d)$ is applied by physical-layer MAP detector, an smart adversary should always choose the non-zero codeword $C_j$ where its $W(C_j)$ is equal to $d_{\min}$, if he wants to maximize the probability that his false data injected packets is missed detected. Therefore it is in the adversary’s interest to change bits such that the Hamming weight of false injection vector $W(F_i)$ is minimized. On the other side, since less bit changed results in less damage to system, this property also constraints damage may caused by adversary. It can be seen this property keep consistent with our previous research in Liu et al. (2018), where the detection accuracy improves if adversary modifies more bits.

Since the signal noise ratio (SNR) must be positive in practice, that is to say any bit error rate, $e_i$, must be less than 0.5, easily we have

$$
\left(\frac{1-e_i}{e_i}\right)^2 - 1 \frac{d}{n} + 1 > 1.
$$

(4.30)

Then it can be infer from (4.29) that its LHS will become larger if $W(Z_i)$ increases, which means a larger LHS of (4.29) results in higher chance to decide $H_1$ and smaller LHS of (4.29) results in lower chance to decide $H_1$. In conclusion, if adversary perform falsified data injection on $i$th packet, the best strategy for him is choosing the non-zero $F_i$ such that the possible value of $W(Z_i)$ is close to zero as much as possible, and by doing so, adversary could maximize the probability of miss detection.

To achieve this goal, adversary has to pick the non-zero codeword whose Hamming weight is $d_{\min}$. The reason is shown as follows: no matter which codeword is picked (recall $F_i = C_j$ and $d = W(C_j)$), the $F_i$ can always be divided into two subvectors like Fig. 4.3 shows. The first subvector is $F_{i1}$, which composed of all the 'bits-1' in $F_i$, with length of $d$ and the other subvector is
Figure 4.3 The structure of $F_i$

$F_{i0}$, which composed of all the 'bits-0' in $F_i$, with length of $n - d$. Since $Z_i = F_i \oplus E_{s,i}$ and the noise equivalently affects (with bit error rate $e_i$) all the bits in $F_{i1}$ and $F_{i0}$ independently, if we define $E_{i1}$ be the subvector of $E_{s,i}$ which is composed of the entries in the positions where $F_i$ is 1 and $E_{i0}$ be the subvector of $E_{s,i}$ which is composed of the entries in the positions where $F_i$ is 0, then the $W(Z_i)$ can also be divided into two subvectors like $F_i$: one is $(F_{i0} \oplus E_{i0})$, where $W(F_{i0} \oplus E_{i0}) \sim \text{Binom}(n - d, e_i)$. The other one is $(F_{i1} \oplus E_{i1})$, where $W(F_{i1} \oplus E_{i1}) \sim \text{Binom}(d, 1 - e_i)$. Also they satisfy

$$W(Z_i) = W(F_{i0} \oplus E_{i0}) + W(F_{i1} \oplus E_{i1}).$$

(4.31)

From Binomial distribution property, easily we obtain the variance of $W(Z_i)$ equals $ne_i(1 - e_i)$, which is constant. And its mean value equals to $ne_i + d(1 - 2e_i)$, which is minimized if $d = d_{\text{min}}$ (because $e_i \leq 0.5$). Therefore, in order to maximize the probability of miss detection, the length of $(F_{i1} \oplus E_{i1})$, $d$, has to be the shortest, i.e. $F_i$ can only be the codeword $(C_j)$ where $W(C_j) = d_{\text{min}}$. The reason behind it is because the subvector $(F_{i1} \oplus E_{i1})$ has higher probability to generate 'bit-1' than the subvector $(F_{i0} \oplus E_{i0})$ does, and their length summation of two subvectors equals to a constant $n$. That is to say, the more transmitted bits are modified, the higher chance this false
data will be detected.

Noted that if $F_i = 0$, $W(Z_i) = W(E_{s,i})$, which is purely controlled by the overhearing noise vector, which means the probability of false alarm can not be influenced by adversary. Hence, by choosing $W(F_i) = d_{\min}$, adversary successfully achieves the goal in maximizing the probability of detection error. On contrary, in the view of detector, he should choosing the $P_d(d)$ such that the probability of detection error can be minimized. Hence, the "game" between detector and adversary becomes a typical MiniMax problem:

$$\min P_E = \min_{P_d(d)} \left\{ \max_{W(Z_i)} \left[ P_E(P_d(d), W(Z_i)) \right] \right\}$$

(4.32)

where $P_E$ is the probability of detection error, $P_d(d)$ and $W(Z_i)$ are its parameters which are controlled by detector and adversary respectively. Since the solution to this MiniMax problem is choosing

$$P_d(d) = \begin{cases} 
\alpha, & \text{if } d = d_{\min}; \\
0, & \text{otherwise.}
\end{cases}$$

(4.33)

By applying solution (4.33) into (4.29), the Physical-Layer MAP Detection is equivalent to

$$\sum_{i=1}^{N} \ln \left[ 1 + \frac{\alpha}{1-\alpha} \left( \frac{e_i}{1-e_i} \right) d_{\min} \left[ \left( \frac{1-e_i}{e_i} \right)^2 - 1 \right] \frac{d_{\min}}{n} + 1 \right] W(Z_i) \right]_{H_1 \leq H_0} \ln 2.$$  

(4.34)

When $N = 1$, (4.34) reduces to

$$W(Z_1) \overset{H_1}{\underset{H_0}{\geq}} \log_\beta \left( \frac{1-\alpha}{\alpha} \right) - d_{\min} \log_\beta \left( \frac{e_i}{1-e_i} \right),$$

(4.35)

where base $\beta$ is given by

$$\beta = \left( \frac{1-e_i}{e_i} \right)^2 - 1 \frac{d_{\min}}{n} + 1.$$ 

(4.36)

This also matches with the our result in Liu et al. (2018) that $W(Z_1) \overset{H_1}{\underset{H_0}{\geq}} \delta$, where $\delta$ is the detection threshold can be chosen such that the probability of detection errors are minimized.

**Remark**

The MAP test in (4.34) requires the information of $\alpha$. Consider it as the secret only known to attacker, the maximum likelihood test corresponding to the case where $\alpha = 0.5$. 


4.4 Probability of Detection Error: Physical-Layer MAP Test

The effectiveness of the physical-layer MAP test can be measured by two metrics: the probability of false alarm which denotes the probability that the detector falsely decides that the received packet $Y_r$ contains modified packets but actually it is not, and the probability of missed detection which denotes the probability that the detector falsely decides that $Y_r$ contains no modified packets when it is not. For convenience, let

$$a_i = \frac{\alpha}{1 - \alpha} \left( \frac{e_i}{1 - e_i} \right)^{d_{\min}}$$

$$b_i = \left[ \left( \frac{1 - e_i}{e_i} \right)^2 - 1 \right] \frac{d_{\min}}{n} + 1,$$

then the detector’s decision statistic is given by

$$Y(Z) = \sum_{i=1}^{N} \ln \left[ 1 + a_i b_i^{W(Z_i)} \right],$$

which will be used to perform physical-layer MAP test in (4.34), i.e. the physical-layer MAP test reduces to

$$Y(Z) \overset{H_1}{\gtrsim} \ln 2.$$  

Clearly, the event of false alarm occurs when $Y(Z) \geq \ln 2$ given $F = 0$, while the event of missed detection occurs when $Y(Z) < \ln 2$ given $F \neq 0$.

Unfortunately, the exact expression for the probability of false alarm and missed detection (which are provided in Appendix 4.8) is too complicate to get insight of how it behaves. Also due to the difficulty in integral for its probability mass function (PMF) given $F \neq 0$ and numerical evaluation. Instead of directly doing analysis based on real probability of false alarm and missed detection expression, we are seeking to do analysis based on a simple closed-form approximation for the probability of false alarm and missed detection: Since $Y(Z)$ is a summation of $N$ independent random variables, whose $N$ is consider as very large number in our targeting network, it can be viewed as Gaussian distributed by the central limit theorem.
4.4.1 Probability of False Alarm

Since \( W(Z_i) = W(E_{s,i}) \) given \( F_i = 0 \) and \( W(E_{s,i}) \) is binomially distributed with parameters \( n \) and \( e_i \), i.e. \( W(E_{s,i}) \sim \text{Binom}(n, e_i) \), the probability mass function of \( W(Z_i) \) given \( F_i = 0 \) is given by

\[
f_0(w) := P(W(Z_i) = w | F_i = 0) = \binom{n}{w} e_i^w (1 - e_i)^{n-w},
\]

for \( 0 \leq w \leq n \). Hence, the mean and variance of \( Y(Z) \) given \( F = 0 \) are given by

\[
\mu_0 = \sum_{i=1}^{N} E \left[ \ln \left(1 + a_i b_i^{W(Z_i)}\right) | F = 0 \right] = \sum_{i=1}^{N} \left[ \sum_{w=0}^{n} \ln(1 + a_i b_i^w) f_0(w) \right]
\]

and

\[
\sigma_0^2 = \sum_{i=1}^{N} E \left[ \ln \left(1 + a_i b_i^{W(Z_i)}\right) | F = 0 \right]^2 - \sum_{i=1}^{N} E^2 \left[ \ln \left(1 + a_i b_i^{W(Z_i)}\right) | F = 0 \right] = \sum_{i=1}^{N} \left[ \sum_{w=0}^{n} \ln(1 + a_i b_i^w)^2 f_0(w) \right] - \sum_{i=1}^{N} \left[ \sum_{w=0}^{n} \ln(1 + a_i b_i^w) f_0(w) \right]^2,
\]

respectively. Hence, the probability of false alarm can be approximated by

\[
P_F = P(Y(Z) > \ln 2 | F = 0) \approx Q \left( \frac{\ln 2 - \mu_0}{\sigma_0} \right),
\]

where

\[
Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du
\]

is the tail distribution function of the standard normal distribution.

4.4.2 Probability of Missed Detection

Since we have shown that \( W(Z_i) = W(F_{i1} \oplus E_{i1}) + W(F_{i0} \oplus E_{i0}) \), \( W(F_{i0} \oplus E_{i0}) \sim \text{Binom}(n - W(F_i), e_i) \) and \( W(F_{i1} \oplus E_{i1}) \sim \text{Binom}(W(F_i), 1 - e_i) \) in Section 4.3.3, then the probability mass
function of $W(Z_i)$ given $F_i \neq 0$ is given by

$$f_1(w) := P(W(Z_i) = w | F_i \neq 0)$$

$$= P(W(F_{i0} \oplus E_{i0}) + W(F_{i1} \oplus E_{i1}) = w)$$

$$= \sum_{m=0}^{w} \binom{W(F_i)}{m} \binom{n - W(F_i)}{w - m} e_i^{W(F_i) + w - 2m} (1 - e_i)^{n - (W(F_i) + w - 2m)},$$

(4.50)

for $0 \leq w \leq n$. Note that $f_1(w) = f_0(w)$ if $W(F_i) = 0$.

Let $M_l$ denote the $l$th non-empty subset of $\{1, 2, \cdots, N\}$ that represents the index of modified packets, where $l = 1, 2, \cdots, 2^N - 1$, and $M_l^c$ denote the complement of $M_l$. For example, $M_l = \{2, 5\}$ denotes the event that the second and the fifth packet are modified. By default, we define $M_0 = \emptyset$ to indicate case no packets are been modified. Since we assume each packet is modified independently with probability $\alpha$, we obtain

$$P(M_l) = \alpha^{|M_l|} (1 - \alpha)^{N - |M_l|},$$

(4.52)

where $|M_l|$ denotes the cardinality of $M_l$. Then, the mean and variance of $Y(Z)$ given $M_l$ are given by

$$\mu_l = E[Y(Z) | M_l]$$

$$= \sum_{i \in M_l} E\left[ \ln \left( 1 + a_i b_i^{W(F_{i} \oplus E_{i}^{w})} \right) \right] + \sum_{i \in M_l^c} E\left[ \ln \left( 1 + a_i b_i^{W(E_{i}^{w})} \right) \right]$$

(4.53)

$$= \sum_{i \in M_l} \left[ \sum_{w=0}^{n} \ln(1 + a_ib_i^w)f_1(w) \right] + \sum_{i \in M_l^c} \left[ \sum_{w=0}^{n} \ln(1 + a_ib_i^w)f_0(w) \right]$$

(4.54)

and

$$\sigma_l^2 = E[Y^2(Z) | M_l] - E^2[Y(Z) | M_l]$$

$$= \sum_{i \in M_l} \left[ \sum_{w=0}^{n} [\ln(1 + a_ib_i^w)]^2 f_1(w) \right] - \sum_{i \in M_l} \left[ \sum_{w=0}^{n} \ln(1 + a_ib_i^w)f_1(w) \right]^2$$

$$+ \sum_{i \in M_l^c} \left[ \sum_{w=0}^{n} [\ln(1 + a_ib_i^w)]^2 f_0(w) \right] - \sum_{i \in M_l^c} \left[ \sum_{w=0}^{n} \ln(1 + a_ib_i^w)f_0(w) \right]^2,$$

(4.55)
respectively. Hence, the probability of missed detection given that the packets in the subset $M_l$ are modified can be approximated by

$$P_{M,l} = P(Y(Z) < \ln 2 | M_l)$$

$$\simeq Q \left( \frac{\mu_l - \ln 2}{\sigma_l} \right) \quad (4.58)$$

$$\simeq Q \left( \frac{\mu_l - \ln 2}{\sigma_l} \right) \quad (4.59)$$

### 4.4.3 Probability of Detection Error

The average probability of detection error is given by

$$P_{E,P} = \frac{2}{N-1} \sum_{l=1}^{2N-1} P_{M,l}P(M_l) + P_F P(M_0)$$

$$\simeq \frac{2}{N-1} \sum_{l=1}^{2N-1} Q \left( \frac{\mu_l - \ln 2}{\sigma_l} \right) \alpha^{|M_l|} (1 - \alpha)^{N-|M_l|} + Q \left( \frac{\ln 2 - \mu_0}{\sigma_0} \right) (1 - \alpha)^N. \quad (4.60)$$

If we let $T_j$ denote the set of $M_l$ whose $|M_l| = j$, i.e. $T_j = \{ M_l : |M_l| = j \}$, the (4.60) can also be approximated by

$$P_{E,P} \simeq \sum_{j=1}^{N} \left[ \sum_{l \in T_j} Q \left( \frac{\mu_l - \ln 2}{\sigma_l} \right) \right] \alpha^j (1 - \alpha)^{N-j} + Q \left( \frac{\ln 2 - \mu_0}{\sigma_0} \right) (1 - \alpha)^N, \quad (4.61)$$

where the summation over $T_j$ contains $\binom{N}{j}$ terms.

### 4.4.4 Upbound of Probability of Detection Error

Since $|M_l| + |M_0| = N$, from (4.55) and (4.57), we obtain

$$\frac{\mu_l}{N} \geq \frac{|M_l|}{N} \min_{1 \leq i \leq N} \left\{ \sum_{w=0}^{n} \ln(1 + a_i b_i^w) f_1(w) \right\} + \frac{|M_0|}{N} \min_{1 \leq i \leq N} \left\{ \sum_{w=0}^{n} \ln(1 + a_i b_i^w) f_0(w) \right\} \quad (4.62)$$

and

$$\frac{\sigma_l^2}{N} \leq \frac{|M_l|}{N} \max_{1 \leq i \leq N} \left\{ \sum_{w=0}^{n} \ln^2(1 + a_i b_i^w) f_1(w) - \left[ \sum_{w=0}^{n} \ln(1 + a_i b_i^w) f_1(w) \right]^2 \right\}$$

$$+ \frac{|M_0|}{N} \max_{1 \leq i \leq N} \left\{ \sum_{w=0}^{n} \ln^2(1 + a_i b_i^w) f_0(w) - \left[ \sum_{w=0}^{n} \ln(1 + a_i b_i^w) f_0(w) \right]^2 \right\}. \quad (4.63)$$
If we define
\[
E_1 := \min_{1 \leq i \leq N, q \in \{0, 1\}} \left\{ \sum_{w=0}^{n} \ln(1 + a_i b_i^w f_q(w)) \right\}
\]
(4.65)
\[
v_1 := \max_{1 \leq i \leq N, q \in \{0, 1\}} \left\{ \sum_{w=0}^{n} \ln^2(1 + a_i b_i^w f_q(w)) - \left[ \sum_{w=0}^{n} \ln(1 + a_i b_i^w f_q(w)) \right]^2 \right\},
\]
(4.66)
then we have
\[
\frac{\mu}{N} > E_1
\]
(4.67)
\[
\frac{\sigma^2}{N} < v_1.
\]
(4.68)

Then from Chenoff bound of Q-function, it can be shown that:
\[
Q\left( \frac{\mu - \ln 2}{\sigma} \right) = Q\left\{ \frac{\mu}{N} - \frac{\ln 2}{\sigma} \right\}
\leq \exp\left\{ -\left( \frac{\mu}{N} - \frac{\ln 2}{N} \right)^2 \right\}
\]
(4.69)
\[
\leq \exp\left\{ -\frac{\left( \frac{\mu}{N} - \frac{\ln 2}{N} \right)^2}{2\frac{\sigma^2}{N^2}} \right\}
\leq \exp\left\{ -\frac{N(E_1 - \ln 2)^2}{2v_1} \right\}.
\]
(4.70)

Inserting (4.71) into (4.61), we have
\[
P_{E,P} < \sum_{l=1}^{2^{N-1}} \exp\left\{ -N\frac{(E_1 - \ln 2)^2}{2v_1} \right\} \alpha^{|M_l|(1 - \alpha)^N - |M_l|} + 1 \cdot (1 - \alpha)^N
\]
(4.72)
\[
< \exp\left\{ -N\frac{(E_1 - \ln 2)^2}{2v_1} \right\} \sum_{l=1}^{2^{N-1}} \alpha^{|M_l|(1 - \alpha)^N - |M_l|} + (1 - \alpha)^N
\]
(4.73)
\[
< \exp\left\{ -N\frac{(E_1 - \ln 2)^2}{2v_1} \right\} + (1 - \alpha)^N.
\]
(4.74)

Because \((E_1 - \ln 2)^2/2v_1\) and \(1 - \alpha\) are constants, it can be infer that the average probability of detection error, \(P_{E,P}\), decreases exponentially with the number of source nodes, \(N\), regardless of \(e_i\). Therefore, the multiplicity of source nodes can be leveraged to mitigate the effect of overhearing errors in the detection of false data injection.
4.5 Selective MAP Test

In this section, we consider performing the MAP test on a subset of the received packets which the destination can overhear well and another separate MAP test on the remaining packets. This is motivated by the fact that the performance of MAP test heavily depends on overhear packets BERs (as shown in Fig. 4.8), when the number of group packets, $N$, is not large. As a result, its accuracy drops significantly if the packets group involves some of the source packets which the destination can hardly overhear, and this issue become severe especially when those bad overhear packets’ BERs are very high compared with good ones. For example, in Fig. 4.5, it can be seen that even only one bad packet can almost ruin all the detection accuracy advantages gained by a large number of well overhear packets in the same checking group. Hence, to prevent the high detection accuracy brought by the well overheard packets from being compromised by bad overheard packets, those packets that the destination can overhear well are selected out to form one MAP test group, while another MAP test group will only include the remaining high BERs packets. This approach will be referred to as selective MAP test and its block diagram is illustrated in Fig. 4.4.

Figure 4.4 Selective MAP test.
Let $e_{(1)} \leq e_{(2)} \leq \cdots \leq e_{(N)}$ denote the ordered BERs of the overhearing channels and let $Y_{r,(1)}, Y_{r,(2)}, \cdots, Y_{r,(N)}$ denote the corresponding received packets. Then, a subset of received packets, $Y'_r := (Y_{r,(1)}, Y_{r,(2)}, \cdots, Y_{r,(N')})$, where $N' \leq N$, are selected to form the well overheard MAP test group, while the remaining packets $Y''_r := (Y_{r,(N'+1)}, \cdots, Y_{r,(N)})$ are applied to another one. Then, the detection outcomes, $\hat{H}_{MAP'}$ on $Y'_r$ and $\hat{H}_{MAP''}$ on $Y''_r$, are combined using the OR rule: if both $Y'_r$ or $Y''_r$ are classified as $H_0$ (unmodified), then $Y_r$ is classified as $H_0$; otherwise, $Y_r$ will be classified as $H_1$. The parameter $N'$ can be chosen to minimize the average probability of detection error on $Y_r$. The MAP test in Section 4.3.2 corresponds to the special case of $N' = N$.

![Figure 4.5](image_url)  

**Figure 4.5** Average probability of detection error versus the number of source nodes, $N$, for different combination of BERs; BCH(63,24), $W(F_i) = d_{\text{min}}$. 


4.5.1 Probability of False Alarm

For selective MAP test, it can be easily known from OR rule that under the condition \( F = 0 \), the false alarm event for \( Y_r \) happens if either of two separate MAP test has false alarmed. Since the separate MAP test in selective method are exact same as general MAP test, but only with different checking packets group. Then following the same steps as \( P_F \) in (4.48), the probability of false alarm for \( Y_r' \), \( P_{F'} \), is given by

\[
P_{F'} = P(Y(Z(Y_r')) > \ln 2 | F = 0) \simeq Q \left( \frac{\ln 2 - \mu'_0}{\sigma'_0} \right),
\]

where

\[
\mu'_0 = \sum_{i=1}^{N'} \left( \sum_{w=0}^{n} \ln(1 + a(i) b_{(i)}^w) \right) f_{0,i}(w)
\]

\[
(\sigma'_0)^2 = \sum_{i=1}^{N'} \left( \sum_{w=0}^{n} \ln(1 + a(i) b_{(i)}^w)^2 f_{0,i}(w) \right) - \sum_{i=1}^{N'} \left( \sum_{w=0}^{n} \ln(1 + a(i) b_{(i)}^w) \right) f_{0,i}(w) \right]^2.
\]

and the probability of false alarm for \( Y_r'' \), \( P_{F''} \), is given by

\[
P_{F''} = P(Y(Z(Y_r''))) > \ln 2 | F = 0) \simeq Q \left( \frac{\ln 2 - \mu''_0}{\sigma''_0} \right),
\]

where

\[
\mu''_0 = \sum_{i=N'+1}^{N} \left( \sum_{w=0}^{n} \ln(1 + a(i) b_{(i)}^w) \right) f_{0,i}(w)
\]

\[
(\sigma''_0)^2 = \sum_{i=N'+1}^{N} \left( \sum_{w=0}^{n} \ln(1 + a(i) b_{(i)}^w)^2 f_{0,i}(w) \right) - \sum_{i=N'+1}^{N} \left( \sum_{w=0}^{n} \ln(1 + a(i) b_{(i)}^w) \right) f_{0,i}(w) \right]^2.
\]

Hence, the probability of detection error caused by false alarm, \( P'_{E,F,A} \), is given by

\[
P'_{E,F,A} = (1 - (1 - P_{F'})(1 - P_{F''})) P(M_0).
\]
4.5.2 Probability of Missed Detection

It can be inferred from OR rule that under the condition $F \neq 0$, if both $Y_r'$ or $Y_r''$ contain modification packets, the missed detection for $Y_r$ happens only when both MAP test have miss detected. If $Y_r'$ contains modification but $Y_r''$ is not, the missed detection event for $Y_r$ happens only when MAP test on $Y_r'$ miss detected and MAP test on $Y_r''$ doesn’t false alarmed. Similar for the case that $Y_r''$ contains modification but $Y_r'$ is not.

To keep consistency, let $M_{l'}$ denote the $l'$th non-empty subset of $\{(1), (2), \cdots, (N')\}$ that records the index of modified packets, where $l' = 1, 2, \cdots, 2^{N'} - 1$, and $M_{l'}^c$ denote the complement of $M_{l'}$. Following the same steps as $P_{M,l'}$ in (4.59), the probability of missed detection for $Y_r'$ given that the packets in $M_{l'}$ are modified, $P_{M,l'}$, is given by

$$P_{M,l'} = P(Y(Z(Y_r')) < \ln 2 | M_{l'}) \simeq Q \left( \frac{\mu_{l'} - \ln 2}{\sigma_{l'}} \right),$$

(4.82)

where

$$\mu_{l'} = \sum_{i \in M_{l'}} \left[ \sum_{w=0}^{n} \ln(1 + a_ib^w_i) f_{1,i}(w) \right] + \sum_{i \in M_{l'}^c} \left[ \sum_{w=0}^{n} \ln(1 + a_ib^w_i) f_{0,i}(w) \right]$$

$$= \sum_{i \in M_{l'}} \left[ \sum_{w=0}^{n} \frac{\ln(1 + a_ib^w_i)^2}{2} f_{1,i}(w) \right] - \sum_{i \in M_{l'}} \left[ \sum_{w=0}^{n} \frac{\ln(1 + a_ib^w_i)^2}{2} f_{0,i}(w) \right]$$

$$+ \sum_{i \in M_{l'}^c} \left[ \sum_{w=0}^{n} \frac{\ln(1 + a_ib^w_i)^2}{2} f_{0,i}(w) \right] - \sum_{i \in M_{l'}^c} \left[ \sum_{w=0}^{n} \frac{\ln(1 + a_ib^w_i)^2}{2} f_{0,i}(w) \right].$$

(4.83)

Similarly, if we let $M_{l''}$ denote the $l''$th non-empty subset of $\{(N'+1), (N'+2), \cdots, (N)\}$ that records the index of modified packets, where $l'' = 1, 2, \cdots, 2^{N-N'} - 1$, and $M_{l''}^c$ denote the complement of $M_{l''}$. The probability of missed detection for $Y_r''$ given that the packets in $M_{l''}$ are modified, $P_{M,l''}$, is given by

$$P_{M,l''} = P(Y(Z(Y_r'')) < \ln 2 | M_{l''}) \simeq Q \left( \frac{\mu_{l''} - \ln 2}{\sigma_{l''}} \right),$$

(4.85)

where

$$\mu_{l''} = \sum_{i \in M_{l''}} \left[ \sum_{w=0}^{n} \ln(1 + a_ib^w_i) f_{1,i}(w) \right] + \sum_{i \in M_{l''}^c} \left[ \sum_{w=0}^{n} \ln(1 + a_ib^w_i) f_{0,i}(w) \right]$$

(4.86)
\[(\sigma''_l)^2 = \sum_{i \in M_{l''}} \left[ \sum_{w=0}^{n} \ln(1 + a_i b^w_i) f_{1,i}(w) \right]^2 - \sum_{i \in M_{l'}} \left[ \sum_{w=0}^{n} \ln(1 + a_i b^w_i) f_{1,i}(w) \right]^2 + \sum_{i \in M_{l''}} \left[ \sum_{w=0}^{n} \ln(1 + a_i b^w_i) f_{0,i}(w) \right] \]  

4.5.3 Average Probability of Detection Error

By inserting (4.75), (4.78) into (4.81) and (4.82), (4.85) into (4.88), the probability of detection error for selective MAP test, \(P'_{E,P} = P'_{E,FA} + P'_{E,MD}\) can be approximated by

\[
P'_{E,P} \simeq \left[ 1 - (1 - Q\left( \frac{\ln 2 - \mu_0'}{\sigma_0'} \right) ) (1 - \alpha)^N \right] \sum_{l'=1}^{2^{N'-1}} P_{M,l'} P(M_{l'}) \]  

Hence the the average probability of detection error for selective MAP test caused by missed detection, \(P'_{E,MD}\), is given by

\[
P'_{E,MD} = \sum_{l=1}^{2^{N'-1}} P_{M,l} P(M_{l}) \sum_{l'=1}^{2^{N'-1}} P_{M,l'} P(M_{l'}) \]  

\[+ (1 - P_{E'}) P(M_0') \sum_{l=1}^{2^{N'-1}} P_{M,l} P(M_{l}) \]  

\[+ (1 - P_{E'}) P(M_0) \sum_{l'=1}^{2^{N'-1}} P_{M,l'} P(M_{l'}).\]  

(4.88)
The optimal choice of $N'$

Even though $N'$ can be any value in the range $[0, N]$, the optimal value of $N'$, denoted by $N_{opt}'$, should be chosen such that $P'_{E,P}$ is minimized. This $N_{opt}'$ can be done by numerical search: changing $N'$ from 0 to $N$ and calculates each $P'_{E,P}$ to find the minimum, i.e.

$$\{N_{opt}' | P'_{E,P}(N_{opt}') = \min_{0 \leq N' \leq N} P'_{E,P}(N')\}.$$  (4.90)

4.6 Numerical Results

![Figure 4.6](image.png)

Figure 4.6  Probability of detection error versus the number of source nodes for different values of $\alpha$, $N$; BCH(31,21), $e_i = 0.1$ for $i = 1, 2, \cdots, N/2$ and $e_i = 0.05$ for $i = N/2 + 1, \cdots, N$, $W(F_i) = d_{min}$.

Fig. 4.6 illustrates the probability of detection error versus the number of packets $N$ for different values of $\alpha$. The circle and solid line are obtained by simulation when the detection rule in (4.22) and (4.40) are applied, respectively. The dotted line is obtained by the Gaussian approximation.
in (4.61). One can observe that the Gaussian approximation is accurate if $N$ is larger than 8. It can also be observed that the detection rule in (4.40) is close to the original MAP detection rule in (4.22). Hence, the simpler detection rule in (4.40) can be used for all $N$ without losing the accuracy of detection. In the remaining part of this paper, we will assume the use of the detection rule is (4.40). It can be seen that the probability of detection error decreases exponentially with $N$, which can also be observed from (4.74). One can also observe that the probability of detection error decreases as $\alpha$ increases, which means the MAP test provides a more accurate detection if more packets are modified by the adversary. This is because the detection statistic $Y(Z)$ in (4.40) increases as $\alpha$ increases, which weighs more towards $H_1$. This enforces the adversary to limit the number of modified packets in order to minimize the probability that the false data injection is detected, which also constrains the effectiveness of attack.

![Figure 4.7 Probability of detection error versus the number of packets $N$](image)

Fig. 4.7 shows the probability of detection error versus the number of packets $N$ for different values of $\alpha$ with its upper bound, $N$; $\text{BCH}(31,21)$, $e_i = 0.1$ for $i = 1, 2, \cdots, N/2$ and $e_i = 0.05$ for $i = N/2 + 1, \cdots, N$, $W(F_i) = d'_{\text{min}}$. 
values of $\alpha$ with its upper bound. It can be observed that, for large $N$, the average probability of detection error, $P_{E,P}$, decreases exponentially with the number of source nodes, $N$, regardless of $e_i$.

Figure 4.8 Probability of detection error versus the overhearing bit error rate for different value of $N$; BCH(63,24), $B = 10$, $\alpha = 0.2$.

Fig. 4.8 illustrates the probability of detection error versus the overhearing bit error rate (BER) for different values of $N$. It can be seen the probability of detection error, $P_{E,P}$, decreases with decreasing BER due to the less noise in overhearing channel. This indicates that the effect of overhearing error can be mitigated by abundance of source nodes: even if the overhearing BER is high, the physical layer MAP test can still provide a lower probability of detection error. This property is particularly desirable for emerging high-density network networks where a real-time decision making is performed based on a large number of short packets that are collected from various sensors.

Fig. 4.9 illustrates the average probability of detection error, $P_{E,P}$, versus the percentage
of modified bits in a packets, $W(F_i)/n$, for different values of $N$. It can be observed that the average probability of detection error monotonically decreases as $W(F_i)$ increases, i.e. more bits are modified. This enforces the adversary to change the minimum number of bits, which is $d_{\text{min}}$, in order to minimize the chance of being detected. In the remaining part, we will assume that $W(F_i) = d_{\text{min}}$. For small $W(F_i)/n$, $P_{E,P}$ is dominated by the false alarm probability, $P_{FP}(M_0)$, while for large $W(F_i)/n$, it is dominated by the missed detection probability $\sum_{l} P_{M,l} P(M_l)$. Both terms decrease with increasing $W(F_i)$, but their slopes are different: the former term decaying slope changes from flat to steep with $W(F_i)$ increases and soon vanishes, but the latter decaying speed keeps an median slope of it and so becomes dominate part in $P_{E,P}$.

Fig. 4.10 illustrates the average probability of detection error versus the number of source
Figure 4.10  Probability of detection error versus the number of source nodes $N$ for different values of $(n, k)$: $\alpha = 0.2$ $e_i = 0.1$ for $i = 1, 2, \cdots, N/2$ and $e_i = 0.05$ for $i = N/2 + 1, \cdots, N$.

nodes, $N$, for different values of $(n, k)$. It is assumed that $W(F_i) = d_{\text{min}}$. It can be seen that the probability of detection error decreases faster with $N$ when the code rate $k/n$ is smaller. This is because the minimum distance $d_{\text{min}}$ increases with decreasing code rate, which enforces the adversary to change more bits i.e. $W(F_i)$ to be increased. The increase of $W(F_i)$ makes the detection more accurate as indicated in Fig. 4.9.

Fig. 4.11 illustrates the average probability of detection error versus the number of source nodes, $N$, for different MAP test schemes. It can be seen that the selective MAP test (dotted line) provides improvement over general MAP test (solid line) when $N$ is in small range ($N < 30$). This is because in selective MAP test, the well overheard packets and the bad overheard packets are excluded from each other, the probability of false alarm and missed detection of low BERs
Figure 4.11 Average probability of detection error versus the number of source nodes, \( N \); BCH(63,24), \( e_i = 0.2 \) for \( i = 1, 2, \ldots, N/3 \) and \( e_i = 0.05 \) for \( i = N/3 + 1, \ldots, N \), \( W(F_i) = d_{\text{min}} \).

\( (e_i = 0.05) \) level and high BERs \( (e_i = 0.2) \) level are captured in \( \{P_{F'}, P_{M,l'}\} \) and \( \{P_{F''}, P_{M,l''}\} \), respectively. However in general MAP test, as have shown in Fig. 4.5, the probability of false alarm and missed detection, captured in \( \{P_F, P_{M,l}\} \), is dominated by the bad overheard packets in the checking group, i.e. \( \{P_F, P_{M,l}\} \approx \{P_{F''}, P_{M,l''}\} \), which makes its probability of detection error very close to the \( P_{E,F} \), where \( e_i = 0.2 \) for \( i = 1, 2, \ldots, N \).

Since \( \sum_{l'} P_{M,l'} P(M_{l'}) \ll \sum_{l''} P_{M,l''} P(M_{l''}) < 1 \) and \( P_{F'} \ll P_{F''} < 1 \) due to the high detection accuracy of low BERs packets, it can be known based on (4.81) and (4.88) that \( P_{E,F,A}' \approx P_{F''} P(M_0) \) and \( P_{E,M,D}' \approx (1 - P_{F'}) P(M_{l'}) \sum_{l''} P_{M,l''} P(M_{l''}) \). That is to say compared with general MAP test’s false alarm errors, \( P_F P(M_0) \) and missed detection errors, \( \sum_l P_{M,l} P(M_l) \); selective MAP test’s false alarm errors, \( P_{F''} P(M_0) \), just increase a little bit, because \( P_F \) is almost as high as \( P_{F''} \). On the
other hand, its missed detection errors, \((1 - P_{F'}) P(M_{0'}) \sum_{l'} P_{M,l'} P(M_{l'})\) is reduced by a small factor, \((1 - P_{F'})(1 - \alpha)^{N'}\). The reason why both lines merge gradually is that the probability of miss detection detection and false alarm both affect the probability of detection error, but the reduce on missed detection errors is large than the increase on false alarm errors when \(N\) is in small range; However, when \(N\) keep increasing, the errors caused by false alarm gradually dominates the whole probability of detection error, the advantages caused by less missed detection rate gradually become insignificant.

4.7 Conclusion

In this chapter, we studied the MAP test for detecting false data in short packets in wireless relay networks. The MAP test exploits noisy information that the destination receives directly from the source nodes as a reference. We found that the average probability of detection error decreases exponentially as the number of source nodes increases, regardless of the reliability of the reference information. The detection is more accurate if more bits in a packet or more packets are modified. This enforces the adversary to limit the number of modified bits in a packet or the number of modified packets, and thereby constraining the effectiveness of attack. We also development selective MAP test which have improvement when the number of packets is small. The proposed scheme would be useful in large-scale networked sensing and decision making systems, such as the IoT, where a decision is made based on a large number of short packets that are collected from multiple source nodes.

4.8 Appendix: The Exact Expression of Probability of False Alarm and Missed Detection for (4.40)

In this appendix, the exact expression of probability of false alarm and missed detection for (4.40) are provided.
4.8.1 Probability of False Alarm

The event of false alarm occurs if \( Y(Z) \geq \ln 2 \), under the situation \( F = 0 \). Since for \( i = 1, 2, \cdots, N \), \( W(Z_i|F = 0) = W(E_{si}) \) and \( W(E_{si}) \) is Binomially distributed with parameters \( n \) and \( e_i \), i.e. \( W(E_{si}) \sim \text{Binom}(n, e_i) \). Hence the probability of false alarm,

\[
P_F = P(Y(Z) \geq \ln 2|F = 0))
\]

\[
P = P\left(\sum_{i=1}^{N} \ln \left[1 + a_i b_i W(E_{si})^i\right] \geq \ln 2|F = 0)\right)
\]

\[
= P\left(\sum_{i=1}^{N} \ln \left[1 + a_i b_i W(E_{si})^i\right] \geq \ln 2\right).
\]

Over the years, the methods to calculate CDF (cumulative distribution function) of a random variable has been studied in many literature. In paper Duby et al. (2013), a concise equation, which based on Gil-Pelaez’s inversion formula Gil-Pelaez (1951), to calculate the CDF of random variable \( V \) is given by

\[
F_V(V \leq v) = \frac{1}{2} - \frac{1}{\pi} \int_{0}^{\infty} \frac{\text{Im}\{e^{-jtv} \varphi_V(t)\}}{t} dt,
\]

where \( \text{Im}\{\cdot\} \) is imaginary part, \( e \) is natural base and \( \varphi_V(t) \) is characteristic function of \( V \), defined as \( \varphi_V(t) = E[e^{jtV}] \). Following (4.93) and (4.94), if we let the \( V \) in (4.94) be

\[
V = \sum_{i=1}^{N} \ln \left[1 + a_i b_i W(E_{si})^i\right],
\]

obviously,

\[
P_F = P(V \geq \ln 2)
\]

\[
= 1 - P(V < \ln 2)
\]

\[
= \frac{1}{2} + \frac{1}{\pi} \int_{0}^{\infty} \frac{\text{Im}\{e^{-j\ln 2 \varphi_V(t)\}}}{t} dt.
\]

Since \( V = \sum_{i=1}^{N} \ln[1 + a_i b_i W(E_{si})^i] \) is the linear combination Duby et al. (2013) of independent random variables \( \ln[1 + a_i b_i W(E_{si})^i] \), if we let \( V_i := \ln[1 + a_i b_i W(E_{si})^i] \), then the characteristic function of \( V \) is the production of \( \varphi_{V_i}(t) \), i.e.

\[
\varphi_V(t) = \prod_{i=1}^{N} \varphi_{V_i}(t).
\]
Because $W(E_{si})$ is Binomial distributed, the characteristic function of $V_i$,

$$
\varphi_{V_i}(t) = E\left[e^{jtV_i}\right] = \sum_{k=w}^{n} \binom{n}{w} e^{w(1-e^{i})^{n-w}e^{jt\ln(1+a_{i}b^{w})}}.
$$

(4.101)

Substituting $\varphi_{V}(t)$ in (4.98) by (4.99) with (4.101), we obtain

$$
P_F = \frac{1}{2} + \frac{1}{\pi} \left( \int_{0}^{\infty} \frac{\text{Im}\left\{ e^{-jt\ln 2} \prod_{i=1}^{N} \sum_{k=0}^{n} \binom{n}{k} e^{k(1-e^{i})^{n-k}e^{jt\ln(1+a_{i}b^{k})}} \right\}}{t} dt. \right)
$$

(4.102)

### 4.8.2 Probability of Missed Detection

On the contrary, the event of missed detection occurs if $Y(Z) < \ln 2$ under the situation $F \neq 0$. For $i = 1, 2, \cdots, N$, we employ $F_{i1}$ to denote the all-one subvector of $F_i$ and $F_{i0}$ denotes the remaining all-zero subvector of $F_i$. For example, if $F_i = 1011001$ then $F_{i1} = 1111$ and $F_{i0} = 000$. Correspondingly, let $E_{i1}$ denotes the subvector of $E_{si}$ which is composed of the entries in the positions where $F_i$ is 1 and $E_{i0}$ denotes the remaining subvector of $E_{si}$ which is composed of the entries in the positions where $F_i$ is 0. For example, if $F_i = 1011001$ and $E_{si} = 0101011$ then $E_{i1} = 0011$ and $E_{i0} = 101$.

Since single bit of $E_{si}$ is 1 with probability $e_i$ and they are independent, let $d_i = W(F_i)$, then we have $W(F_{i1} \oplus E_{i1}) \sim \text{Binom}(d_i, 1-e_i)$ and $W(F_{i0} \oplus E_{i0}) \sim \text{Binom}(n-d_i, e_i)$, therefore, the probability of missed detection given that the packets in the subset $M_l$ are modified is given by

$$
P_{M,l} = P(Y(Z) < \ln 2|M_l)
$$

(4.103)

$$
P_{M,l} = P\left( \sum_{i=1}^{N} \ln \left[ 1 + a_{i}b^{i}W(F_{i}) \right] < \ln 2|M_l \right)
$$

(4.104)

$$
P_{M,l} = P\left( \sum_{i=1}^{N} \ln \left[ 1 + a_{i}b^{i}W(F_{i0} \oplus E_{i0}) + W(F_{i1} \oplus E_{i1}) \right] < \ln 2 \right).
$$

(4.105)

Similarly, if we let the $V$ in (4.94) be

$$
V = \sum_{i=1}^{N} \ln \left[ 1 + a_{i}b^{i}W(F_{i0} \oplus E_{i0}) + W(F_{i1} \oplus E_{i1}) \right],
$$

(4.106)
following (4.94) and (4.105), we obtain

\[ P_{M,l} = P(V < \ln 2) \]

\[ = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\text{Im}\{e^{-jt\ln 2} \varphi_V(t)\}}{t} dt. \]  

(4.107)

(4.108)

Further, from (4.51), we know that the probability mass function of \( W(F_{i0} \oplus E_{i0}) + W(F_{i1} \oplus E_{i1}) \) if \( F_i \neq 0 \) is given by

\[ f_1(w) = \sum_{m=0}^w \binom{W(F_i)}{m} \left( \frac{n - W(F_i)}{w - m} \right) e_i^{W(F_i) + w - 2m} (1 - e_i)^{n - (W(F_i) + w - 2m)} \]

(4.109)

if we let \( V_i := \ln[1 + a_i b_i^{W(F_{i0} \oplus E_{i0}) + W(F_{i1} \oplus E_{i1})}] \) the characteristic function of \( V_i \) are given by

\[ \varphi_{V_i}(t) = E[e^{jtV_i}] \]

\[ = \sum_{w=0}^n f_1(w) e^{jt \ln(1 + a_i b_i^w)} \]

(4.110)

(4.111)

\[ = \sum_{w=0}^n \sum_{m=0}^w \binom{W(F_i)}{m} \left( \frac{n - W(F_i)}{w - m} \right) e_i^{W(F_i) + w - 2m} (1 - e_i)^{n - (W(F_i) + w - 2m)} e^{jt \ln(1 + a_i b_i^w)} \]

(4.112)

Likewise, the characteristic function of \( V \) is the production of \( \varphi_{V_i}(t) \), i.e.

\[ \varphi_V(t) = \prod_{i=1}^N \varphi_{V_i}(t) \]

\[ = \prod_{i \in M_i} \sum_{w=0}^n f_1(w) e^{jt \ln(1 + a_i b_i^w)} \prod_{i \in M_i} \sum_{k=w}^n \binom{n}{w} (1 - e_i)^{n-w} e^{jt \ln(1 + a_i b_i^w)} \]

(4.113)

(4.114)

Then replacing \( \varphi_V(t) \) in (4.108) with (4.114), we got

\[ P_M = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{t} \text{Im}\left\{ e^{-jt\ln 2} \prod_{i \in M_i} \sum_{w=0}^n f_1(w) e^{jt \ln(1 + a_i b_i^w)} \right\} e_i^{W(F_i) + w - 2m} (1 - e_i)^{n - (W(F_i) + w - 2m)} e^{jt \ln(1 + a_i b_i^w)} dt, \]

(4.115)

where \( f_1(w) \) is given by (4.109). The (4.102) and (4.115) can be calculated through trapezoid Strawderman (2004) method.
4.9 References


CHAPTER 5. PACKET RECOVERY OF PHYSICAL LAYER MAP TEST
IN WIRELESS RELAY NETWORKS

5.1 Introduction

Once the MAP detection decides \( H_1 \) (packets group contains modified ones), discarding the entire group (total of \( N \) packets) is the safest way to protect destination from being misguided by adversary’s falsified information. Even though the destination may keep sending Automatic Repeat Request (ARQ) Safak (2017) until the crypto-aided MAP detection result shows \( \hat{H}_{CP} = H_0 \) (packets group contains no modified packets), in some scenarios, this mechanism may cause a significant drop on the throughput. For instance, if adversary knows the detection method, he may merely modifies a few specific packets obviously, which would result in the other unmodified packets also wasted at the destination end. This type of attack may become very critical especially when size of group packets, \( N \), is a very large number. Therefore, an ‘Recovery’ method that could prevent the most of unmodified ones from being discarded is very necessary. In this chapter we discuss a packet recovery mechanism after the MAP test (integrity check) is completed. The packet recovery mechanism attempts to repair false alarmed packets, i.e. unmodified packets that are mistakenly declared as modified (corrupted). The proposed recovery mechanism relies on the likelihood ratio test (LRT) that conveys the confidence that a particular packet is modified or not. It allows the destination to determine, with no additional feedback or information from the source nodes, which packets are likely to be correct and to recover the correct packets. The goal is to avoid discarding unmodified packets, thereby improving the throughput.

5.2 Likelihood Ratio Test

The proposed packet recovery mechanism is illustrated in Fig. 5.1. Firstly, the MAP test is performed on \( N \) received packets, \( \mathbf{Y}_r = (Y_{r,1}, \cdots , Y_{r,N}) \). If the test outcome is negative (\( H_0 \)), then
the destination accepts $Y_r$. If the test outcome is positive ($H_1$), then the LRT is performed on each packet to identify the modified packets. Those packets that are identified as modified are discarded, while the remaining ones are accepted.

![Diagram](image.png)

**Figure 5.1** Packet recovery flow.

The likelihood ratio of $F_i$ given $Z_i$ is given by

$$
\lambda_i := \frac{P(F_i \neq 0 | Z_i)}{P(F_i = 0 | Z_i)}
$$

(5.1)

$$
= \frac{P(F_i \neq 0, Z_i)}{P(F_i = 0, Z_i)}
$$

(5.2)

$$
= \frac{\sum_{j=1}^{2^k-1} P(Z_i | F_i = C_j) P(F_i = C_j)}{P(Z_i | F_i = 0) P(F_i = 0)}.
$$

(5.3)

From Section 4.3.2, we know

$$
\sum_{j=1}^{2^k-1} P(Z_i | F_i = C_j) P(F_i = C_j) = \sum_{j=1}^{2^k-1} e_i^{W(Z_i \oplus C_j)} (1 - e_i)^{n-W(Z_i \oplus C_j)} P(F_i = C_j),
$$

(5.4)

and

$$
P(Z_i | F_i = 0) P(F_i = 0) = e_i^{W(Z_i)} (1 - e_i)^{n-W(Z_i)} (1 - \alpha).
$$

(5.5)

Therefore, $\lambda_i$ in (5.3) equals to

$$
\lambda_i = \frac{1}{1 - \alpha} \sum_{j=1}^{2^k-1} \left( \frac{e_i^{W(Z_i \oplus C_j) - W(Z_i)}}{1 - e_i} \right) P(F_i = C_j).
$$

(5.6)
It can be easily shown that \( \ln(\lambda_i + 1) \) is exactly the summation term of (4.22). Then follow the analysis in Section 4.3.3, the \( \ln(\lambda_i + 1) \) can be simplified as

\[
\ln(\lambda_i + 1) = \ln \left[ 1 + \frac{\alpha}{1 - \alpha} \left( \frac{e_i}{1 - e_i} \right)^{d_{\min}} \left[ \left( \frac{1 - e_i}{e_i} \right)^{2} - 1 \right] \frac{d_{\min}}{n} + 1 \right]^{W(Z_i)},
\]

(5.7)

so \( \lambda_i \) can be expressed as

\[
\lambda_i = a_i b_i^{W(Z_i)},
\]

(5.8)

where

\[
a_i = \frac{\alpha}{1 - \alpha} \left( \frac{e_i}{1 - e_i} \right)^{d_{\min}}
\]

(5.9)

\[
b_i = \left[ \left( \frac{1 - e_i}{e_i} \right)^{2} - 1 \right] \frac{d_{\min}}{n} + 1.
\]

(5.10)

Then, we decide \( Y_{r,i} \) is unmodified (\( H_0 \)) if \( \lambda_i < \gamma_i \) for some threshold \( \gamma_i \) and, otherwise, decide \( Y_{r,i} \) is modified (\( H_1 \)). It can be shown that the LRT is equivalent to

\[ W(Z_i) \overset{H_1}{\gtrless} \tau_i \]

(5.11)

where \( \tau_i = \log_{10}(\gamma_i/a_i) \). Also from 4.4.1, we know that the probability mass function of \( W(Z_i)|H_0 \) is \( f_0(w) \) given by (4.42). Because false alarm event on \( Y_{r,i} \) occurs if \( W(Z_i) \geq \tau_i \) given \( H_0 \) (\( F_i = 0 \)), the probability of false alarm on \( Y_{r,i} \) is given by

\[ Q_{F,i} = \sum_{w=\tau_i}^{n} f_0(w) \]

(5.12)

\[ = \sum_{w=\tau_i}^{n} \left( \frac{n}{w} \right) e_i^w (1 - e_i)^{n-w}. \]

(5.13)

Similarly, it can be known from 4.4.2 that the probability mass function of \( W(Z_i)|H_1 \) is \( f_1(w) \) given by (4.51). Since the miss detection event on \( Y_{r,i} \) occurs if \( W(Z_i) < \tau_i \) given \( H_1 \) (\( F_i \neq 0 \)), if we assume \( W(F_i) = d_{\min} \), which maximizes the likelihood of bypassing the MAP test, then the probability of missed detection on \( Y_{r,i} \) is given by

\[ Q_{M,i} = \sum_{w=0}^{\tau_i-1} f_1(w) \]

(5.14)

\[ = \sum_{w=0}^{\tau_i-1} \left[ \min(w, d_{\min}) \left( \frac{d_{\min}}{m} \right) \left( n - d_{\min} \right) \right] \times e_i^{w+d_{\min}-2m} (1 - e_i)^{n-(w+d_{\min})+2m}. \]

(5.15)
Remark

As we can see, applying likelihood ratio test of recovery scheme won't cause extra computational cost, because $\lambda_i$ has already been calculated in MAP detection and it can be directly used from the memory.

5.3 Normalized Throughput Analysis for Different Schemes

The normalized throughput is defined as the average number of unmodified packets that are accepted by the destination per packet transmission.

**MAP+LRT:** The normalized throughput for the MAP test followed by the LRT is given by

$$\eta_{MAP+LRT} = \frac{1}{N} \left[ \left( (1 - P_F)N + P_F \sum_{i=1}^{N} (1 - Q_{F,i}) \right)(1 - \alpha)^N \right.$$\left.
+ \sum_{l=1}^{2^N-1} (1 - P_{M,l}) \left( \sum_{i=1}^{N-|M_l|} (1 - Q_{F,(i)}) \right) P(M_l) \right], \quad (5.16)
$$
where $P_F$ and $P_{M,l}$ are given by (4.48) and (4.59), respectively, and $Q_{F,(i)}$ is the probability of false alarm on the $i$th packet in the unmodified set $M^c_l$. Here, the first term is the average number of unmodified packets that are accepted by the destination when all packets in $Y_r$ are unmodified, and the second term is that when at least one packet in $Y_r$ is modified.

The probability that a modified packet in $Y_r$ is mistakenly accepted by the destination is given by

$$\rho_{MAP+LRT} = \frac{1}{N} \sum_{l=1}^{2^N-1} \left[ |M_l|P_{M,l} + \left( \sum_{i=1}^{|M_l|} Q_{M,(i)} \right)(1 - P_{M,l}) \right] P(M_l), \quad (5.17)
$$
where $Q_{M,(i)}$ is the probability of missed detection on the $i$th packet in the modified set $M_l$. Here, the first term is the average number of missed detection on the $i$th packet in the modified set $M_l$. Here, the first term is the average number of modified packets that are undetected by the MAP test and the second term is the average number of modified packets that are detected by the MAP test but undetected by the LRT.
MAP: If no LRT (packet recovery) is performed after the MAP test, then all packets in $Y_r$ that are declared $H_1$ by the MAP test will be considered as modified ($H_1$). That is, $Q_{F,i} = 1$ and $Q_{M,i} = 0$ for all $i$. Therefore, the normalized throughput for the MAP test only (without packet recovery) is given by (5.16) with $Q_{F,i} = 1$, which yields

$$\eta_{MAP} = (1 - P_F)(1 - \alpha)^N,$$  \hspace{1cm} (5.18)

and the probability that a modified packet in $Y_r$ is mistakenly accepted is given by (5.17) with $Q_{M,i} = 0$ for all $i$, which yields

$$\rho_{MAP} = \frac{1}{N} \sum_{l=1}^{2^N-1} |M_l|P_{M,l}P(M_l).$$ \hspace{1cm} (5.19)

LRT: If no MAP test is performed prior to the LRT, then all packets in $Y_r$ will be applied to the LRT. This corresponds to the case that the MAP test always declares $H_1$ in Fig. 5.1, namely $P_F = 1$ and $P_{M,l} = 0$. Therefore, the normalized throughput with LRT only (without the MAP test) is given by (5.16) with $P_F = 1$ and $P_{M,l} = 0$, which yields

$$\eta_{LRT} = \frac{1}{N} \sum_{l=0}^{2^N-1} \left( \sum_{i=1}^{N-|M_l|} (1 - Q_{F,i}) \right) P(M_l) \hspace{1cm} (5.20)$$

and the probability that a modified packet in $Y_r$ is mistakenly accepted is given by (5.17) with $P_{M,l} = 0$, which yields

$$\rho_{LRT} = \frac{1}{N} \sum_{l=1}^{2^N-1} \sum_{i=1}^{|M_l|} Q_{M,(i)}P(M_l) \hspace{1cm} (5.22)$$

$$= \frac{\alpha}{N} \sum_{i=1}^N Q_{M,i}.$$

The proof of (5.21) and (5.23) is provided in Appendix 5.6.
**MAC:** Message authentication code (MAC) algorithm, sometimes called a keyed (cryptographic) hash function Paar et al. (2010b), validates the received packet $Y_{r,i}$ by recomputing the MAC based on $Y_{r,i}$ and the key, $h(key,Y_{r,i})$, and comparing it to the MAC $h_i$ which is received from R: if $h(key,Y_{r,i}) \neq h_i$ then $Y_{r,i}$ is classified as “modified”, otherwise, it is classified as “unmodified”. If the length of $h(key,X_i)$ is $B$ bits, then the probability that the adversary randomly picks a hash sequence and a modified packet is validated by the verifier is $2^{-B}$. However, the true packet is always validated regardless of $B$. Therefore, the probability that a true packet is declared positive (modified) is given by

$$Q_{F,C} = 0$$  \hspace{1cm} (5.24)

and the probability that a false packet is declared as negative (unmodified) is given by

$$Q_{M,C} = 2^{-B}.$$  \hspace{1cm} (5.25)

Therefore, the normalized throughput with MAC is given by by (5.21) with $Q_{F,i} = 0$ and $Q_{M,i} = 2^{-B}$, which yields

$$\eta_{MAC} = \frac{n}{N(n+B)} \left[ \sum_{l=0}^{2^{N-1}} (N - |M_l|)P(M_l) \right]$$  \hspace{1cm} (5.26)

$$= \frac{n(1-\alpha)}{n+B}$$  \hspace{1cm} (5.27)

where the factor $n/(n+B)$ is to account for the overhead $B$ bits for the MAC, and the probability that a modified packet in $Y_r$ is mistakenly accepted is given by (5.23) with $Q_{M,(i)} = 2^{-B}$ for all $i$, which yields

$$\rho_{MAC} = 2^{-B} \alpha.$$  \hspace{1cm} (5.28)

### 5.3.1 Compared with Existing Tracing Bits techniques

**Tracing bits:** In tracing bits technique Mao and Wu (2007); Khalaf et al. (2014), tracing bits are embedded in the information data, i.e. the number of $t$ bits out of $k$ bits are used to are used to detect whether the transmission packets is changed or not by relays. It assume those tracing
bits positions and its pattern are secrets pre-share only between each $S_i$ with $D$, also it assumes relay in between never known the transmission packet contains tracing bits.

The detector validates the received packet $Y_{r,i}$ by extracting its tracing bits out of the information bits after decoding, and comparing it to the "ground truth", which has been pre-shared. If extracted tracing bits is different from the "ground truth", then $Y_{r,i}$ is classified as "modified", otherwise, it is classified as "unmodified". The performance of tracing bits technique highly depends on the number of tracing bits, $t$, because the adversary may happen to choose the $F_i \neq 0$, which makes the decoded codeword, $\text{Decode}(X_i + F_i)$ has the exact tracing bits with $\text{Decode}(X_i)$. If that so, this modified packet will still be validated by the verifier (However, the true packet is always validated regardless of number of tracing bits $t$). To increasing the accuracy, the number of tracing bits has to be large, which on other hand may reduce the throughput. Above all, the probability that a true packet is declared positive (modified) is given by

$$Q_{F,T} = 0.$$  \hfill (5.29)

If adversary perform false data injection, in averaging sense, each bit in the transmission packet (n-bits codeword) has same chance to be flipped. Since a smart attacker only change $d_{\text{min}}$ bits, the missed detection happens if those $d_{\text{min}}$ bits position does not affect tracing bits. Hence the probability that a false packet is declared as negative (unmodified) can be approximated by

$$Q_{M,T} \simeq \prod_{\omega=0}^{d_{\text{min}}-1} \frac{n-t-\omega}{n-\omega},$$  \hfill (5.30)

which is obtained by following the fact that, the first modified bit among those $d_{\text{min}}$ position has probability of $(n-t)/n$ in not affecting tracing bits, the second modified bit among those $d_{\text{min}}$ position has probability of $(n-t-1)/(n-1)$ in not affecting tracing bits, and so forth.

**Simulation method to find $Q_{M,T}$**

Let $N_{\text{trial}}$ denotes the number of trails, for each trail: the sender first generate a $k$-bits length message, $m = (m_1, m_2, \cdots, m_k)$, where $(m_1, m_2, \cdots, m_t) \ (t \leq k)$ are assumed to be tracing bits.
And each $m_j$ has the probability of $1/2$ to be 1 and the probability of $1/2$ to be 0, i.e.

$$
\begin{aligned}
Pr(m_j = 1) &= \frac{1}{2}, \\
Pr(m_j = 0) &= \frac{1}{2},
\end{aligned}
$$

(5.31)

for $j = 1, 2, 3, \cdots, k$. Then $m$ is encoded by $(n, k)$ linear code, denoted as $X$, which is a $n$-bits length codeword and sent to relay. After receive $X$, relay may perform false data injection attack, which is equivalent to say, relay will transmit $X' = X \oplus F$ to destination, with the probability of $1 - \alpha$ that $F = 0$ and the probability of $\alpha$ that $F \neq 0$. When $F \neq 0$, $F$ will be chosen from codebook $\{C_j\}, j = 1, 2, 3, \cdots, 2^k - 1$ such that $W(F) = d_{\min}$, if there are multiple choice of such codewords, $F$ is picked randomly, i.e. each $F$ candidate is picked with equal-probability. Finally, destination will decode $X'$ into $m' = (m'_1, m'_2, \cdots, m'_k)$ and checking the tracing bits:

$$
I_{\text{trial}} = \begin{cases} 
1, & \text{if}(m'_1, m'_2, \cdots, m'_t) = (m_1, m_2, \cdots, m_t) \\
0, & \text{otherwise}
\end{cases}
$$

(5.32)

If $N_{\text{trial}}$ is pretty large, the probability that a false packet is declared as negative (unmodified) can be approximated by

$$
Q_{M,T} \simeq \frac{1}{N_{\text{trial}}} \sum_{\text{trial}=1}^{N_{\text{trial}}} I_{\text{trial}}.
$$

(5.33)

Therefore, the normalized throughput with tracing bits technique is given by

$$
\eta_{\text{Tracing}} = \frac{k - t}{Nk} \left[ \sum_{l=0}^{2^N - 1} (N - |M_l|)P(M_l) \right]
$$

(5.34)

$$
= \frac{(k-t)(1-\alpha)}{k}
$$

(5.35)

where the factor $(k-t)/k$ is to account for the overhead that there has the number of $t$ tracing bits out of $k$ information bits and the probability that a modified packet in $Y_r$ is mistakenly accepted
is given by

\[
\rho_{\text{Tracing}} = \frac{1}{N} \sum_{i=1}^{2^N-2} \sum_{i=1}^{[M_i]} Q_{M,T} P(M_i) = \frac{\alpha}{N} \sum_{i=1}^{N} Q_{M,T} = Q_{M,T} \alpha.
\]

5.4 Numerical Results

![Normalized Throughput vs. Probability of Accepting a Modified Packet](image)

Figure 5.2 Normalized throughput versus probability of mistakenly accepting a modified packet; BCH(63,24), \( \alpha = 0.2 \), \( N = 20 \), \( e_i = 0.1 \) for \( i = 1, \cdots, N/2 \) and \( e_i = 0.05 \) for \( i = N/2 + 1, \cdots, N \).

Fig. 5.3 shows comparison between our MAP+LRT scheme with the existing tracing bit technique. It can be seen that our MAP test detection with LRT data recovery method has significant
improvement over tracing scheme. This is because when a smart adversary perform false data injection attack, he always chooses $F$, which has minimum hamming weight. The less bits modified in encoded codeword ($n$-bits length) usually results in less bits changed in decoded message ($k$-bits length). Since the performance of tracing bits technique highly depends on the number of tracing bits, $t$, if $t$ is small, the highly chance the missed detection would happen; on the other hand, if $t$ is large, the throughput would be reduced by the factor $(k - t)/k$. Hence, our proposed scheme can outperform tracing bit scheme in this minimum false data injection scenario.

Fig. 5.3 illustrates the normalized throughput, $\eta$, versus the probability of mistakenly accepting a modified packet, $\rho$. Figure 5.3 Normalized throughput versus probability of mistakenly accepting a modified packet, $\rho$; BCH(63,24), $\alpha = 0.2$, $N = 20$, $e_i = 0.1$ for $i = 1, \cdots , N/2$ and $e_i = 0.05$ for $i = N/2 + 1, \cdots , N$.

a modified packet, $\rho$, for different detection schemes. $\tau_i$ is varied from 0 to $n$ for MAP+LRT and
LRT, and $B$ is varied from 0 to 128 for MAC. It can be seen that the packet recovery based the MAP test followed by the LRT, denoted by MAP+LRT, provides a significantly higher throughput than the MAP test only (no packet recovery), denoted by MAP. This is due to discarding the entire $N$ packets when even one packet in $Y_r$ is modified and detected by the MAP test. Most of unmodified packets can be recovered, thereby providing the normalized throughput of $1 - \alpha$, by MAP+LRT if the probability of accepting a modified packet is allowed to be $10^{-4}$ or higher. It can also be seen that the throughput can be improved by performing the MAP test prior to the LRT and that the improvement provided by MAP+LRT over LRT is more significant at smaller $\rho$. Also, shown in the figure is the normalized throughput with MAC, denoted by MAC. We can see that MAC can provide a higher throughput than MAP+LRT if $\rho$ is below a threshold and, otherwise, MAP+LRT provides a higher throughput than MAC. It should be noted that MAC requires an additional overhead (beyond the transmission of $B$ bits) for key exchanges between $S_i$ and $D$, $i = 1, \cdots, N$, through separate secure channels, which was not taken into account in the comparison.

5.5 Conclusion

In this chapter, the packet recovery scheme by using likelihood ratio test is developed. We found that the packet recovery based on the MAP test followed by the LRT can provide a significantly higher throughput than that based on the MAP test only. The proposed scheme would be useful in large-scale networked sensing and decision making systems, such as the IoT, where a decision is made based on a large number of short packets that are collected from multiple source nodes.

5.6 Appendix

In this appendix, we provide the proof of (5.21) and (5.23). If we let

$$I_{i,l} = \begin{cases} 1, & i \notin M_l \\ 0, & i \in M_l \end{cases}$$

(5.39)
then \( \sum_{i=1}^{N-|M|} (1 - Q_{F,i}) = \sum_{i=1}^{N} (1 - Q_{F,i})I_{i,l}P(M_l) \). Hence, it follows from (5.20) that

\[
\eta_{\text{LRT}} = \frac{1}{N} \sum_{l=0}^{2^{N-1}} \sum_{i=1}^{N} (1 - Q_{F,i})I_{i,l}P(M_l) \tag{5.40}
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} (1 - Q_{F,i}) \sum_{l=0}^{2^{N-1}} I_{i,l}P(M_l) \tag{5.41}
\]

\[
= \frac{1 - \alpha}{N} \sum_{i=1}^{N} (1 - Q_{F,i}) \tag{5.42}
\]

Similarly, if we let

\[
I_{i,l}^c = \begin{cases} 
1, & i \in M_l \\
0, & i \notin M_l 
\end{cases}
\]

(5.43)

then it follows from (5.22) that

\[
\rho_{\text{LRT}} = \frac{1}{N} \sum_{l=1}^{2^{N-1}} \sum_{i=1}^{N} Q_{M,i}I_{i,l}^cP(M_l) \tag{5.44}
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} Q_{M,i} \sum_{l=1}^{2^{N-1}} I_{i,l}^cP(M_l) \tag{5.45}
\]

\[
= \frac{\alpha}{N} \sum_{i=1}^{N} Q_{M,i} \tag{5.46}
\]

5.7 References


CHAPTER 6. ENHANCE MAP TEST PERFORMANCE WITH CRYPTOGRAPHY FOR SHORT LENGTH, AGGREGATED DATA NETWORKS

6.1 Introduction

To enhance the performance of physical layer MAP test, in this chapter, we propose crypto-aided maximum *a posteriori* probability (MAP) test for detecting false data injection in wireless data collection networks. This scheme catenates physical-layer MAP test after a lightweight cryptographic hash check, by utilizing *a priori* information obtained from cryptographic hash bits, to get the synergy of the two complementary solutions. Since the transmission packets prone to be short and aggregated in the emerging wireless networks, especially as fifth-generation (5G) networks Butun et al. (2013), the traditional integrity check techniques which works for the long data packets may not suitable anymore due to overhead and computation cost it incurs. Hence, our purpose is to provide a high accuracy efficient detection scheme to guarantee data integrity for those high-density wireless networks while operating under resource constraints Zhang et al. (2017); Hu and Qian (2014). The results show that the crypto-aided MAP test has significant improvement over conventional cryptographic methods. This technique will benefit a wide variety of limited resources wireless applications where the traffic type are tend to be short and massive, meanwhile the information needs to be communicated in trustworthy and timely manner.

6.2 Cryptographic Check

As we discussed in Section 3.2, MAC accepts input as a secret key and a message $X_i$ to be authenticated, and outputs a MAC tag, $h(key, X_i)$. The Fig. 6.1 illustrates how does MAC works in multi-resources wireless relay networks, the receiver validates the received packet $Y_{r,i}$ by recomputing the MAC,$h(key, Y_{r,i})$, which is based on $Y_{r,i}$ and the private key, and then comparing
it to the MAC, $h_i$, which is sent by $R$: if $h(key,Y_{r,i}) = h_i$ then $Y_{r,i}$ is considered correct, otherwise, it is considered incorrect.

![Diagram of the Cryptographic Check algorithm]

Figure 6.1 The Cryptographic Check algorithm.

6.2.1 Probability of Detection Error for Cryptographic Check

If the length of $h(key,X_i)$ is $B$ bits, the probability that the adversary randomly picks a hash sequence and gets a modified packet validated by the receiver is $2^{-B}$. However, the true packet is always validated regardless of $B$. Therefore, the probability that a true packet is declared positive (false alarmed) is given by

$$P_{F,C} = 0$$

and the probability that a false packet is declared negative (missed detection) is given by

$$P_{M,C} = 2^{-B}. \quad (6.2)$$
Since the event of missed detection occurs if all the modified packets are classified as unmodified ($H_0$), the average probability of detection error for cryptographic hash check, $P_{E,C}$, is given by

$$P_{E,C} = \sum_{l=1}^{2^N-1} (2^{-B})^{|M_l|} P(M_l) + P_{F,C} P(M_0)$$

(6.3)

$$= \sum_{l=1}^{2^N-1} (2^{-B})^{|M_l|} \alpha^{|M_l|}(1 - \alpha)^{N-|M_l|}. \quad (6.4)$$

Since there are $\binom{N}{j}$ terms of $|M_l| = j$ in (6.4), we obtain

$$P_{E,C} = (1 - \alpha + \alpha 2^{-B})^N - (1 - \alpha)^N. \quad (6.5)$$

Typically, the MAC length $B$ is much smaller than the packet size $n$ in order to make the transmission overhead insignificant. Therefore, for short packets, $B$ needs to kept small. But the drawback of short MAC is that the probability of missed detection is high. Therefore, a further verification step is necessary for the case of $h(key,Y_{r,i}) = h_i$ if $B$ is not sufficient long.

### 6.3 Crypto-Aided MAP Test

![Diagram of Crypto-aided MAP test](image)

**Figure 6.2** Crypto-aided MAP test.
The drawback of short MAC can be addressed by applying the physical-layer MAP test after the cryptographic check to detect any remaining false data that the cryptographic check misses. The proposed architecture, called crypto-aided MAP test, is illustrated in Fig. 6.2. The \( i \)th sender \( S_i \) sends the packet \( X_i \) to the relay \( R \) along with a MAC, \( h(key, X_i) \), of length \( B \) bits. The verifier \( D \) validates the correctness of the received packet \( Y_{r,i} \) by recomputing the MAC \( h(key, Y_{r,i}) \) and comparing it with the MAC, \( h_i \), which is received from \( R \). If \( h(key, Y_{r,i}) = h_i \), the cryptographic check outcome \( \hat{H}_C \) is \( H_0 \), otherwise, \( \hat{H}_C = H_1 \). If \( \hat{H}_C = H_1 \) for at least one received packet, then the number of \( N \) aggregated packets \( Y_r = (Y_{r,1}, Y_{r,2}, \ldots, Y_{r,N}) \) are declared containing modified packets. This is motivated by

\[
Pr(Y_{r,i} \neq X_i|h(key, Y_{r,i}) \neq h(key, X_i)) = 1. \tag{6.6}
\]

Due to the property (6.6), physical-layer MAP in (4.40) test will only be performed if \( \hat{H}_C = H_0 \) for all received packets.

### 6.3.1 a priori information

Before \( D \) starts doing MAP test, in order to apply MAP detection criterion (4.34), the information of \( \alpha \) is necessary to calculate parameter \( b_i \). Recall the discussion in Section 4.3.3: \( \alpha \) is usually consider as a secret only known to attacker, so the maximum likelihood test corresponding to the case where \( \alpha = 0.5 \) is applied. However, the \( \alpha = 0.5 \) in MAP test (4.34) can not be directly used for the crypto-aided MAP test detection part. This is due to the fact that rechecking packets in second phase have already passed the cryptographic hash check, which makes the following MAP test essentially conditional on hash check outcome \( \hat{H}_C = H_0 \). Therefore the attack probability \( \alpha \) \( (P(F_i \neq 0)) \) should be replaced by the posterior probability of the cryptographic check, \( P(F_i \neq 0|\hat{H}_C = H_0) \), which is given by

\[
P(F_i \neq 0|\hat{H}_C = H_0) = \frac{P(\hat{H}_C = H_0, F_i \neq 0)}{P(\hat{H}_C = H_0, F_i \neq 0) + P(\hat{H}_C = H_0, F_i = 0)} \tag{6.7}
\]

\[
= \frac{1}{1 + 2^B(1 - \alpha)/\alpha}. \tag{6.8}
\]
This \textit{a posteriori} probability can be used as the \textit{a priori} probability in the physical-layer MAP test. That is to say, the cryptographic check provides the \textit{a priori} information for the physical-layer MAP test in the proposed crypto-aided MAP test. Hence, by applying maximum likelihood test case, i.e. $\alpha = 1/2$ in (6.8), we obtain \textit{a priori} as

$$\alpha_C = \frac{1}{1 + 2B}. \quad (6.9)$$

Then replacing $\alpha$ in (4.37) by $\alpha_C$, the physical-layer MAP test detection metric of crypto-aided MAP test, $Y_C(Z)$, is given by

$$Y_C(Z) = \sum_{i=1}^{N} \ln \left[ 1 + a_i b_i W(Z_i) \right], \quad (6.10)$$

where

$$a_{i,C} = \frac{\alpha_C}{1 - \alpha_C} \left( \frac{e_i}{1 - e_i} \right)^{d_{\min}} \quad (6.11)$$
$$b_i = \left[ \left( \frac{1 - e_i}{e_i} \right)^2 - 1 \right] \frac{d_{\min}}{n} + 1 \quad (6.12)$$

And we will see later that the performance loss caused by using $\alpha_C$ instead of (6.8) is negligible.

\subsection*{6.3.2 MAP combining rule}

After the physical-layer MAP test, the final decision of crypto-aided MAP test is made based on the cryptographic check outcome $\hat{H}_C$ and the physical-layer MAP test outcome $\hat{H}_P$ using the MAP combining rule:

$$P(H_1|\hat{H}_C, \hat{H}_P) \overset{\hat{H}_{CP}=H_1}{\geq} P(H_0|\hat{H}_C, \hat{H}_P). \quad (6.13)$$

If $(\hat{H}_C, \hat{H}_P) = (H_0, H_0)$, then the crypto-aided MAP test outcome $\hat{H}_{CP} = H_0$, which follows from $P(H_1|(H_0, H_0)) < P(H_0|(H_0, H_0))$. If $(\hat{H}_C, \hat{H}_P) = (H_1, H_0)$ or $(H_1, H_1)$, then $\hat{H}_{CP} = H_1$, which follows from (6.6). If $(\hat{H}_C, \hat{H}_P) = (H_0, H_1)$, then the MAP combining rule in (6.13) is given by

$$\sum_{l=1}^{2^{N-1}} P(\hat{H}_C, \hat{H}_P, M_l) \overset{\hat{H}_{CP}=H_1}{\geq} P(\hat{H}_C, \hat{H}_P, M_0). \quad (6.14)$$
Since
\[ P(\hat{H}_C, \hat{H}_P, M_l) = P(\hat{H}_C = H_0, \hat{H}_P = H_1 | M_l) P(M_l) \] (6.15)
\[ = (2^{-B}|M_l|(1 - Q_{M,l})\alpha|M_l|(1 - \alpha)^N|M_l|, \] (6.16)
\[ P(\hat{H}_C, \hat{H}_P, M_0) = P(\hat{H}_C = H_0, \hat{H}_P = H_1 | M_0) P(M_0) \] (6.17)
\[ = Q_F(1 - \alpha)^N. \] (6.18)

Inserting (6.17) and (6.18) into (6.14), we obtain
\[ 2^{-N-1} \sum_{l=1}^{2N-1} 2^{-B|M_l|}\left(\frac{\alpha}{1 - \alpha}\right)^{|M_l|} \left(\frac{1 - Q_{M,l}}{Q_F}\right)^{\frac{1}{\hat{H}_{CP}=H_1}} \leq 1, \] (6.19)
where \( Q_F := P(\hat{H}_P = H_1 | M_0) \) and \( Q_{M,l} := P(\hat{H}_P = H_0 | M_l) \) are defined in (4.47) and (4.58), respectively. Then, follow the same steps in Section 4.4.1, \( Q_F \) can be approximated by
\[ Q_F \simeq Q\left(\frac{\ln 2 - \mu_0}{\sigma_0}\right), \] (6.20)
where
\[ \mu_0 = \sum_{i=1}^{N} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)f_0(w) \right] \] (6.21)
and
\[ \sigma_0^2 = \sum_{i=1}^{N} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)^2 f_0(w) \right] - \sum_{i=1}^{N} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)f_0(w) \right]^2, \] (6.22)

Similarly, follow the same steps in Section 4.4.2, \( Q_{M,l} \) can be approximated by
\[ Q_{M,l} \simeq Q\left(\frac{\mu_l - \ln 2}{\sigma_l}\right). \] (6.23)
where
\[ \mu_l = \sum_{i \in M_l} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)f_1(w) \right] + \sum_{i \in M_0^l} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)f_0(w) \right] \] (6.24)
and
\[ \sigma_l^2 = \sum_{i \in M_l} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)^2 f_1(w) \right] - \sum_{i \in M_l} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)f_1(w) \right]^2 \]
\[ + \sum_{i \in M_0^l} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)^2 f_0(w) \right] - \sum_{i \in M_0^l} \left[ \sum_{w=0}^{n} \ln(1 + a_{i,C}b_i^w)f_0(w) \right]^2. \] (6.25)
The probability of distribution function (PDF) \( f_0(w) \) and \( f_1(w) \) have been derived in (4.42) and (4.51), respectively. For convenience, the results after MAP combining rule is summarized Table 6.1.

### Table 6.1 MAP combining rule in crypto-aided MAP test.

<table>
<thead>
<tr>
<th>( \hat{H}_C )</th>
<th>( \hat{H}_P )</th>
<th>( \hat{H}_{CP} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_1 )</td>
<td>( H_1 )</td>
<td>( H_1 )</td>
</tr>
<tr>
<td>( H_1 )</td>
<td>( H_0 )</td>
<td>( H_1 )</td>
</tr>
<tr>
<td>( H_0 )</td>
<td>( H_1 )</td>
<td>(6.19)</td>
</tr>
<tr>
<td>( H_0 )</td>
<td>( H_0 )</td>
<td>( H_0 )</td>
</tr>
</tbody>
</table>

The reason behind combining scheme is that, even though the false alarm rate of cryptographic hash check is low as 0, its miss detection rate is relatively high, especially when \( B \) is small. That is to say the detection outcome \( H_{0,C} \) is unreliable to some extent and there may still be some modified packets in those valid ones. Hence, the physical-layer MAP test is performed after cryptographic check. On the other hand, cryptographic hash check detection results can provide a priori information of to help physical layer MAP test in making decision. That is to say, MAP test and cryptographic hash check could mutually benefit from each other.

### 6.4 Probability of Detection error: Crypto-Aided MAP Test

In this section, we derive the probability of detection error for the crypto-aided MAP test. It can be seen from Table 6.1 and (6.19) that

\[
\hat{H}_{CP} = \hat{H}_C \text{ if } \sum_{i=1}^{2^N-1} 2^{-B|M_i|} \left( \frac{\alpha}{1 - \alpha} \right)^{|M_i|} \left( \frac{1 - Q_{M,i}}{Q_F} \right) < 1 \quad (6.26)
\]

Since \( \hat{H}_C = H_0 \) when all \( |M_i| \) modified packets are classified as unmodified \( (H_0) \) given \( M_l \), the probability of missed detection of the crypto-aided MAP test are given by

\[
P_{M_i,CP} = (2^{-B})^{M_i}. \quad (6.27)
\]

And due to \( P_{F,C} = 0 \), the probability of false alarm of the crypto-aided MAP test are given by

\[
P_{F,CP} = 0 \quad (6.28)
\]
Therefore, the average probability of detection error for the crypto-aided MAP test, $P_{E,CP}$, is given by

$$P_{E,CP} = \sum_{l=1}^{2^N-1} P_{M,l,CP}P(M_l) + P_{F,CP}P(M_0)$$  \hspace{1cm} (6.29)$$

$$= (1 - \alpha + \alpha 2^{-B})^N - (1 - \alpha)^N$$  \hspace{1cm} (6.30)$$

for $\sum_{l=1}^{2^N-1} 2^{-B|M_l|}(\frac{\alpha}{1-\alpha})^{|M_l|}(1-\frac{Q_{M,l}}{Q_F}) < 1$, which is also equivalent to (6.5).

It also can be known that if $\sum_{l=1}^{2^N-1} 2^{-B|M_l|}(\frac{\alpha}{1-\alpha})^{|M_l|}(1-\frac{Q_{M,l}}{Q_F}) > 1$, then the MAP combining rule is equivalent to the OR combining rule because of $\hat{H}_{CP} = H_1$. Therefore, the probability of missed detection and false alarm of the crypto-aided MAP test are given by

$$P_{M,l,CP} := P(\hat{H}_C = H_0, \hat{H}_P = H_0|M_l)$$  \hspace{1cm} (6.31)$$

$$= Q_{M,l}(2^{-B}|M_l|)$$  \hspace{1cm} (6.32)$$

and

$$P_{F,CP} := 1 - P(\hat{H}_C = H_0, \hat{H}_P = H_0|M_0)$$  \hspace{1cm} (6.33)$$

$$= Q_F,$$  \hspace{1cm} (6.34)$$

respectively. Therefore, the average probability of detection error for the crypto-aided physical MAP test is given by

$$P_{E,CP} = \sum_{l=1}^{2^N-1} P_{M,l,CP}P(M_l) + P_{F,CP}P(M_0)$$  \hspace{1cm} (6.35)$$

$$= \sum_{l=1}^{2^N-1} Q_{M,l}(2^{-B}|M_l|\alpha^{|M_l|}(1 - \alpha)^N - |M_l|) + Q_F(1 - \alpha)^N,$$  \hspace{1cm} (6.36)$$

for $\sum_{l=1}^{2^N-1} 2^{-B|M_l|}(\frac{\alpha}{1-\alpha})^{|M_l|}(1-\frac{Q_{M,l}}{Q_F}) > 1$.

### 6.5 Selective Crypto-Aided MAP Test

In this section, we consider performing the physical-layer MAP test after the cryptographic check on a subset of received packets, which are selected based on the reliability (BER) of overhearing
channels. This is motivated by high probability of detection error when some of the overhearing channels are unreliable. As Fig. 6.7 have shown that a minor decreases in bit error rate will lead to significant drop on probability of detection error. Consider such circumstance, sometimes sending all the overhearing packets $Y_s$ to physical-layer MAP test process may not provide optimal detection accuracy, this issue may become critical especially when some of the overhearing channels have very bad communication quality, i.e. where the $e_i$ is pretty high. Hence, we develop another detection method call "Selective Crypto-Aided MAP Test" based on the previous crypto-aided map test algorithm to accommodate such situation.

The architecture of selective crypto-aided map test is shown in Fig. 6.3. The cryptographic check process keeps unchanged: $S_i$ still sends the packet $X_i$ to the relay $R$ with $h(key, X_i)$ and the verifier $D$ validates the correctness of the received packet $Y_{r,i}$ by recomputing $h(key, Y_{r,i})$ and comparing it with $h_i$. Then the physical-layer MAP test is performed only if $\hat{H}_C = H_0$ for all received packets. Otherwise the number of $N$ aggregated packets, $\{Y_{r,1}, Y_{r,2}, \cdots, Y_{r,N}\}$, are declared
containing modified packets. Let \( e(1) \leq e(2) \leq \cdots \leq e(N) \) denote the ordered BERs of overhearing channels for the received packets, \( Y_{r(1)}, Y_{r(2)}, \cdots, Y_{r(N)} \), once they all pass the cryptographic check, the physical-layer MAP test is performing on \( Y'_r := (Y_{r(1)}, Y_{r(2)}, \cdots, Y_{r(N)}) \), where \( N' \leq N \); the remaining packets \( Y''_r := (Y_{r(N'+1)}, Y_{r(N'+2)}, \cdots, Y_{r(N)}) \) are classified as "unmodified." Then the MAP combining is performed on \( Y'_r \) only. The crypto-aided MAP test in Section 6.3 corresponds to a special case of \( N' = N \).

### 6.5.1 Average Probability of Detection Error

Compared with Crypto-Aided MAP Test in Section 6.3, the only difference between selective crypto-aided map test and crypto-aided map test is that, instead of deliver the entire number of \( N \) packets, \( \{Y_{r,1}, Y_{r,2}, \cdots, Y_{r,N}\} \), to physical layer MAP test, we only deliver the number of \( N' \) packets, \( \{Y_{r(1)}, Y_{r(2)}, \cdots, Y_{r(N')}\} \), whose overhearing packets’ BER are ranked within top \( N' \) lowest group, to the physical layer MAP test. Hence, the average probability of detection error for selective crypto-aided map test, \( P_{E,SCP} \), share the same form as \( P_{E,CP} \) (6.36), which is given by

\[
P_{E,SCP} = \sum_{l=1}^{2^{N-1}} P_{M_l,SCP}P(M_l) + P_{F,SCP}P(M_0),
\]

(6.37)

where \( P_{M_l,SCP} \) denotes probability of missed detection for under the condition \( M_l \) and \( P_{F,SCP} \) denotes probability of false alarm, for selective crypto-aided map test.

Recall the discussion we had in Section 6.4, similarly, if \( \sum_{l=1}^{2^{N'-1}} 2^{-B|M_l|} (\frac{\alpha}{1-\alpha}|M_l|) (1-\frac{Q_{M|M}}{Q_F}) < 1 \), then the \( \hat{H}_{CP} \) is exactly equal to \( \hat{H}_C \), then selective crypto-aided map test reduces to cryptographic check. Therefore, the average probability of detection error for selective crypto-aided map test,

\[
P_{E,SCP} = \sum_{l=1}^{2^{N'-1}} 2^{-B|M_l|} P(M_l) + 0 \cdot P(M_0)
\]

(6.38)

\[
= (1 - \alpha + \alpha 2^{-B})^N - (1 - \alpha)^N.
\]

(6.39)

For \( \sum_{l=1}^{2^{N'-1}} 2^{-B|M_l|} (\frac{\alpha}{1-\alpha}|M_l|) (1-\frac{Q_{M|M}}{Q_F}) > 1 \), the false alarm event happens if \( \hat{H}_{CP} = H_1 \) under the condition \( M_0 \). Since the unmodified packets can always pass the cryptographic check \( (\hat{H}_C = H_0) \), the false alarm event occurs only if the physical-layer MAP test decision is \( \hat{H}_P = H_1 \). Then the
probability of false alarm for selective crypto-aided MAP test is given by

\[ P_{F,SCP} : = P(\hat{H}_{CP} = H_1|M_0) \] (6.40)
\[ = P(\hat{H}_C = H_0|M_0)P(\hat{H}_P = H_1|M_0) \] (6.41)
\[ = Q_F. \] (6.42)

Similarly, the Missed Detection event happens if \( \hat{H}_{CP} = H_0 \) under the condition \( M_1 \), that is to say the cryptographic check decision is \( \hat{H}_C = H_0 \) and so does the physical layer test decision \( \hat{H}_P = H_0 \). Then the probability of missed detection under the condition \( M_l \) for the selective crypto-aided MAP test is given by

\[ P_{M,l,SCP} : = P(\hat{H}_{CP} = H_0|M_1) \] (6.43)
\[ = P(\hat{H}_C = H_0|M_1)P(\hat{H}_P = H_0|M_1) \] (6.44)
\[ = 2^{-B|M_l|}Q_{M,l}. \] (6.45)

Therefore, the average probability of detection error for selective crypto-aided map test,

\[ P_{E,SCP} = \sum_{l=1}^{2^{N'-1}} Q_{M,l}(2^{-B}|M_l|\alpha^{|M_l|}(1 - \alpha)^{N-|M_l|} + Q_F(1 - \alpha)^N, \] (6.46)

for \( \sum_{l=1}^{2^{N'-1}} 2^{-B|M_l|}(\frac{\alpha}{1 - \alpha})^{|M_l|}(\frac{1-Q_{M,l}}{Q_F}) > 1 \). It should be aware that in (6.46), the \( Q_F \) and \( Q_{M,l} \) still can be evaluated by (6.20) and (6.23), respectively, but with \( Y_r \) change to \( Y'_r \).

6.5.2 The optimal choice of \( N' \)

Even though \( N' \) can be any value \( \in [0, N] \), the optimal value of \( N' \) should be chosen such that the probability of detection error, \( P_{E,SCP} \), can be minimized, i.e.

\[ \min P_{E,SCP} = \min_{0 \leq N' \leq N} \left[ P_{E,SCP}(N') \right], \] (6.47)
where

\[
P_{E,SCP}(N') = \begin{cases} 
(1 - \alpha + \alpha 2^{-B})N - (1 - \alpha)^N, & \text{if } \sum_{l=1}^{2^{N'}-1} 2^{-B|M_l|} (\alpha |M_l| (\frac{1-Q_{M,l}}{Q_F}) < 1 \\
\sum_{l=1}^{2^{N}-1} Q_{M,l} (2^{-B}) |M_l| (1-\alpha)^{N-|M_l|} + Q_F (1-\alpha)^N, & \text{otherwise}
\end{cases}
\]

This process can be done by numerical search: changing \(N'\) from 0 to \(N\) to find smallest \(P_{E,SCP}(N')\), i.e. \(N_{opt}'\) satisfies

\[
P_{E,SCP}(N_{opt}') \leq P_{E,SCP}(N') \text{ for } N' = 0, 1, 2, \cdots, N.
\]

\(6.49\)

### 6.6 Numerical Results

Fig. 6.4 shows probability of detection error versus number of packets, \(N\), for different detection schemes. It can be seen that the probability of detection error for the physical-layer MAP test and the crypto-aided MAP test decreases with increasing \(N\), which indicates that the detection of false data injection becomes more accurate if more packets are collected and tested altogether. This follows from the fact that the probability distribution of the decision statistic \(Y(Z)\) under \(H_0\) and \(H_1\) are separated further from each other and concentrated more on their mean values as \(N\) increases. However, the probability of detection error with the cryptographic integrity check increases with increasing \(N\). This is because the cryptographic check is performed on each packet. The crypto-aided MAP test provides a significantly lower probability of detection error than the cryptographic check and the physical-layer MAP test.

Fig. 6.5 illustrates how the MAP combining rule in (6.14) depends on the bit error rate (BER) given \((\hat{H}_C, \hat{H}_P) = (H_0, H_1)\). One can see that LHS of (6.14) \(\sum_{l=1}^{2^{N}-1} P(\hat{H}_C, \hat{H}_P, M_l)\) decreases monotonically with increasing \(e\), while the RHS of (6.14) \(P(\hat{H}_C, \hat{H}_P, M_0)\) increases monotonically with increasing \(e\). If we let \(e^*\) denote the BER where LHS=RHS, then the MAP combiner decides \(\hat{H}_{CP} = H_1\) if \(e < e^*\) and \(\hat{H}_{CP} = H_0\) if \(e > e^*\). That is, the MAP combiner selects the physical-layer MAP detection output if \(e < e^*\) and, otherwise, selects the cryptographic detection output.

Fig. 6.6 illustrates the probability of detection error versus the MAC length, \(B\), for different
values of $\alpha$. It can be observed that the probability of detection error decreases monotonically with increasing $B$ and its decaying speed keeps changing for smaller $B$ ($B < 12$) but becomes stable for larger $B$ ($B \geq 12$). The reason behind this phenomenon is that: the detection errors caused by the false alarm, $P_{F,CP}P(M_0)$, dominates the probability of detection error for smaller $B$, but its decaying speed keeps increasing with $B$ increases. However the detection errors caused by the missed detection, $P_{M,ICP}P(M_l)$, decaying speed keeps unchanged, so after former term’s decaying speed surpasses it, the detection errors of missed detection soon become dominate part. And that the improvement provided by the cryptographic check is more significant for smaller $B$, which is suitable for short packets. It can also be seen that the probability of detection error of the crypto-aided MAP test with $\alpha_C = 1/(1 + 2^B)$, which does not require $\alpha$, is almost identical.

Figure 6.4  Probability of detection error versus number of source nodes, $N$, for different detection schemes; BCH(63,24), $B = 10$, $\alpha = 0.5$, $e_i = 0.1$ for $i = 1, 2, \ldots, N/2$ and $e_i = 0.05$ for $i = N/2 + 1, \ldots, N$. 

to that with $\alpha_C = 1/(1 + 2^B(1 - \alpha)/\alpha)$, which requires to know $\alpha$. It means that the additional term, $(1 - \alpha)/\alpha$, is insignificant for the crypto-aided MAP test. In addition, if $B$ is sufficiently large, the performance of the crypto-aided MAP test converges to that of the cryptographic check, because the accuracy of cryptographic check is approaching to 100%. Also due to the property of physical-layer MAP test, the larger $\alpha$ results in more accurate detection.

Fig. 6.7 illustrates the probability of detection error versus the overhearing bit error rate (BER) for different values of $N$. The probability of detection error for the crypto-aided MAP test increases with increasing BER since the performance of the physical-layer MAP test deteriorates with increasing BER. Note that the probability of detection error with the crypto-aided MAP decreases faster for larger $N$. This indicates that the effect of overhearing error can be mitigated.
by abundance of source nodes. even if the overhearing BER is high, the crypto-aided MAP test can not have higher probability of detection error than the cryptographic check. This property is particularly desirable for emerging high-density network networks where a real-time decision making is performed based on a large number of short packets that are collected from various sensors.

6.7 Conclusion

In this chapter, the crypto-aided MAP test, which combines physical-layer MAP test with a lightweight cryptographic hash check, to address the false data injection issues in wireless data
Figure 6.7  Probability of detection error versus the overhearing bit error rate for different value of $N$; BCH(63,24), $B = 10$, $\alpha = 0.2$.

collection networks are described. The physical-layer MAP test is performed based on the cryptographic hash check outcomes, by exploiting its a priori information and the impaired information from the overheard packets, the proposed technique can provide high detection accuracy over conventional cryptographic methods. We found that if the number of collected packets are sufficiently large, the probability of detection error can be reduced to zero regardless of the bit error rate in overhearing channel. This characteristic is very suitable for the emerging traffic type where the transmission packets tend to be short and massive, such as 5G communication and IoT. In the future, the proposed technique can be extended to multi-hops relay scenario, which is also a very important infrastructure in wireless relay networks.
6.8 References

