DEFENDING AGAINST MULTIFACETED ATTACKS IN WIRELESS NETWORKS AND POWER GRID NETWORKS

BY

WENKAI WANG

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING

UNIVERSITY OF RHODE ISLAND

2011
It is well known that cyber security plays a critical role in ensuring functionality and reliability of increasingly ubiquitous communication systems and critical infrastructures. Compared with the traditional attacks that target individual users, protocols, or components in the complicated systems, the emerging attacks can 1) exploit collaboration among multiple users or network nodes, 2) exploit vulnerabilities in multiple protocols, which can be in different protocol layers, simultaneously and coordinately, and 3) exploit relationship among multiple network components aiming to cause cascading failures, in which the failure of one or several components propagates to other components.

In this dissertation, we investigate attacks that exploit multiple user, multiple protocols, or multiple components, referred to as multifaceted attacks, in networking systems. We focus on studying two systems: the cognitive radio networks and the power grid networks. In particular, this dissertation has four parts.

Secure Collaborative Spectrum Sensing

Cognitive radio is a revolutionary paradigm to migrate the spectrum scarcity problem in wireless networks. The basic idea is to allow secondary users to use the spectrum that is allocated to the primary user when the primary user is absent. For example, when a TV transmitter (primary user) is not using the allocated spectrum, some mobile users (secondary users) can use this spectrum to exchange data among themselves such as in the mobile ad hoc networks. Therefore, an important task in cognitive radio networks is to detect whether the primary user exists or not. In cognitive radio networks, collaborative spectrum sensing is considered as an effective method to improve the performance of primary user detection through multiple user collaboration. For current collaborative spectrum sensing schemes, secondary users are usually assumed to report their sensing information honestly.
However, it is known that wireless devices can be compromised by malicious parties. Compromised nodes can send false sensing information to mislead the system and undermine the collaboration. In this part, we propose defense methods that can detect untrustworthy secondary users in cognitive radio networks. Compared with existing defense methods, the proposed scheme can effectively differentiate malicious nodes and honest nodes. As a result, it can significantly improve the performance of collaborative spectrum sensing. For example, when there are 10 secondary users, with the primary user detection rate equals to 0.99, one malicious user can make the false alarm rate ($P_f$) increase to 72%. The proposed scheme can reduces it to 5%. Two malicious users can make $P_f$ increase to 85%, the proposed scheme reduces it to 8%.

**Cross Layer Attack and Defense in Cognitive Radio Networks**

The existing research on security issues in cognitive radio networks mainly focuses on attack and defense in individual network layers. However, the attackers do not necessarily restrict themselves within the boundaries of network layers. In this work, we design cross-layer attack strategies that can largely increase the attackers’ power or reducing their risk of being detected. As a case study, we investigate the coordinated report-false-sensing-data attack (PHY layer) and small-back-off-window attack (MAC layer). Furthermore, we propose a trust-based cross-layer defense framework that relies on abnormal detection in individual layers and cross-layer trust fusion. Simulation results demonstrate that the proposed defense framework can significantly reduce the maximum damage caused by attackers.

**Modeling of Cascading Failures in Power Systems - Part I: Foundations, Models, and Assessment Metrics**

With the continuous growing energy demand and environmental concerns, it has recently attracted significant attention of academia, industry, and governments
in the development of a smart electric power grid to provide affordable, reliable, efficient, and secure supply of electricity. Among many enabling technologies toward such a smart grid, security has been widely identified as one of the key components for such a complex system. In Part I of this two-part series study, we present a comprehensive analysis of the foundations, system models, and assessment metrics for power system cascading failures. The proposed models and metrics carefully consider the relationship among multiple components (e.g., substations) in the power grid systems and how such relationship affects the propagation of failures from one component to other parts of the network. The objective in this part is to understand the limitations of traditional largest-load based attack strategies and provide critical insights to understand the power grid behavior subject to complex attacks.

Modeling of Cascading Failures in Power Systems - Part II: Attack Strategies and Simulation Analysis

This is the second part of a two-part study addressing topology-based modeling of cascading failures in power systems. Part I presents a comprehensive analysis of the foundations, system models, and assessment metrics to understand this problem. In Part II, we study specific attack strategies and analyze their simulation results based on the Western North American power grid benchmark under two representative topology based models. The goal is to analyze the power grid behavior and find effective attack strategies when the attacker can take down one or multiple nodes. The first model we investigated is the non-recoverable model, in which overloaded nodes fail to operate, and the second network model is recoverable model, in which overloaded nodes are still in function but their performance in power delivery is reduced. In both network models, the proposed attack strategies, which represent novel ways for joint consideration of load and topology, are
much more destructive than the traditional load based strategies.
ACKNOWLEDGMENTS

My deepest gratitude and earnest appreciate is to my advisor, Dr. Yan Sun. I am very lucky to have this knowledgeable, responsible, considerable, and approachable professor as my advisor. Dr. Sun always gives good guidance on research topic selection. She keeps track of the latest research topics in our area and creatively thinks of new directions. In the meantime, she gives me enough freedom to choose the topics that I am interested in. Dr. Sun is very patient in training our skills in many aspects. She asks us to do presentations in group meeting and provides insightful feedbacks for us to improve. She reads every of my publication drafts and made very detailed edits to make the writing more smooth in both structural and linguistical level. Lastly, Dr. Sun is a friend. She cares about our studies as well as our life. No matter I was stuck in a research problem or I had an internship job, I am always willing to share it with her and listen to her suggestions. She would view the situations from different perspectives and give me useful advice. Without her help, I would not be able to accomplish this dissertation.

I have been working with Dr. Haibo He on the power grid security project since last year. I was greatly influenced by his enthusiasm in research and dedication to work. He is very active in exploiting interesting research problems and striving to find effective approach to solve them. During this period, I learned a lot from him both in problem identifying and solving skills. I would like to thank him for giving me useful advices in research and the inspiration on work.

I would also like to thank Prof. Lisa DiPippo, Prof. Qing Yang, and Prof. Lubos Thoma for serving as my dissertation committee member. Their extraordinary expertise on computer science and mathematics encouraged me to continue studying. I benefit a lot from their insightful comments about my research and dissertation. I appreciate them for taking time to participate in my comprehensive
exam and thesis defense when they are extremely busy with their daily works.

I thank Prof. Zhu Han, University of Houston and Prof. Husheng Li, University of Tennessee. The collaboration with them on secure cognitive radio networks project was pleasant. They taught me a lot on both theoretical derivation and practical simulation parameter settings. Without their help, I would not be able to finish the project.

A special thank you to all the faculty member of the Electrical, Computer, and Biomedical Engineering department at URI. Specifically, I would like to thank Prof. Boudreaux-Bartels, Prof. Godi Fisher, and Prof. J. C. Lo for their help. Also I want to thank my fellow students Yuhong Liu, Xuan Li, Xiang Liu, Yihai Zhu, Yongbo Zeng, and Qiao Cai.

Finally and most importantly, I would like to thank my family. Firstly, I want to thank my mother, Xirong Wu. She is a great woman who never gives up the support for me even in the worst circumstances. Without her, I cannot imagine where I end up with today. I also want to thank my father, Youjin Wang, for his contribution and dedication to the family. Lastly, I want to thank my wife, Yue Zhao. Her lovely smile has always been my joy. The continuous concern, support, encouragement, and love from them is the driving force for me to fulfill this dissertation.
This dissertation is organized in the manuscript format consisting of four manuscripts. The corresponding publications are as follows.

- **Manuscripts 1:**

- **Manuscripts 2:**
  Wenkai Wang, Yan Sun, Husheng Li, and Zhu Han, ”Cross-layer Attack and Defense in Cognitive Radio Networks”, in Proceedings of IEEE Global Communication Conference (Globecom’10), Miami, FL, Dec 2010.

- **Manuscripts 3:**

- **Manuscripts 4:**
  Wenkai Wang, Haibo He, and Yan Sun, ”Topology based Modeling of Cascading Failures in Power Systems - Part II: Attack Strategies and Simulation Analysis”, in submission to IEEE Transactions on Information Forensics and Security, 2011.²

¹This journal paper consists of two parts: Part I and Part II. This is Part I. The earlier conference paper was accepted by IEEE Global Communication Conference in July 2011.
²This is Part II of the journal paper. The earlier conference paper was accepted by IEEE Global Communication Conference in July 2011.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>PREFACE</td>
<td>viii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
<tr>
<td><strong>MANUSCRIPT</strong></td>
<td></td>
</tr>
<tr>
<td>1 Secure Collaborative Spectrum Sensing</td>
<td>2</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>3</td>
</tr>
<tr>
<td>1.2 System Model</td>
<td>6</td>
</tr>
<tr>
<td>1.2.1 Collaborative Spectrum Sensing</td>
<td>7</td>
</tr>
<tr>
<td>1.2.2 Attack Model</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Secure Collaborative Sensing</td>
<td>10</td>
</tr>
<tr>
<td>1.3.1 Single Malicious User Detection</td>
<td>10</td>
</tr>
<tr>
<td>1.3.2 Multiple Malicious Users Detection</td>
<td>15</td>
</tr>
<tr>
<td>1.3.3 Optimal Attack</td>
<td>20</td>
</tr>
<tr>
<td>1.4 Simulation Results</td>
<td>23</td>
</tr>
<tr>
<td>1.4.1 Single Attacker</td>
<td>23</td>
</tr>
<tr>
<td>1.4.2 Multiple Attackers</td>
<td>25</td>
</tr>
<tr>
<td>1.4.3 Dynamic Behaviors</td>
<td>28</td>
</tr>
<tr>
<td>1.4.4 Optimal Attack</td>
<td>31</td>
</tr>
<tr>
<td>1.4.5 Impact of Position Estimation Error upon Performance</td>
<td>36</td>
</tr>
</tbody>
</table>
3.2.3 Load Redistribution Models .................................. 71
3.3 Limitations of Traditional Largest-Load Based Attack Strategy . 73
3.4 Assessment Metrics ................................................. 76
   3.4.1 Assessment Metrics for Non-Recoverable Model ........... 76
   3.4.2 Assessment Metrics for Recoverable Model ................. 83
3.5 Conclusion .......................................................... 85
List of References ..................................................... 86

4 Modeling Cascading Failures in Power Systems - Part II ........ 90
   4.1 Introduction ....................................................... 90
   4.2 Attack Strategies under Non-recoverable Model ............. 94
      4.2.1 Single-node Attack Strategies .......................... 94
      4.2.2 Multi-node Attack Strategies ......................... 95
   4.3 Attack Strategies under Recoverable Model .................. 98
      4.3.1 Single-node Attack Strategies .......................... 98
      4.3.2 Multi-node Attack Strategies ......................... 100
   4.4 Simulation Results ............................................... 103
      4.4.1 Simulation Results for Non-recoverable Model .......... 103
      4.4.2 Simulation Results for Recoverable Model .............. 107
   4.5 Conclusion and Future Work .................................... 112
List of References ..................................................... 113

BIBLIOGRAPHY .......................................................... 116
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Collaborative Spectrum Sensing</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>ROC curves ($P_a = 100%$, FA Attack)</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>ROC curves ($P_a = 50%$, FA Attack)</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>ROC curves ($P_a = 100%$, FAMD Attack)</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>ROC curves ($P_a = 50%$, FAMD Attack)</td>
<td>28</td>
</tr>
<tr>
<td>6</td>
<td>ROC curves (FA Attack, Two Attackers)</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>ROC curves (FA Attack, Three Attackers)</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>ROC curves (FAMD Attack, Two Attackers)</td>
<td>29</td>
</tr>
<tr>
<td>9</td>
<td>ROC curves (FAMD Attack, Three Attackers)</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>Dynamic trust value (a user attacks during time (50, 90))</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>Dynamic trust value (a user attacks during time (50, 55))</td>
<td>31</td>
</tr>
<tr>
<td>12</td>
<td>Dynamic suspicious level (two users attack during time (20, 100))</td>
<td>32</td>
</tr>
<tr>
<td>13</td>
<td>Dynamic suspicious level (two users attack during time (5, 15))</td>
<td>33</td>
</tr>
<tr>
<td>14</td>
<td>Region that detection is impossible</td>
<td>34</td>
</tr>
<tr>
<td>15</td>
<td>Damage in region</td>
<td>35</td>
</tr>
<tr>
<td>16</td>
<td>Impact of Position Estimation Error</td>
<td>37</td>
</tr>
<tr>
<td>17</td>
<td>Cross-Layer Defense Framework</td>
<td>54</td>
</tr>
<tr>
<td>18</td>
<td>Multipath Trust Value Combing</td>
<td>55</td>
</tr>
<tr>
<td>19</td>
<td>Channel Availability - PHY Layer Attack &amp; Defense</td>
<td>56</td>
</tr>
<tr>
<td>20</td>
<td>Transmission Rate - MAC Layer Attack &amp; Defense</td>
<td>57</td>
</tr>
<tr>
<td>21</td>
<td>Transmission Rate - Cross Layer Attack, Single Layer Defense</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>22</td>
<td>Transmission Rate - Cross Layer Attack, Cross Layer Defense</td>
<td>58</td>
</tr>
<tr>
<td>23</td>
<td>Load redistribution example under non-recoverable model</td>
<td>74</td>
</tr>
<tr>
<td>24</td>
<td>Multi-node attack example under recoverable model</td>
<td>75</td>
</tr>
<tr>
<td>25</td>
<td>Percentage-of-failure under three fixed system tolerance values</td>
<td>77</td>
</tr>
<tr>
<td>26</td>
<td>Histograms for Percentage-of-failure, RED, and RIF</td>
<td>78</td>
</tr>
<tr>
<td>27</td>
<td>Percentage-of-failure for two nodes under different T value</td>
<td>79</td>
</tr>
<tr>
<td>28</td>
<td>Required Redundancy (RED) v.s. Risk if Failure (RIF)</td>
<td>81</td>
</tr>
<tr>
<td>29</td>
<td>Initial Load Distribution and RED v.s. Initial Load</td>
<td>82</td>
</tr>
<tr>
<td>30</td>
<td>Attack scenarios for two victim nodes</td>
<td>96</td>
</tr>
<tr>
<td>31</td>
<td>Topology of top 20 largest nodes and their one-hop neighbors</td>
<td>98</td>
</tr>
<tr>
<td>32</td>
<td>Summary of attack strategies</td>
<td>103</td>
</tr>
<tr>
<td>33</td>
<td>RED v.s. K (Single node attack)</td>
<td>105</td>
</tr>
<tr>
<td>34</td>
<td>Percentage-of-failure v.s. K (Single node attack)</td>
<td>106</td>
</tr>
<tr>
<td>35</td>
<td>Percentage-of-failure v.s. System Tolerance (Multi-node attack)</td>
<td>107</td>
</tr>
<tr>
<td>36</td>
<td>Network efficiency changes</td>
<td>108</td>
</tr>
<tr>
<td>37</td>
<td>Average network efficiency v.s. $M$, topology snapshot 1</td>
<td>110</td>
</tr>
<tr>
<td>38</td>
<td>Average network efficiency v.s. $M$, topology snapshot 2</td>
<td>111</td>
</tr>
</tbody>
</table>
Securing Collaborative Spectrum Sensing against Untrustworthy Secondary Users in Cognitive Radio Networks

Wenkai Wang‡, Husheng Li†, Yan Sun‡, and Zhu Han⋆

‡ Department of Electrical, Computer, and Biomedical Engineering
University of Rhode Island, Kingston, RI 02881

† Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN 37996

⋆ Department of Electrical and Computer Engineering
University of Houston, Houston, TX 77004

Corresponding Author: Wenkai Wang

Electrical, Computer, and Biomedical Engineering
University of Rhode Island
4 East Alumni Ave.
Kingston, RI 02881, USA
Phone:+1-401-874-5872
Email address:wenkai@ele.uri.edu
Securing Collaborative Spectrum Sensing against Untrustworthy Secondary Users in Cognitive Radio Networks

Abstract

Cognitive radio is a revolutionary paradigm to migrate the spectrum scarcity problem in wireless networks. In cognitive radio networks, collaborative spectrum sensing is considered as an effective method to improve the performance of primary user detection. For current collaborative spectrum sensing schemes, secondary users are usually assumed to report their sensing information honestly. However, it is known that wireless devices can be compromised by malicious parties. Compromised nodes can send false sensing information to mislead the system. In this paper, we study the detection of untrustworthy secondary users in cognitive radio networks. We first analyze the case when there is only one compromised node in collaborative spectrum sensing schemes. Then we investigate the scenario that there are multiple compromised nodes. Defense schemes are proposed to detect malicious nodes according to their reporting histories. We calculate the suspicious level of all nodes based on their reports. Node will be considered as malicious when its suspicious level goes beyond certain threshold. The reports from malicious nodes will be excluded in decision-making. Results show that even a single malicious node can significantly degrade the performance of spectrum sensing and multiple malicious nodes can make the performance much worse. Compared with existing defense methods, the proposed scheme can effectively differentiate malicious nodes and honest nodes. As a result, it can significantly improve the performance of collaborative sensing. For example, when there are 10 secondary users, with the primary user detection rate equals to 0.99, one malicious user can make the false alarm rate ($P_f$) increase to 72%. The proposed scheme can reduces
it to 5%. Two malicious users can make $P_f$ increase to 85%, the proposed scheme reduces it to 8%.

1.1 Introduction

Nowadays the available wireless spectrum becomes more and more scarce due to increasing spectrum demand for new wireless applications. It is obvious that current static frequency allocation policy cannot meet the needs of emerging applications. Cognitive radio networks [1, 2, 3], which has been widely studied recently, is considered as a promising technology to migrate the spectrum shortage problem. In cognitive radio networks, secondary users are allowed to opportunistically access spectrums which have already been allocated to primary users, given that they do not cause harmful interference to the operation of primary users. In order to access available spectrums, secondary users have to detect the vacant spectrum resources by themselves without changing the operations of primary users. Existing detection schemes include matched filter, energy detection, cyclostationary detection and wavelet detection [2, 3, 4, 5, 6]. Among these schemes, energy detection is commonly adopted because it does not require priori information of primary users.

It is known that wireless channels are subject to fading and shadowing. When secondary users experience multipath fading or happen to be shadowed, they may fail to detect the existence of primary signal. As a result, it will cause interference to primary users if they try to access this occupied spectrum. To cope with this problem, collaborative spectrum sensing [7, 8, 9, 10, 11, 12] is proposed. It combines sensing results of multiple secondary users to improve the probability of primary user detection. There are many works that address the cooperative spectrum sensing schemes and challenges. The performance of hard-decision combining scheme and soft-decision combining scheme is investigated in [7, 8]. In these schemes, all secondary users send sensing reports to a common decision center. Coopera-
tive sensing can also be done in a distributed way, where secondary users collect reports from their neighbors and make the decision individually [13, 14, 15]. Optimized cooperative sensing is studied in [16, 17]. When the channel that forwards sensing observations experiences fading, the sensing performance degrades significantly. This issue is investigated in [18, 19]. Furthermore, energy efficiency in collaborative spectrum sensing is addressed in [20].

There are some works that address the security issues of cognitive radio networks. Primary user emulation attack is analyzed in [21, 22]. In this attack, malicious users transmit fake signals which have similar feature of primary signal. In this way attacker can mislead legitimate secondary users to believe that primary user is present. The defense scheme in [21] is to identify malicious user by estimating location information and observing received signal strength (RSS). In [22], it uses signal classification algorithms to distinguish primary signal and secondary signal. Primary user emulation attack is an outsider attack, targeting both collaborative and non-collaborative spectrum sensing. Another type of attack is insider attack that targets collaborative spectrum sensing. In current collaborative sensing schemes, secondary users are often assumed to report their sensing information honestly. However, it is quite possible that wireless devices are compromised by malicious parties. Compromised nodes can send false sensing information to mislead the system. A natural defense scheme [23] is to change the decision rule. The revised rule is, when there are $k - 1$ malicious nodes, the decision result is on only if there are at least $k$ nodes reporting on. However, this defense scheme has three disadvantages. First, the scheme does not specify how to estimate the number of malicious users, which is difficult to measure in practice. Second, the scheme will not work in soft-decision case, in which secondary users report sensed energy level instead of binary hard decisions. Third, the scheme has very high false alarm rate.
when there are multiple attackers. This will be shown by the simulation results in Section 1.4. The problem of dishonest users in distributed spectrum sensing is discussed in [24]. The defense scheme in this work requires secondary users to collect sensing reports from their neighbors when confirmative decision cannot be made. The scheme is also only applied to hard-decision reporting case. Finally, current security issues in cognitive radio networks, including attacks and corresponding defense schemes, are concluded in [25].

In this paper, we develop defense solutions against one or multiple malicious secondary users in soft-decision reporting collaborative spectrum sensing. We first analyze the single malicious user case. The suspicious level of each node is estimated by their reporting histories. When the suspicious level of a node goes beyond certain threshold, it will be considered as malicious and its report will be excluded in decision-making. Then, we extend this defense method to handle multiple attackers by using an “onion-peeling approach”. The idea is to detect malicious users in a batch-by-batch way. The nodes are classified into two sets, honest set and malicious set. Initially all users are assumed to be honest. When one node is detected to be malicious according to its accumulated suspicious level, it will be moved into malicious set. The way to calculate suspicious level will be updated when the malicious node set is updated. This procedure continues until no new malicious node can be found.

Extensive simulations are conducted. We simulate the collaborative sensing scheme without defense, the straightforward defense scheme in [23], and the proposed scheme with different parameter settings. We observe that even a single malicious node can significantly degrade the performance of spectrum sensing when no defense scheme is employed. And multiple malicious nodes can make the performance even much worse. Compared with existing defense methods, the
proposed scheme can effectively differentiate honest nodes from malicious nodes and significantly improve the performance of collaborative spectrum sensing. For example, when there are 10 secondary users, with the primary user detection rate equals to 0.99, one malicious user can make the false alarm rate ($P_f$) increase to 72%. While a simple defense scheme can reduce $P_f$ to 13%, the proposed scheme reduces it to 5%. Two malicious users can make $P_f$ increase to 85%, the simple defense scheme can reduce $P_f$ to 23%, the proposed scheme reduces it to 8%. We study the scenario that malicious nodes dynamically change their attack behaviors. Results show that the scheme can effectively capture the dynamic change of nodes. For example, if a node behaves well for a long time and suddenly turns bad, the proposed scheme rapidly increase the suspicious level of this node. If it only behaves badly for a few times, the proposed scheme allows slow recovery of its suspicious level.

The rest of paper is organized as follows. Section 1.2 describes the system model. Attack models and the proposed scheme are presented in Section 1.3. In Section 1.4, simulation results are demonstrated. Conclusion is drawn in Section 1.5.

1.2 System Model

Studies show that collaborative spectrum sensing can significantly improve the performance of primary user detection [7, 8]. While most collaborative spectrum sensing schemes assume secondary users are trustworthy, it is possible that attackers compromise cognitive radio nodes and make them send false sensing information. In this section, we describe the scenario of collaborative spectrum sensing and present two attack models.
1.2.1 Collaborative Spectrum Sensing

In cognitive radio networks, secondary users are allowed to opportunistically access available spectrum resources. Spectrum sensing should be performed constantly to check vacant frequency bands. For the detection based on energy level, spectrum sensing is to perform the hypothesis test:

\[ y_i = \begin{cases} n_i, & H_0 \quad \text{(channel is idle)}, \\ h_i s + n_i, & H_1 \quad \text{(channel is busy)}, \end{cases} \tag{1} \]

where \( y_i \) is the sensed energy level at the \( i \)-th secondary user, \( s \) is the signal transmitted by the primary user, \( n_i \) is the additive white Gaussian noise (AWGN), and \( h_i \) is the channel gain from the primary transmitter to the \( i \)-th secondary user.

We denote by \( Y_i \) the sensed energy for the \( i \)-th cognitive user in \( T \) time slots, \( \gamma_i \) the received signal-to-noise-ratio (SNR), and \( TW \) the time-bandwidth product. According to \([7]\), \( Y_i \) follows centralized \( \chi^2 \) distribution under \( H_0 \) and non-centralized \( \chi^2 \) distribution under \( H_1 \):

\[ Y_i \sim \begin{cases} \chi^2_{2TW}, & H_0, \\ \chi^2_{2TW}(2\gamma_i), & H_1. \end{cases} \tag{2} \]

From (2), we can see that under \( H_0 \) the probability \( P(Y_i = y_i|H_0) \) depends on \( TW \) only. Under \( H_1 \), \( P(Y_i = y_i|H_1) \) depends on \( TW \) and \( \gamma_i \). Recall that \( \gamma_i \) is the received SNR of secondary user \( i \), which can be estimated according to path loss model and location information.

By comparing \( y_i \) with a threshold \( \lambda_i \), secondary user makes a decision about whether the primary user is present. As a result, the detection probability \( P_d^i \) and false alarm probability \( P_f^i \) are given by

\[ P_d^i = P(y_i > \lambda_i|H_1), \tag{3} \]

and

\[ P_f^i = P(y_i > \lambda_i|H_0), \tag{4} \]
respectively.

Notice that equation (3) and (4) are detection rate and false rate for single secondary user. In practice it is known that wireless channels are subject to multipath fading or shadowing. The performance of spectrum sensing degrades significantly when secondary users experience fading or happen to be shadowed \[7, 8\]. Collaborative sensing is proposed to alleviate this problem. It combines sensing information of several secondary users to make more accurate detection. For example, considering collaborative spectrum sensing with \(N\) secondary users. When OR-rule, i.e., the detection result of primary user is on if any secondary user reports on, is the decision rule, the detection probability and false-alarm probability for collaborative sensing are \[7, 8\],

\[
Q_d = 1 - \prod_{i=1}^{N} (1 - P^i_d),
\]

(5)

and

\[
Q_f = 1 - \prod_{i=1}^{N} (1 - P^i_f),
\]

(6)

respectively. A scenario of collaborative spectrum sensing is demonstrated in Fig. 1. We can see that with OR rule, decision center will miss detect the existence of primary user only when all secondary users miss detect it.
1.2.2 Attack Model

The compromised secondary users can report false sensing information to the decision center. According to the way they send false sensing reports, attackers can be classified into two categories: selfish users and malicious users. The selfish users report *yes* or high energy level when their sensed energy level is low. In this way they intentionally cause false alarm such that they can use the available spectrum and prevent others from using it. The malicious users report *no* or low signal level when their sensed energy is high. They will reduce the detection rate, which yields more interference to the primary user. When the primary user is not detected, the secondary users may transmit in the occupied spectrum and interfere with the transmission of the primary user. In this paper, we investigate two attack models, *False Alarm (FA) Attack* and *False Alarm & Miss Detection (FAMD) Attack*, as presented in [26, 27].

In energy spectrum sensing, secondary users send reports to decision center in each round. Let $X_n(t)$ denote the observation of node $n$ about the existence of the primary user at time slot $t$. The attacks are modeled by three parameters: the attack threshold ($\eta$), attack strength ($\Delta$), and attack probability ($P_a$). The two attack models are,

- **False Alarm (FA) Attack**: For time slot $t$, if sensed energy $X_n(t)$ is higher than $\eta$, it will not attack in this round, and just report $X_n(t)$; Otherwise it will attack with probability $P_a$ by reporting $X_n(t) + \Delta$. This type of attack intends to cause false alarm.

- **False Alarm & Miss Detection (FAMD) Attack**: For time slot $t$, attacker will attack with probability $P_a$. If it does not choose to attack this round, it will just report $X_n(t)$; Otherwise it will compare $X_n(t)$ with $\eta$. If $X_n(t)$ is higher than $\eta$, the attacker reports $X_n(t) - \Delta$; Otherwise, it reports $X_n(t) + \Delta$. This
type of attack causes both false alarm and miss detection.

1.3 Secure Collaborative Sensing

In this paper, we adopt the centralized collaborative sensing scheme in which
$N$ cognitive radio nodes report to a common decision center. Among these $N$ cogni-
tive radio nodes, one or more secondary users might be compromised by attackers.
We first study the case when only one secondary node is malicious. By calculating
the suspicious level, we propose a scheme to detect malicious user according to
their report histories. Then we extend the scheme to handle multiple attackers.
As we will discuss later, malicious users can change their attack parameters to
avoid being detected, so the optimal attack strategy is also analyzed.

1.3.1 Single Malicious User Detection

In this section, we assume that there is at most one malicious user. Define

$$\pi_n(t) \triangleq P(T_n = M|\mathcal{F}_t) \quad (7)$$

as the suspicious level of node $n$ at time slot $t$, where $T_n$ is the type of node, which
could be H(Honest) or M(Malicious), and $\mathcal{F}_t$ is observations collected from time
slot 1 to time slot $t$. By applying Bayesian criterion, we have

$$\pi_n(t) = \frac{P(\mathcal{F}_t|T_n = M)P(T_n = M)}{\sum_{j=1}^{N} P(\mathcal{F}_t|T_j = M)P(T_j = M)}. \quad (8)$$

Suppose that $P(T_n = M) = \rho$ for all nodes. Then, we have

$$\pi_n(t) = \frac{P(\mathcal{F}_t|T_n = M)}{\sum_{j=1}^{N} P(\mathcal{F}_t|T_j = M)}. \quad (9)$$
It is easy to verify

\[
P(F_i|T_n = M) = \prod_{\tau=1}^{t} P(X(\tau)|T_n = M, F_{\tau-1})
\]

\[
= \prod_{\tau=1}^{t} \left[ \prod_{j=1, j \neq n}^{N} P(X_j(\tau)|T_j = H) \right] P(X_n(\tau)|F_{\tau-1})
\]

\[
= \prod_{\tau=1}^{t} \rho_n(\tau), \tag{10}
\]

where

\[
\rho_n(t) = P(X_n(t)|F_{t-1}) \prod_{j=1, j \neq n}^{N} P(X_j(t)|T_j = H), \tag{11}
\]

which represents the probability of reports at time slot \( t \) conditioned that node \( n \) is malicious. Note that the first equation in (10) is obtained by repeatedly applying the following equation

\[
P(F_i|T_n = M) = P(X(t)|T_n = M, F_{t-1})P(F_{t-1}|T_n = M). \tag{12}
\]

Let \( p_B \) and \( p_I \) denote the observation probabilities under busy and idle states, respectively, i.e.,

\[
p_I(X_j(t)) = P(X_j(t)|S(t) = I), \tag{13}
\]

\[
p_B(X_j(t)) = P(X_j(t)|S(t) = B). \tag{14}
\]

Note that calculation in (13) and (14) is based on the fact that the sensed energy level follows centralized \( \chi^2 \) distribution under \( H_0 \) and non-centralized \( \chi^2 \) distribution under \( H_1 \) [7]. The \( \chi^2 \) distribution is stated in (2), in which the channel gain \( \gamma_i \) should be estimated based on (a) the distance between the primary
transmitter and secondary users and (b) the path loss model. We assume that the primary transmitter (TV tower, etc.) is stationary and the position of secondary users can be estimated by existing positioning algorithms [28, 29, 30, 31, 32]. Of course, the estimated distance may not be accurate. In Section 1.4.5, the impact of distance estimation error on the proposed scheme will be investigated.

Therefore, the honest user report probability is given by

\[
P(X_j(t) | T_j = H) = P(X_j(t), S(t)_B | T_j = H) + P(X_j(t), S(t)_I | T_j = H)
= p_B(X_j(t))q_B(t) + p_I(X_j(t))q_I(t)
\]

(15)

The malicious user report probability, \( P(X_n(t) | \mathcal{F}_{t-1}) \), depends on the attack model. When FA attack is adopted, there are two cases that malicious user will report \( X_n(t) \) in round \( t \). In the first case, \( X_n(t) \) is the actual sensed result, which means \( X_n(t) \) is greater than \( \eta \). In the second case, \( X_n(t) \) is the actual sensed result plus \( \Delta \). So the actual sensed energy is \( X_n(t) - \Delta \) and is less than \( \eta \). In conclusion, the malicious user report probability under FA is,

\[
P(X_n(t) | \mathcal{F}_{t-1}) = P(X_n(t), S(t)_B | \mathcal{F}_{t-1}) + P(X_n(t), S(t)_I | \mathcal{F}_{t-1})
= p_B(X_n(t))P(X_n(t) \geq \eta)q_B(t) +
\]

\[
p_B(X_n(t) - \Delta)P(X_n(t) < \eta + \Delta)q_B(t) +
\]

\[
p_I(X_n(t))P(X_n(t) \geq \eta)q_I(t) +
\]

\[
p_I(X_n(t) - \Delta)P(X_n(t) < \eta + \Delta)q_I(t).
\]

(16)
Similarly, when \textit{FAMD} attack is adopted,

\[
P (X_n(t) | \mathcal{F}_{t-1}) = P (X_n(t), S(t)_B | \mathcal{F}_{t-1}) + P (X_j(t), S(t)_I | \mathcal{F}_{t-1})
\]

\[
= p_B(X_n(t) + \Delta) P (X_n(t) \geq \eta - \Delta) q_B(t) + \\
p_B(X_n(t) - \Delta) P (X_n(t) < \eta + \Delta) q_B(t) + \\
p_I(X_n(t) + \Delta) P (X_n(t) \geq \eta - \Delta) q_I(t) + \\
p_I(X_n(t) - \Delta) P (X_n(t) < \eta + \Delta) q_I(t).
\]

(17)

In (15-17), \(q_B(t)\) and \(q_I(t)\) are the priori probabilities of whether the primary user is present or not, which can be obtained through a two-state Markov chain channel model [33]. The observation probabilities, \(p_B(X_j(t))\), \(p_B(X_n(t) - \Delta)\), and other similar terms can be calculated by equation (13) and (14). \(P(X_n(t) \geq \eta), P(X_n(t) < \eta + \Delta)\), and similar terms, are detection probabilities or false alarm probabilities, which can be evaluated under specific path loss model [7, 8]. Therefore, we can calculate the value of \(\rho_n(t)\) in (11) as long as \(\Delta, \eta, q_B(t), q_I(t), TW,\) and \(\gamma_i\) are known or can be estimated. In this derivation, we assume that the common receiver has the knowledge of the attacker’s policy. This assumption allows us to obtain the performance upper bound of the proposed scheme and reveal insights of the attack/defense strategies. In practice, the knowledge about the attacker’s policy can be obtained by analyzing previous attacking behaviors. For example, if attackers were detected previously, one can analyze the reports from these attackers and identify their attack behavior and parameters. Investigation on the unknown attack strategies will be investigated in the future work.

The computation of \(\pi_n(t)\) is given by

\[
\pi_n(t) = \frac{\prod_{\tau=1}^{t} \rho_n(\tau)}{\sum_{j=1}^{N} \prod_{\tau=1}^{t} \rho_j(\tau)}.
\]

(18)
We convert suspicious level $\pi_n(t)$ into trust value $\phi_n(t)$ as

$$\phi_n(t) = 1 - \pi_n(t).$$  

(19)

Trust value is the measurement for honesty of secondary users. But this value alone is not sufficient to determine whether a node is malicious or not. In fact, we find that trust values become unstable if there is no malicious user at all. The reason is that above deduction is based on the assumption that there is one and only one malicious user. When there is no attacker, the trust values of honest users become unstable. To solve this problem, we define trust consistency value of user $n$ (i.e. $\psi_n(t)$) as

$$\mu_n(t) = \left\{ \begin{array}{ll} \frac{\sum_{t=1}^{L} \phi_n(t)}{L}, & t < L \\ \frac{\sum_{t=L+1}^{\infty} \phi_n(t)}{L}, & t \geq L, \end{array} \right.$$  

(20)

$$\psi_n(t) = \left\{ \begin{array}{ll} \sum_{\tau=1}^{t} (\phi_n(t) - \mu_n(t))^2, & t < L \\ \sum_{\tau=1-L+1}^{\infty} (\phi_n(t) - \mu_n(t))^2, & t \geq L, \end{array} \right.$$  

(21)

where $L$ is the size of the window in which the variation of recent trust values is compared with overall trust value variation.

**Procedure 1** Primary user detection

1: receive reports from $N$ secondary users.

2: calculate trust values and consistency values for all users.

3: for each user $n$ do

4: \hspace{1em} if $\phi_n(t) < \text{threshold}_1$ and $\psi_n(t) < \text{threshold}_2$ then

5: \hspace{2em} the report from user $n$ is removed

6: \hspace{1em} end if

7: end for

8: perform primary user detection algorithm based on the remaining reports.

Procedure 1 shows the process of by applying the trust value $\phi_n(t)$ and the consistency value $\psi_n(t)$ in primary user detection algorithm. The basic idea is to eliminate the reports from users who have consistent low trust values. The
value of $threshold_1$ and $threshold_2$ can be chosen dynamically. This procedure can be used together with many existing primary user detection algorithms such as hard decision combing and soft decision combing. The study in [23] have shown that hard decision performs almost the same as soft decision in terms of achieving performance gain when the cooperative users (10-20) face independent fading. For simplicity, in this paper, we will use the hard decision combining algorithm in [7, 8] to demonstrate the performance of the proposed scheme and other defense schemes.

1.3.2 Multiple Malicious Users Detection

The detection of single attacker is to find the node that has the largest probability to be malicious. We can extend this method to multiple attackers case. The idea is enumerating all possible malicious nodes set and trying to identify the set with the largest suspicious level. We call this method “ideal malicious node detection”. However, as we will discuss later, this method faces the curse of dimensionality when the number of secondary users $N$ is large. As a result, we propose a heuristic scheme named “Onion-peeling approach” which is applicable in practice.

Ideal Malicious Node Detection

For any $\Omega \subset \{1, ..., N\}$ (note that $\Omega$ could be an empty set, i.e. there is no attacker), we define

$$\pi_\Omega(t) \triangleq P(T_n = M, \forall n \in \Omega, T_m = H, \forall m \notin \Omega | \mathcal{F}_t),$$

as the belief that all nodes in $\Omega$ are malicious nodes while all other nodes are honest.

Given any particular set of malicious nodes $\Theta$, by applying Bayesian criterion, we have

$$\pi_\Omega(t) = \frac{P(\mathcal{F}_t | \Omega)P(\Omega)}{\sum_\Theta P(\mathcal{F}_t | \Theta)P(\Theta)}.$$
Suppose that $P(T_n = M) = \rho$ for all nodes. Then, we have

$$
P(\Omega) = \rho^{|\Omega|} (1 - \rho)^{N - |\Omega|},
$$

(24)

where $|\Omega|$ is the cardinality of $\Omega$.

Next, we can calculate

$$
P(\mathcal{F}_t|\Omega) = \prod_{\tau = 1}^t \prod_{j \notin \Omega} P(X_j(\tau)|T_j = H) \prod_{j \in \Omega} P(X_j(\tau)|F, \mathcal{F}_{\tau-1})
$$

$$
= \prod_{\tau = 1}^t \rho_n(\tau),
$$

(25)

where

$$
\rho_n(t) = \prod_{j \notin \Omega} P(X_j(\tau)|T_j = H) \prod_{j \in \Omega} P(X_j(\tau)|F, \mathcal{F}_{\tau-1}).
$$

(26)

For each possible malicious node set $\Omega$, using equation (23)-(26), we can calculate the probability that this $\Omega$ contains only malicious users and no honest users. And we can find the $\Omega(t)$ with largest $\pi_\Omega(t)$ value. Then compare this $\pi_\Omega(t)$ with certain threshold, if it is beyond this threshold, the nodes in $\Omega$ are considered to be malicious.

However, for a cognitive radio network with $N$ secondary users, there are $2^N$ different choices of set $\Omega$. Thus, the complexity grows exponentially with $N$. So this ideal detection of attackers faces the curse of dimensionality. When $N$ is large, we have to use approximation.

**Onion-Peeling Approach**

To make the detection of multiple malicious nodes feasible in practice, we propose a heuristic “onion-peeling approach” that detects the malicious user set in a batch-by-batch way. Initially all nodes are assumed to be honest. We calculate suspicious level of all users according to their reports. When the suspicious level
of a node is beyond certain threshold, it will be considered as malicious and moved into the malicious user set. Reports from nodes in malicious user set are excluded in primary user detection. And the way to calculate suspicious level is updated once the malicious node set is updated. We continue to calculate the suspicious level of remaining nodes until no malicious node can be found.

In the beginning, we initialize the set of malicious nodes, Ω, as an empty set. In the first stage, compute the a posteriori probability of attacker for any node n, which is given by

\[
\pi_n(t) = \frac{P(T_n = M | \mathcal{F}_t)}{P(\mathcal{F}_t | T_n = M) P(T_n = M) + P(\mathcal{F}_t | T_n = H) P(T_n = H)}
\]

(27)

where we assume that all other nodes are honest when computing \( P(\mathcal{F}_t | T_n = M) \) and \( P(\mathcal{F}_t | T_n = H) \). In equation (27) we only calculate the suspicious level for each node rather than that of a malicious nodes set, the computation complexity is reduced from \( O(2^N) \) to \( O(N) \).

Recall that \( X(t) \) denote the collection of \( X_n(t) \), i.e. reports from all secondary nodes at time slot \( t \). It is easy to verify

\[
P(\mathcal{F}_t | T_n = M) = \prod_{\tau=1}^{t} P(X(\tau) | T_n = M, \mathcal{F}_{\tau-1})
\]

\[
= \prod_{\tau=1}^{t} \left[ \prod_{j=1, j \neq n}^{N} P(X_j(\tau) | T_j = H) \right] P(X_n(\tau) | \mathcal{F}_{\tau-1})
\]

\[
= \prod_{\tau=1}^{t} \rho_n(\tau),
\]

(28)

where

\[
\rho_n(t) = P(X_n(t) | \mathcal{F}_{\tau-1}) \prod_{j=1, j \neq n}^{N} P(X_j(t) | T_j = H).
\]

(29)
Here, $P(F_t|T_n = M)$ means the probability of reports at time slot $t$ conditioned that node $n$ is malicious. Note that the first equation in (28) is obtained by repeatedly applying equation (12).

Similarly, we can calculate $P(F_t|T_n = H)$ by

\[
P(F_t|T_n = H) = \prod_{\tau=1}^{t} P(X(\tau)|T_n = H, F_{\tau-1})
\]

\[
= \prod_{\tau=1}^{t} \left[ \prod_{j=1}^{N} P(X_j(\tau)|T_j = H) \right]
\]

\[
= \prod_{\tau=1}^{t} \theta_n(\tau),
\]

(30)

where

\[
\theta_n(t) = \prod_{j=1}^{N} P(X_j(t)|T_j = H).
\]

(31)

As mentioned before, $q_B(t)$ and $q_I(t)$ are the priori probabilities of whether the primary user exists or not, $p_B(X_j(t))$ and $p_I(X_j(t))$ are the observation probabilities of $X_j(t)$ under busy and idle states. An honest user’s report probability can be calculated by equation (15).

Then for each reporting round, we can update each node’s suspicious level based on above equations. We set a threshold $\xi$ and consider $n_1$ as a malicious node when $n_1$ is the first node such that

\[
P(T_{n_1} = M|F_t) \geq \text{threshold}_3.
\]

(32)

Then, add $n_1$ into $\Omega$.

Through equation (27) - (32), we have shown how to detect the first malicious node. In the $k$-th stage, we compute the a posteriori probability of attacker in the same manner of (27). The only difference is that when computing $P(F_t|T_n = M)$
and $P(\mathcal{F}_t|T_n = H)$, we assume that all nodes in $\Omega$ are malicious. Equation (29) and (31) now become (33) and (34) respectively, and they can be seen as the special cases of (33) and (34) when $\Omega$ is empty.

$$
\rho_n(t) = \mathcal{P}(X_n(t)|\mathcal{F}_t\tau^{-1}) \left( \prod_{j=1, j \neq n \in \Omega} P(X_j(t)|T_j = H) \right).
$$

$$
\theta_n(t) = \left( \prod_{j=1, j \notin \Omega} P(X_j(t)|T_j = H) \right) \left( \prod_{j=1, j \in \Omega} P(X_j(t)|T_j = M) \right),
$$

Add $n_k$ to $\Omega$ when $n_k$ is the first node (not in $\Omega$) such that

$$
P(T_{n_k} = M|\mathcal{F}_t) \geq \text{threshold}_3.
$$

Repeat the procedure until no new malicious node can be found. Based on the above discussion, the primary user detection process is shown in Procedure 2. The basic idea is to exclude the reports from users who have suspicious level higher than threshold. In this procedure, $\text{threshold}_3$ can be chosen dynamically. This procedure can be used together with many existing primary user detection algorithms. As discussed in Section 1.3.1, hard decision performs almost the same as soft decision in terms of achieving performance gain when the cooperative users (10-20) face independent fading. So for simplicity, we still use the hard decision combining algorithm in [7, 8] to demonstrate the performance of the proposed scheme.
Procedure 2 Primary user detection

1: initialize the set of malicious nodes.
2: collect reports from $N$ secondary users.
3: calculate suspicious level for all users.
4: for each user $n$ do
5:    if $\pi_n(t) \geq \text{threshold}_3$ then
6:        move node $n$ to malicious nodes set, the report from user $n$ is removed
7:    exit loop
8: end if
9: end for
10: perform primary user detection algorithm based on nodes that are currently assumed to be honest.
11: go to step 2 and repeat the procedure

1.3.3 Optimal Attack

As presented in Section 3.2.2, the attack model in this paper has three parameters: the attack threshold ($\eta$), attack strength ($\Delta$), and attack probability ($P_a$). These parameters determine the power and covertness of the attack. Here, the power of attack can be described by the probability that the attack is successful (i.e. causing false alarm and/or miss detection). The covertness of the attack can be roughly described by the likelihood that the attack will not be detected.

Briefly speaking, when $\eta$ or $P_a$ increases, the attack happens more frequently. When $\Delta$ increases, the attack goal is easier to achieve. Thus, the power of attack increase with $\eta$, $P_a$, and $\Delta$. On the other hand, when the attack power increases, the covertness reduces. Therefore, there is the tradeoff between attack power and covertness.

The attacker surely prefers maximum attack power and maximum covertness. Of course, these two goals cannot be achieve simultaneously. Then, what is the “best” way to choose attack parameters from the attacker’s point of view? In this section, we define a metric called damage that considers the tradeoff between attack power and covertness, and find the attack parameters that maximize the damage. To simplify the problem, we only consider one attacker case in this study.
We first make the following arguments.

- The attacker can damage the system if it achieves the attack goal and is not detected by the defense scheme. Thus, the total damage can be described by *the number of successful attacks before the attacker is detected*.

- Through experiments, we found that the defense scheme cannot detect some conservative attackers, who use very small $\eta$, $\Delta$, and $P_a$ values. It can be proved that all possible values of $\{\eta, \Delta, P_a\}$ that will not trigger the detector form a continuous 3D region, referred to as the *undetectable region*.

- Thus, maximizing the total damage is equivalent to finding attack parameters in the undetectable region that maximize the probability of successful attack.

Based on above arguments, we define damage $D$ as the probability that the attacker achieves the attack goal (i.e. causing false alarm) in one round of collaborative sensing. Without loss generality, we only consider FA attack in this section. In FA attack, when sensed energy $y$ is below attack threshold $\eta$, the attacker will report $\Delta + y$ with probability $P_a$. When $\Delta + y$ is greater than the decision threshold $\lambda$ and the primary user does not present, the attacker causes false alarm and the attack is successful. Thus, the damage $D$ is calculated as:

\[
D = P_a P(y < \eta) P(y + \Delta \geq \lambda | y < \eta) \\
= P_a (\tilde{P}_I P(y < \eta | H_0) P(y + \Delta \geq \lambda | H_0, y < \eta) + \\
\tilde{P}_B P(y < \eta | H_1) P(y + \Delta \geq \lambda | H_1, y < \eta))
\]  

(36)

where $\tilde{P}_I$ is the priori probability that channel is idle and $\tilde{P}_B$ is the priori probability that channel is busy.
From the definition of $P_d$ and $P_f$ in (3) and (4), we have,

$$P(y < \eta | H_0) = 1 - P_f(\eta)$$  
(37)

$$P(y < \eta | H_1) = 1 - P_d(\eta)$$  
(38)

Similarly,

$$P(y + \Delta \geq \lambda | H_1, y < \eta) = P(\lambda - \Delta \leq y < \eta | H_0)$$

$$= P_f(\lambda - \Delta) - P_f(\eta),$$  
(39)

$$P(y + \Delta > \lambda | H_1, y < \eta) = P(\lambda - \Delta \leq y < \eta | H_1)$$

$$= P_d(\lambda - \Delta) - P_d(\eta),$$  
(40)

Substitute (37) - (40) to (36), we have,

$$D = P_a(\tilde{P}_I(1 - P_f(\eta))(P_f(\lambda - \Delta) - P_f(\eta)) + \tilde{P}_B(1 - P_d(\eta))(P_d(\lambda - \Delta) - P_d(\eta))).$$  
(41)

Under the attack models presented in this paper, the attacker should choose the attack parameters that maximize $D$ and are in the undetectable region.

Finding optimal attack has two purposes. First, with the strongest attack (in our framework), we can evaluate the worst-case performance of the proposed scheme. Second, it reveals insights of the attack strategies. Since it is extremely difficult to obtain the close form solution of the undetectable region, we will find undetectable region through simulations and search for optimal attack parameters using numerical methods. Details will be presented in Section 1.4.4.
1.4 Simulation Results

We simulate a cognitive radio network with $N (=10)$ secondary users. Cognitive radio nodes are randomly located around the primary user. The minimum distance from them to primary transmitter is 1000m and maximum distance is 2000m. The time-bandwidth product $[7, 8]$ is $m = 5$. Primary transmission power and noise level are 200mw and -110dBm respectively. The path loss factor is 3 and Rayleigh fading is assumed. Channel gains are updated based on node’s location for each sensing report. The attack threshold is $\eta = 15$, the attack strength is $\Delta = 15$, and the attack probability $P_a$ is 100% or 50%. We conduct simulations for different choices of thresholds. Briefly speaking, if trust value threshold $\text{threshold}_1$ is set too high or suspicious level threshold $\text{threshold}_3$ is set too low, it is possible that honest nodes will be regarded as malicious. If trust consistency value $\text{threshold}_2$ is set too low, it will take more rounds to detect malicious users. In simulation, for single malicious node detection, we choose the trust value threshold $\text{threshold}_1 = 0.01$, the consistency value threshold $\text{threshold}_2 = 0.1$, and the window size for calculating consistency value is $L = 10$. For multiple malicious users detection, the suspicious level threshold $\text{threshold}_3$ is set to 0.99.

1.4.1 Single Attacker

Three schemes of primary user detection are compared.

- OR Rule: The presence of primary user is detected if one or more secondary users’ reported value is greater than certain threshold. This is the most common hard fusion scheme.

- Ki Rule: The presence of primary user is detected if $i$ or more secondary users’ reported value is greater than certain threshold. This is the straightforward defense scheme proposed in [23].
Proposed Scheme: Use OR rule after removing reports from malicious nodes.

Performance of these schemes are shown by Receiver Operating Characteristic (ROC) curves, which is a plot of the true positive rate vs. the false positive rates as its discrimination threshold is varied. Fig. 2 - 5 show ROC curves for primary user detection in 6 cases when only one secondary user is malicious. Case 1 is for OR rule with $N$ honest users. Case 2 is for OR rule with $N - 1$ honest users. In Case 3 - 6, there are $N - 1$ honest users and one malicious user. Case 3 is for OR rule. Case 4 is for K2 rule. Case 5 is for the proposed scheme with $t = 250$, where $t$ is the index of detection rounds. Case 6 is for the proposed scheme with $t = 500$.

When the attack strategy is the FA Attack, Fig. 2 and Fig. 3 show the ROC curves when the attack probability is 100% and 50%, respectively. The following observations are made.

- By comparing the ROC for Case 1 and Case 3, we see that the performance of primary user detection degrades significantly even when there is only one malicious user. This demonstrates the vulnerability of collaborative sensing, which leads inefficient usage of available spectrum resource.

- The proposed scheme demonstrates significant performance gain over the scheme without defense (i.e. OR rule) and the straightforward defense scheme (i.e. K2 rule). For example, Table 1 shows the false alarm rate ($P_f$) for two given detection rate ($P_d$), when attack probability ($P_a$) is 1. When the attack probability is 0.5, the performance advantage is smaller but still large.

- In addition, as $t$ increases, the performance of the proposed scheme gets close to the performance of Case 2, which represents perfect detection of the malicious nodes.
Figure 2. ROC curves for different collaborative sensing schemes ($P_a = 100\%$, False Alarm Attack)

<table>
<thead>
<tr>
<th></th>
<th>OR Rule</th>
<th>Ki Rule</th>
<th>Proposed ($t = 250$)</th>
<th>Proposed ($t = 500$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA, $P_a = 1$</td>
<td>0.72</td>
<td>0.13</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>FA, $P_a = 0.5$</td>
<td>0.36</td>
<td>0.07</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>FAMD, $P_a = 1$</td>
<td>0.74</td>
<td>0.20</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>FAMD, $P_a = 0.5$</td>
<td>0.37</td>
<td>0.10</td>
<td>0.06</td>
<td>0.04</td>
</tr>
</tbody>
</table>

1.4.2 Multiple Attackers

Fig. 6 - 9 are the ROC curves for six cases when there are multiple attackers. Similarly, Case 1 is $N$ honest users, no malicious node, and OR rule. Case 2 is $N - 2$ (or $N - 3$) honest users, no attacker, and OR rule. Case 3-6 is $N - 2$ (or $N - 3$) honest users and 2 (or 3) malicious users. OR rule is used in Case 3 and Ki rule is used in case 4. Case 5 and Case 6 are with the proposed scheme with different detection rounds. Case 5 is the performance evaluated at round $t = 500$ and Case 6 is at round $t = 1000$.

When the attack strategy is the FA Attack, Fig. 6 and Fig. 7 show the ROC curves when the attacker number is 2 and 3, respectively. We still compare
Figure 3. ROC curves for different collaborative sensing schemes ($P_a = 50\%$, False Alarm Attack)

the three schemes described in Section 1.4.1. Similarly, following observations are made.

- By comparing the ROC curves for Case 1 and Case 3, we see that the performance of primary user detection degrades significantly when there are multiple malicious users. And the degradation is much more severe than single malicious user case.

- The proposed scheme demonstrates significant performance gain over the scheme without defense (i.e. OR rule) and the straightforward defense scheme (i.e. Ki rule). Table 2 shows the false alarm rate ($P_f$) when detection rate is $P_d = 99\%$.

<table>
<thead>
<tr>
<th></th>
<th>OR Rule</th>
<th>Ki Rule</th>
<th>Proposed ($t = 500$)</th>
<th>Proposed ($t = 1000$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA,2 Attackers</td>
<td>0.85</td>
<td>0.23</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>FA,3 Attackers</td>
<td>0.88</td>
<td>0.41</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>FAMD,2 Attackers</td>
<td>0.88</td>
<td>0.31</td>
<td>0.15</td>
<td>0.09</td>
</tr>
<tr>
<td>FAMD,3 Attackers</td>
<td>0.89</td>
<td>0.50</td>
<td>0.26</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Figure 4. ROC curves for different collaborative sensing schemes ($P_a = 100\%$, False Alarm & Miss Detection Attack)

- When there are three attackers, false alarm rates for all these schemes become larger, but the performance advantage of the proposed scheme over other schemes is still large.

- In addition, as $t$ increases, the performance of the proposed scheme becomes close to the performance of Case 2, which is the performance upper bound.

Fig. 4 and Fig. 5 show the ROC performance when the malicious user adopts the FAMD attack. We observe that the FAMD attack is stronger than FA. In other words, the OR rule and K2 rule have worse performance when facing the FAMD attack. However, the performance of the proposed scheme is almost the same under both attacks. That is, the proposed scheme is highly effective under both attacks, and much better than the traditional OR rule and the simple defense K2 rule. The example false alarm rates are listed as follows.

Fig. 8 and Fig. 9 shows the ROC performance when the schemes face the FAMD attack for multiple malicious users. We observe that the FAMD attack is stronger than FA. Compared to the cases with FA attack, performance of the OR rule and Ki rule is worse when facing the FAMD attack. However, the performance
of the proposed scheme is almost the same under both attacks. That is, the proposed scheme is highly effective under both attacks, and much better than the traditional OR rule and the simple defense Ki rule. The examples of false alarm rate are listed in Table 1.

1.4.3 Dynamic Behaviors

We also analyze the dynamic change in behavior of malicious nodes for FAMD attack. Fig. 10 and Fig. 11 are for single malicious user. In Fig. 10, the malicious

![Figure 5: ROC curves for different collaborative sensing schemes (P_a = 50%, False Alarm & Miss Detection Attack)](image)

![Figure 6: ROC curves (FA Attack, Two Attackers)](image)
user changes the attack probability from 0 to 1 at $t = 50$ and from 1 to 0 at time $t = 90$. The dynamic change of trust value can be divided into three intervals. In Interval 1, $t \in [0, 50]$, malicious user does not attack. The trust value of malicious user and honest user are not stable since there is no attacker. Note that the algorithm will not declare any malicious nodes because the trust consistency levels are high. In Interval 2, $t \in [50, 65]$, malicious user starts to attack, and its trust value quickly drops when it turns from good to bad. In Interval 3, where $t > 60$, the trust value of malicious user is consistently low. In Fig. 11, one user behaves badly in only 5 rounds starting at $t = 50$. We can have similar observations. In
Figure 9. ROC curves (FAMD Attack, Three Attackers)

Figure 10. Dynamic trust value in proposed scheme (a user attacks during time [50, 90], $P_a = 1.$)

Interval 1, malicious user does not attack. It has high trust value. Please note that these dynamic figures are just snap shots of trust values. In Fig. 11, the trust value in Region 1 does not fluctuate as frequently as that in Fig. 10. This is also normal. The reason for unstable trust value may due to channel variation or unintentional errors. In Interval 2, $t \in [50, 55]$, malicious user starts to attack, its trust value drops quickly. In Interval 3, where $t > 55$, trust value of malicious user recovers very slowly.
Similarly, we also make observations for dynamic change in behaviors for multiple attackers. Suspicious level of honest users and malicious users are shown in Fig. 12 and 13. Please note that we only demonstrate suspicious level curve for one honest node. The malicious user adopts the FA attack and dynamically chooses which round to start attack and which round to stop attack. In Fig. 12, the malicious users start to attack at $t = 20$ and stop to attack at time $t = 100$. In Fig. 13, one user behaves badly in only 10 rounds starting at $t = 5$. Similar observations can be made. We can see that the suspicious level of malicious nodes increases steadily when nodes turn from good to bad. And the scheme allows slow recovery of suspicious level for occasional bad behaviors.

1.4.4 Optimal Attack

As discussed in Section 1.3.3, given the defense scheme, the attacker can find the optimal attack parameters that maximize the damage. In this set of experiments, we find the optimal attack parameters and evaluate the worst performance of the proposed scheme.

We assume that there are $N = 10$ cognitive radio nodes performing collabo-
we refer to additive sensing. We set the decision threshold $\lambda$ so that the overall detection rate $P_d$ is 99\% when all users are honest. When OR rule is used, $\lambda = 28$ leads to $P_d = 99\%$.

Obviously, the practical values of $\eta$ and $\Delta$ cannot be over certain range. Within the range, for each pair of $(\eta, \Delta)$, we run simulations to identify the maximum attack probability $P_a$ that the attacker can use and avoid being detected. In particular, binary search is used to find the maximum $P_a$. We first try an initial $P_a$, which is usually the $P_a$ value of a neighbor pair. For example, if we already obtain the $P_a$ for pair $(\eta - 1, \Delta)$ through simulation, then normally the maximum $P_a$ for pair $(\eta, \Delta)$ is a little bit smaller than that of pair $(\eta - 1, \Delta)$. Then, we run the simulation for 2000 rounds. If the attacker is not detected within 2000 rounds, we will search the middle value of range $(P_a, 1)$, otherwise we search the middle value of range $(0, P_a)$. The search continues until the maximum $P_a$ is found. Then, the boundary of undetectable region is determined. We would like to point out that there exists more computational efficient ways to search for the undetectable region, which can be exploited in the future work.
Fig. 13. Dynamic suspicious level in proposed scheme (two malicious nodes perform FA attack during time [5, 15].)

Fig. 14 shows the undetectable region when $N = 10$ and other simulation parameters are the same as these in Section 1.4. The X-axis and Y-axis are attack threshold $\eta$ and attack strength $\Delta$ respectively, and Z-axis is attack probability $P_a$. The following observations are made. When $\eta$ and $\Delta$ are small, $P_a$ can be as large as 100%. This is easy to understand. If $\eta$ is small, the probability that sensed energy is below $\eta$ is small. If $\Delta$ is small, the reporting values are just a little higher than true sensed values. Thus, when both $\eta$ and $\Delta$ are small, the behavior of malicious node is not very different from that of honest nodes. Each attack is very weak and the attacker can do more attacks (i.e. larger $P_a$) without triggering the detector. As $\eta$ or $\Delta$ increases, the maximum allowed attack probability $P_a$ decreases. When both $\eta$ and $\Delta$ are large, $P_a$ should be very small (0-5%).

According to (41), we know that the maximum damage will occur at the boundary of the undetectable region. Using (41), we can find the point (i.e. attack parameters) that maximizes the damage in the undetectable region. In this experiment, the optimal attack parameters are $\eta = 16$, $\Delta = 23$, and $P_a = 0.05$, the maximum damage is 0.02.
We also plot the damage in Figure 15. The X-axis and Y-axis are $\eta$ and $\Delta$ respectively, and Z-axis is damage $D$. The damage value is calculated for the boundary points of the undetectable region. We do not show the $P_a$ value because each $(\eta, \Delta)$ pair corresponds to one $P_a$ value on the boundary. From this figure, we can see that when $\eta$ and $\Delta$ are low, the damage is 0. The attacker can cause larger damage by choosing relatively large $\eta$ and $\Delta$ values and small $P_a$ values.

With the optimal attack parameters, for decision threshold $\lambda = 28$, the overall false alarm rate will increase from 1% to 3%. Recall that the decision threshold was determined to ensure 99% detection rate. This is the worst-case performance of the proposed scheme. Please note that this is the worst case when the attackers are undetectable. When malicious users can be detected, as discussed in Section 1.4.1, the performance will get close to upper bound (the performance of $N - 1$ honest nodes) as detection round $t$ increases.

For K2 rule with $N = 10$ secondary users, to maintain overall detection rate $P_d$ being 99%, the decision threshold $\lambda$ should be decreased to 22. Because K2 rule does not try to detect malicious users, attacker has no risk of being detected even they launch the strongest attack. For our attack model, they can set attack
probability $P_a$ to 1, and set attack threshold $\eta$ and attack strength $\Delta$ as large as possible. For K2 rule, when two or more secondary users report on, the decision result is on. The attacker can launch the strongest attack which is similar to report on in hard-decision reporting case. But only when another one or more honest nodes also make false alarm, the attacker can mislead the decision center. So the overall false alarm rate is not 1. In the simulation, we set $P_a$ to 1, $\eta$ and $\Delta$ both to 1000. The overall false alarm rate is 17.5% for K2 rule under these settings, which is much larger than the worst case of the proposed scheme. For OR rule, the overall false alarm rate is 1. This result is summarized in Table 3. In this table, the ideal case means all $N$ secondary users are honest, and other three columns are the worse performance for different schemes when one of the $N$ cognitive radio nodes is malicious.

<table>
<thead>
<tr>
<th></th>
<th>Ideal Case</th>
<th>Proposed Scheme</th>
<th>K1 Rule</th>
<th>OR Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.03</td>
<td>0.175</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Finally, we would like to point out that the optimal attack is only optimal under certain attack model and certain defense scheme. The method of finding
the optimal attack can be extended to study other attack models. We believe the proposed scheme will still work very well under many other attack models, since the attacker’s basic philosophies are similar.

1.4.5 Impact of Position Estimation Error upon Performance

Recall that the proposed scheme needs to know the channel gains that are estimated based on the position of secondary nodes. There are many existing schemes that estimate the location of wireless devices in sensor networks [27-31]. These schemes can be classified into two categories: range based and range free. The range based methods first estimate the distances between pairs of wireless nodes and then calculate the position of individual nodes. Examples of range based schemes are Angle of Arrival (AoA) [28], Received Signal Strength Indicator (RSSI) [29], Time of Arrival (ToA) [30], and Time Difference of Arrival (TDoA) [31]. The range free methods usually use connectivity information to identify the beacon nodes within radio range and then estimate the absolute position of non-beacon nodes [32].

The performance of these schemes are measured by the location estimate error, which is usually normalized to the units of node radio transmission range (R). Most current algorithms can achieve the accuracy that the estimation error is less than one unit of radio transmission range [28, 29, 30, 31, 32].

In this section, we study the impact of position estimation error on the proposed scheme. The simulation settings are mostly the same as the settings in previous experiments. We choose the decision threshold $\lambda = 28$ to ensure the overall detection rate $P_d$ be 99% when there are no malicious nodes. The radio transmission range is set to 50m, which is a typical value for wireless sensor nodes. Both FA attack and FAMD attack with single attacker are simulated.

The proposed scheme needs a certain number of rounds to detect the malicious
users. When the positions of secondary users are not accurate, it can be envisioned that the number of rounds needed to detect the malicious user will increase. In Fig. 16, the horizontal axis is the normalized position estimation error, and the veridical axis is the averaged number of rounds needed to detect the malicious node. In particular, when the normalized position estimation error value is $e$ and the actual distance between primary transmitter and secondary user $i$ is $r_i$, we simulate the case that the estimated distance between the secondary users and the primary transmitter is Gaussian distributed with mean being $r_i$ and variance being $(eR)^2$. From Fig. 16, the following observations are made.

- The average number of rounds to detect malicious node is very stable when the position estimation error is within 4 units of radio range. Recall that most positioning estimate algorithms have the estimation error around 1 unit of radio range. Thus, the performance of the proposed scheme is stable given realistic positioning estimation errors.

- When estimation error goes beyond 4 units of radio range, it would take much more rounds to detect the malicious node.
• The position estimation error has similar impact on the FA attack and the FAMD attack.

In conclusion, the performance of the proposed scheme is not sensitive to the position estimate error as long as it is within a reasonable range. This reasonable range can be achieved by existing positioning algorithms.

1.5 Conclusions

Untrustworthy secondary users can significantly degrade the performance of collaborative spectrum sensing. We propose two attack models, FA attack and FAMD attack. The first attack intends to cause false alarm and the second attack causes both false alarm and miss detection. To deal with these attacks, we first propose a defense scheme to detect single malicious user. The basic idea is to calculate the trust value of all secondary nodes based on their reports. Only reports from nodes that have consistent high trust value will be used in primary user detection. Then we extend the method for single attacker to multiple attacker case. This defense scheme uses an onion-peeling approach and does not need prior knowledge about the attacker number. Finally, we define the damage metric and investigate the attack parameters that maximize the damage.

Comprehensive simulations are conducted to study the ROC curves and suspicious level dynamics for different attack models, attacker numbers and different collaborative sensing schemes. The proposed schemes demonstrate significant performance advantage. For example, when there are 10 secondary users, with the primary user detection rate equals to 0.99, one malicious user can make the false alarm rate ($P_f$) increases to 72%. Whereas the K2 rule defense scheme can reduce $P_f$ to 13%, the proposed scheme reduces $P_f$ to 5%. Two malicious users can make the false alarm rate ($P_f$) increases to 85%. Whereas the K3 defense scheme can reduce $P_f$ to 23%, the proposed scheme reduces $P_f$ to 8%. Furthermore, when a good
user suddenly turns bad, the proposed scheme can quickly increase the suspicious level of this user. If this user only behaves badly for a few times, its suspicious level can recover after a large number of good behaviors. For single attacker case, we find optimal attack parameters for the proposed scheme. When facing the optimal attack, the proposed scheme yield 3% false alarm rate, with 99% detection rate. On the other hand, when the K2 rule scheme faces the strongest attack against the K2 rule, the false alarm rate can be 17.5% with 99% detection rate. With the proposed scheme, the impact from malicious users is greatly reduced even if the attacker adopts optimal attack parameters and remains undetected.

List of References


Manuscript-II

Published in Proceedings of IEEE Global Communication Conference, 2010

Cross-layer Attack and Defense in Cognitive Radio Networks

Wenkai Wang‡, Yan Sun‡, Husheng Li†, and Zhu Han⋆

‡ Department of Electrical, Computer, and Biomedical Engineering
University of Rhode Island, Kingston, RI 02881

† Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN 37996

⋆ Department of Electrical and Computer Engineering
University of Houston, Houston, TX 77004

Corresponding Author: Wenkai Wang
Electrical, Computer, and Biomedical Engineering
University of Rhode Island
4 East Alumni Ave.
Kingston, RI 02881, USA
Phone:+1-401-874-5872
Email address:wenkai@ele.uri.edu
Cross-Layer Attack and Defense in Cognitive Radio Networks

Abstract

The existing research on security issues in cognitive radio networks mainly focuses on attack and defense in individual network layers. However, the attackers do not necessarily restrict themselves within the boundaries of network layers. In this paper, we design cross-layer attack strategies that can largely increase the attackers’ power or reducing their risk of being detected. As a case study, we investigate the coordinated report-false-sensing-data attack (PHY layer) and small-back-off-window attack (MAC layer). Furthermore, we propose a trust-based cross-layer defense framework that relies on abnormal detection in individual layers and cross-layer trust fusion. Simulation results demonstrate that the proposed defense framework can significantly reduce the maximum damage caused by attackers.

2.1 Introduction

Cross-layer design, a concept introduced to increase network efficiency through information exchange among different layers, has brought revolutionary view change to the networking research community in the past. Nowadays, the increasingly ubiquitous and distributed networking systems are facing brutal and intelligent attacks that exploit almost all network protocols and surely do not restrict themselves within the boundaries of network layers. Attackers have the capability to launch attacks in multiple layers simultaneously [1, 2]. Smart attackers can coordinate the attack activities in different layers to better achieve their goals. The capability of attackers is even strengthened by cognitive radio [3, 4] technology, which makes network protocols more dynamic, adaptive and programmable.

A cognitive radio device can dynamically program its transmission parameters
according to the surrounding wireless channel conditions. As a result, it can make use of the under-utilized frequency bands and mitigate the increasing demand for spectrum resources. Meanwhile, it also brings new vulnerabilities. In the physical (PHY) layer, there are primary user emulation attack [5] and reporting false sensing data attack [6, 7]. In the MAC layer, there are small-back-off-window attack [8], reporting false selection frame[9], and common control channel denial-of-service attacks [9]. Furthermore, many higher layer attacks against traditional wireless networks can also apply to cognitive radio networks.

In current literature, the effectiveness of these attacks and their defense methods are mostly studied independently. However, a smart attacker can launch several attacks in different layers coordinately, which is referred as the cross-layer attack in this paper. Can attackers significantly increase the damage or reduce the risk of being detected by launching the cross-layer attacker? What are the effective defense strategies?

In this paper, we gain insights for answering the above questions by investigating a cross-layer attack in cognitive radio networks. We choose the reporting false sensing data attack [6, 7] in PHY layer and the small-back-off-window attack in MAC layer [8]. The goal of the attack is to reduce channel utilization. Particularly, we

- propose a cross-layer defense architecture, which relies on trust evaluation in individual layers and trust fusion across multiple-layers;
- modify/develop anomaly detection in individual layers;
- design trust fusion algorithm that considers the diverse performance of anomaly detection in different layers;
- demonstrate the significant increase of the attackers’ power due to cross-layer
attack strategies, as well as the effectiveness of cross-layer defense in terms of reducing maximum damage caused by attackers.

The rest of the paper is organized as follows. Section 2.2 summarizes the related works and background. Section 2.3 describes the details of single layer attacks and the defense in individual layers, the cross-layer attack strategy, and the cross-layer defense architecture. Simulation results are shown in Section 4.4 and conclusion is drawn in Section 4.5.

2.2 Background and Related Work

The proposed cross-layer defense architecture is applicable to any distributed networking systems, in which coordinated attack activities can occur simultaneously in multiple layers. In this paper, we study PHY layer and MAC layer in cognitive radio networks to demonstrate the proposed ideas.

**PHY Layer Attack** In cognitive radio network, secondary users (without license) are allowed to access the licensed spectrum if primary users (having license) are not present. To protect the priority of primary users, secondary users must quit the spectrum when primary users emerge. Therefore, secondary users need to carry out *spectrum sensing* to detect the existence of primary users. In PHY layer, there are two types of attack against spectrum sensing: Primary User Emulation attack (PUE) [5] and Reporting False Sensing Data Attack (RFSD) [6, 7]. In this work, we choose RFSD attack in the PHY layer to demonstrate cross-layer attack/defense strategies. RFSD attack targets *collaborative spectrum sensing*, in which the final spectrum sensing results are based on the sensing reports from multiple secondary users[10]. In RFSD attack, malicious users mislead the final sensing results by sending false sensing reports, which may result in inefficient usage of spectrum resource or interference to primary user.

**MAC Layer Attack** Some MAC layer protocols [11] adopt a spectrum access
scheme similar to IEEE 802.11 DCF, the CSMA/CA protocol. That is, in sensing period secondary users scan channels and get the availability information. In transmission period they back off a random time and then select some channels to transmit. If there is a collision, they will double the backoff window size and retransmit. However, a greedy node can use a small backoff window and gain priority on channel access over other nodes, referred to as the small-back-off-window (SBW) attack in this paper. Although there are many other attacks in the MAC layer, we choose SBW attack to demonstrate cross-layer attack/defense strategies.

**Attack/Defense in Multiple Layers** The study on handling simultaneous attacks in multiple layers is rare. One such work is cross-layer intrusion detection [1], which examines features from multiple layers but does not consider the correlation between attacks in different layers. It will be difficult for them to capture the attacks that introduce minor misbehavior in individual layers but cause big overall damage. Not to mention the difficulty of obtaining training data for cross-layer attacks.

### 2.3 Cross-Layer Attack & Defense in Cognitive Radio Networks

#### 2.3.1 Cross-layer Attack Strategies and Defense Overview

We argue that the *coordination of attack activities in multiple layers* can (1) reduce the attacker’s probability of being detected, (2) reduce the cost to conduct the attack successfully and/or (3) achieve the attack goals that may not be feasible through attack activities in a single layer. For effective coordination, the attacker should have a clearly defined goal, which determines how the attack activities in different layers are jointly organized or even optimized. We propose a definition of cross-layer attack as

A cross-layer attack is a collection of attack activities that are conducted
coordinately in multiple network layers in order to achieve specific attack goals.

In cognitive radio networks, we identify two representative cross-layer attacks, which have not been reported in the current literature.

**Attack A1** Attackers can reduce channel utilization by

- making honest secondary users wrongly believe the existence of primary user when the primary user is absent, through PUE attack, RFSD attack, etc, in PHY layer;
- reducing the probability of honest secondary users utilizing the channel in MAC layer, through common control channel denial-of-service attack, SBW attack, etc.

These attacks have an “OR” relationship. That is, one attack alone can achieve the attack goal (i.e. reducing channel utilization). In this case, the attacker can simply use these attacks alternatively in the time domain, such that the probability of being detected in individual layers is greatly reduced.

**Attack A2** If the cognitive radio nodes near the primary user transmit data while the primary user is on, they cause interference to the primary user. To achieve this attack goal, two conditions need to be satisfied. First, the cognitive radio nodes fail to detect the existence of the primary user, which can be done through the RFSD attack. Second, they have data to transmit. This can be achieved through attacking routing protocols in the network layer. In particular, malicious nodes can route the packets toward the secondary users who are close to the primary user. Although this type of attack in the network layer has not been reported in current literature, its feasibility is obvious. In this case, the attack in physical layer and network layer have an “AND” relationship, That is, one attack activity alone cannot achieve the attack goal effectively. These attacks should be conducted simultaneously.
To address the cross-layer attacks, the defense solution should be effective such that the attackers cannot benefit from the coordination among attack activities in different layers, and compatible with the existing layered network organization. To satisfy these requirements, one effective way is to introduce “a common language” that can describe and integrate defense in different layers. We choose trust as this common language and build cross-layer trust framework.

In this paper, we will demonstrate the proposed framework in the circumstance that the attackers conduct cross-layer attack A1 with RFSD in PHY layer and SBW in MAC layer. In the rest of this section, we will introduce how to evaluate the trustworthiness of secondary users in PHY layer (Section 2.3.2) and MAC layer (Section 2.3.3), describe cross-layer attack A1 (Section 2.3.4), and present the cross-layer defense framework (Section 2.3.5).

2.3.2 PHY Layer Attack and Defense

PHY Layer Attack Model As discussed in Section II, malicious users can report false sending data to the common receive (i.e. fusion center) such that they can mislead the results of collaborative spectrum sensing. For example, an attacker can report high energy level when the actual sensed energy is low. If the fusion result by the common receiver is on (primary user is present), the attack is successful.

Before demonstrating the RFSD attack model, we review the characteristics of sensing reports from honest secondary users. Let $E_i$ denote the sensing energy for the $i^{th}$ cognitive user in each sensing period, the distribution of $E_i$ [10] is,

$$E_i \sim \begin{cases} 
\chi^2_{2m}, & H_0, \\
\chi^2_{2m}(2\gamma_i), & H_1.
\end{cases}$$

(42)

where $\chi$ stands for chi-square distribution, $m$ is time-bandwidth product, $\gamma_i$ is the received signal to noise ratio (SNR) for node $i$, and $H_0$ ($H_1$) means primary user is absent (present).
\[ f_0(x_i) = \begin{cases} U[0, 32], & \text{w.p. } 1 - p_c \\ U[0, 32] + U[0, 64], & \text{w.p. } p_c (1 - p_c) \\ U[0, 32] + U[0, 64] + U[0, 128], & \text{w.p. } p_c^2 (1 - p_c) \\ U[0, 32] + U[0, 64] + U[0, 128] + U[0, 256], & \text{w.p. } p_c^3 (1 - p_c) \\ U[0, 32] + U[0, 64] + U[0, 128] + U[0, 256] + U[0, 512], & \text{w.p. } p_c^4 (1 - p_c) \\ U[0, 32] + U[0, 64] + U[0, 128] + U[0, 256] + U[0, 512] + \sum_{i=5}^{n_c} U[0, 1024], & \text{w.p. } p_c^n (1 - p_c) \end{cases} \]

For RFSD attack, we consider the always-yes attack strategy proposed in [7]. When the malicious node senses energy level \( E_m \), it will honestly report \( E_m \) if \( E_m \geq \xi \) and dishonestly report \( E_m + \Delta \) if \( E_m < \xi \), where \( \xi \) is the attack threshold and \( \Delta \) is the bias introduced by the attacker. In this attack, malicious users report higher energy level when it estimates that the primary user is not present. Let \( G_0 \) denote honest reporting and \( G_1 \) denote dishonest reporting. The probability density function of \( E_i \) with attackers becomes,

\[
E_i \sim \begin{cases} \chi^2_{2m}, & G_0, H_0, \\ \chi^2_{2m}(\frac{2\Delta}{\sigma^2}), & G_1, H_0, \\ \chi^2_{2m}(2\gamma_i), & G_0, H_1, \\ \chi^2_{2m}(2\gamma_i), & G_1, H_1. \end{cases}
\]

(44)

where \( \sigma^2 \) is noise power.

**PHY Layer Defense Scheme** Currently there are some defense methods that address the RFSD attack [6, 7]. The scheme in [6] can only apply to hard fusion case, i.e., secondary users reporting binary detection results to the fusion center. A defense scheme that deals with soft fusion, i.e, the reporting value is sensed energy level, is proposed in [7]. In this scheme, suspicious level of each secondary user is calculated based on its reporting history. Inspired by this scheme, we develop a scheme that has lower computation complexity and is easier to fit into the cross-layer defense architecture.

The proposed scheme is composed of three steps. In the **first** step, for each node \( j \), the common receiver conducts hypothesis test 1 to detect the presence of the primary user using sensing reports from other secondary users. The Neyman-
Pearson lemma can be written as

\[
\prod_{i=1, i \neq j}^{N} \frac{P(E_i = e_i|H_1)}{P(E_i = e_i|H_0)} \xrightarrow{H_1 \sim H_0} \eta ,
\]  
(45)

where \( \eta \) is the detection threshold for hypothesis test 1 and \( N \) is the number of secondary users.

In the **second** step, hypothesis test 2 is performed to check whether a secondary user (e.g. \( j \)) is lying or not. From (44), we can see that attackers will not lie under \( H_1 \). So we only perform hypothesis test 2 when the detection result of hypothesis test 1 is \( H_0 \). Hypothesis test 2 is given by

\[
\frac{P(E_j = e_j|G_1, H_0)}{P(E_j = e_j|G_0, H_0)} \xrightarrow{G_1 \sim G_0} \zeta ,
\]  
(46)

where \( \zeta \) is the detection threshold for hypothesis test 2.

Through hypothesis test 2 we have the binary opinion about whether a node is lying or not in each sensing period. In the **third** step, if we observe that a node has reported \( r \) honest reports and \( s \) dishonest reports in the past, by the beta function trust model[12] we calculate the PHY layer trust value of the node as,

\[
\pi_1 = \frac{r + 1}{r + s + 2}.
\]  
(47)

### 2.3.3 MAC Layer Attack and Defense

**MAC Layer Attack Model** As mentioned in Section 2.2, some MAC layer protocols in cognitive radio networks are similar to IEEE 802.11 DCF protocol. When the channel becomes idle for a time equal to a distributed interframe space (DIFS), secondary users that have packets to send start to transmit. If the channel is sensed busy during the DIFS period, the nodes should defer their transmission by a random backoff time.

The random backoff time is uniformly distributed between \([0, CW]\), where \( CW \) is current contention window. For the first backoff, \( CW \) is set to \( CW_{\text{min}} \).
After each unsuccessful transmission (collision or packet lost), the value of $CW$ is doubled until it reaches $CW_{max}$. It will be reset to $CW_{min}$ after a successful transmission. For a typical IEEE 802.11 DCF protocol, $CW_{min}$ is 32 and $CW_{max}$ is 1024. In SBW attack, malicious nodes use a small $CW$ value to gain channel access priority over other nodes. Attacks can be conducted with different intensity. For example, an aggressive attacker can set $CW_{min} = CW_{max} = 2$ and a moderate attacker can set $CW_{min} = 16$, $CW_{max} = 512$. In this paper we set $CW_{min} = CW_{max} = 8$ for attackers.

**MAC Layer Defense Scheme** To defend against the SBW attack, we develop a defense method based on the scheme in [8], which checks whether the observed backoff window size distribution follows the real distribution. In [8], Kolmogoriv-Smirnov (K-S) test is used to compute the difference between distributions. However, as K-S test only considers the maximum value of the CDF difference, it is known to be sensitive near the center of the distribution. To improve the scheme in [8], we replace the K-S test by a modified Cramer-von Mises (C-M) test [13]. The proposed scheme is described as follows.

First, the backoff window size of each node is observed. For the RTS/CTS access in 802.11 DCF protocol, all nodes within the range of the observed node can have the knowledge about: the end time of last transmission $t_{i-1}$, current time of RTS packet $t_i$, and the time elapsed ($T_O$) between $t_{i-1}$ and $t_i$ when there is a collision or other nodes are transmitting. Then the backoff window size can be calculated as [8],

$$x_i = \frac{t_i - t_{i-1} - T_{DIFS} - T_O}{\delta},$$

(48)

where is $T_{DIFS}$ is the length of DIFS frame and $\delta$ is the time unit of backoff window. With the observed backoff window size, we can obtain the empirical distribution of backoff window size and its cumulative distribution $F_1$. 

For a typical IEEE 802.11 DCF with \( CW_{\text{min}} = 32 \) and \( CW_{\text{max}} = 1024 \), the backoff window size distribution of normal nodes \( f_0 \) is given in Eq. (43) [8], where \( U \) is uniform distribution and \( p_c \) is collision probability which can be estimated by the observation of successful transmission count and collision count. Then we can obtain the CDF \( F_0 \).

If we have \( K \) observations \( x_1, ..., x_K \) and \( L \) sample data \( y_1, ..., y_L \) generated from real distribution, we can conduct the C-M test as,

\[
\theta = \frac{KL}{(K + L)^2} \left( \sum_{i=1}^{K} [F_0(x_i) - F_1(x_i)]^2 + \sum_{j=1}^{L} [F_0(y_j) - F_1(y_j)]^2 \right).
\tag{49}
\]

Note that in (49), it only measures the absolute difference between the two distributions, which cannot distinguish whether the observed CDF \( F_1 \) is above the real CDF \( F_0 \) or under \( F_0 \). For the case that \( F_1 \) is mostly under \( F_0 \), which means the observed backoff window size is greater than normal size, the node should not be classified as misbehaving. To take this case into consideration, we modify the C-M test to,

\[
\theta = \frac{KL}{(K + L)^2} \left( \sum_{i=1}^{K} \text{sgn}(F_0(x_i) - F_1(x_i))[F_0(x_i) - F_1(x_i)]^2 + \sum_{j=1}^{L} \text{sgn}(F_0(y_j) - F_1(y_j))[F_0(y_j) - F_1(y_j)]^2 \right),
\tag{50}
\]

where \( \text{sgn}(x) \) is sign function.

Define \( D = \max\{\theta, 0\} \), we calculate trust value of MAC layer as,

\[
\pi_2 = e^{-D^2}.
\tag{51}
\]

When \( \theta \) is negative, (i.e. the area of \( F_1 \) is mostly under that of \( F_0 \)), \( D \) is 0, and trust value \( \pi_2 \) is 1. It indicates that the node is completely trusted. Otherwise, when \( \theta \) is positive (i.e. the backoff window size is smaller than normal value), \( \pi_2 \)
decreases as the distribution difference increases. If $\pi_2$ is below threshold $\lambda_2$, the node is detected as a malicious node.

### 2.3.4 Cross-Layer Attack

In this paper, we investigate the cross-layer attack strategy $A_1$ described in Section 2.3.1. In particular, malicious users choose to conduct RFSD attack with probability $P_1$ in each reporting round in the PHY layer, and conduct SBW attack with probability $P_2$ after a successful transmission or collision in the MAC layer. Here, $P_1$ and $P_2$ are called *attack probability* in PHY and MAC layer, respectively.

Because of the defense schemes described in Section 2.3.2 and 2.3.3, the aggressive attackers, who attack all the time, can be easily detected. Smart attackers should behave well and badly alternatively, and carefully choose the attack probabilities. As we will demonstrate in Section 4.4, there exists optimal attack probabilities such that attackers can cause maximum damage without being detected. The performance of the single layer defense, cross-layer attack, and cross-layer defense will be evaluated in the worst- or best-case scenario in which the attackers choose the optimal attack probabilities.

### 2.3.5 Cross-Layer Defense

To address the cross-layer attack in wireless networks, we propose a cross-layer defense framework shown in Fig. 17. The framework has the following components:
- Single Layer Monitoring & Trust Calculation (SLMTC): collecting observations from network protocols, and evaluating \textit{in-layer trust values} (e.g. $\pi_1$ and $\pi_2$).

- Trust Fusion: taking in-layer trust values as inputs and calculating an overall trust value $T$ for each node.

- Abnormal Detection: identifying malicious nodes based on the overall trust values and how they change with time.

The SLMTC schemes have been discussed in Section III.B and III.C. In this subsection, we will describe trust fusion and abnormal detection, which are jointly referred to as the cross-layer trust manager (CTM). Trust fusion can be modeled as a \textit{multipath trust propagation} problem shown in Fig. 18, in which PHY layer trusts node $i$ with level $\pi_1$, MAC layer trusts node $i$ with level $\pi_2$, and the cross-layer trust manager (CTM) trusts the PHY layer results with level $w_1$ and the MAC layer results with level $w_2$. The goal is to determine how much CTM should trust node $i$, which is just the total trust value $T$.

For the multipath combining, we adopt the method in [14],

$$T = w_1 \pi_1 + w_2 \pi_2,$$

where $w_1$ and $w_2$ are the weights, and $\pi_1$ and $\pi_2$ are in-layer trust values calculated in Section 2.3.2 and 2.3.3.
We argue that $w_1$ and $w_2$ should describe the effectiveness of in-layer trust values in terms of differentiating good nodes and bad nodes. When the variance of honest nodes' in-layer trust values is large, it is more difficult to separate the malicious nodes and the honest nodes with low trust. Denote by $v_1$ the variance of $\pi_1$ and $v_2$ the variance of $\pi_2$. We define

$$w_1 = \frac{v_2}{v_1 + v_2}, \quad w_2 = \frac{v_1}{v_1 + v_2}. \quad (53)$$

We use a heuristic method to calculate $v_j$, $j = 1, 2$. Denote $v_j^i$ the variance of the sequence $\{\pi_j^i(1) \ldots \pi_j^i(M)\}$, where $i$ is node ID, $j$ is layer ID, and $M$ is current detection round. Let $v_j^m$ be the median of $\{v_j^i, 1 \leq i \leq N\}$. The variance value $v_j$ is calculated as

$$v_j = \frac{1}{C_j} \sum_{i, \forall 1 \leq i \leq N \; \& \; \frac{v_j^i}{v_j^m} \leq \rho} v_j^i, \quad (54)$$

where $\rho$ is a threshold to filter out variance values that are far away from normal variance values and $C_j$ is the number of nodes whose trust variance satisfies $\frac{v_j^i}{v_j^m} \leq \rho$. The reason for adding this filter is to make sure that the calculation in (54) is based on trust variance from honest nodes.
With Eq. (52) - (54), we can obtain an overall trust value $T$ for each node. If the overall trust value of a node is below threshold $\lambda_3$, the node is detected as malicious. Then actions can be taken in the PHY layer and MAC layer such that it cannot cause further damage.

### 2.4 Simulation Results

#### 2.4.1 Simulation Setup

We consider a cognitive radio network of $N (=10)$ secondary users and two of them are malicious. We assume the attackers do not collude.

For PHY layer simulation, the time-bandwidth product $m$ is 5. Primary transmission power is 200mw and noise level is -110dBm. Path loss factor is 3. For the RFSD attack, the attack threshold $\xi$ is 15 and attack bias $\Delta 15$. The threshold in hypothesis test 1 is set to $\eta = 1$ and that in hypothesis test 2 is set to $\zeta = 1.6$.

In MAC layer, for honest users, the minimum backoff window is 32 and maximum is 1024. Attackers use a fixed backoff window with size 8. The length of observed data $K$ is 15 and that of sample data $L$ is 1000. We simulate the scenario that the secondary users are saturating, i.e., they always have packets to send. For the case that some cognitive radio nodes are not saturating, for example, the...
nodes will wait a certain time for new packets after a successful transmission, their observed backoff window CDF $F_1$ will be absolutely under $F_0$, then they will still be classified as legitimate terminals.

For fair comparison, in all defense schemes, the trust value threshold is set such that the false alarm rate $P_f$ of malicious user detection is 0.001. Specifically, for single layer attack/defense in Section 2.4.2, the PHY layer trust threshold $\lambda_1$ is set to 0.8 and MAC layer trust threshold $\lambda_2$ is set to 0.4. For cross-layer attack/defense in Section 2.4.4, trust threshold $\lambda_3$ is set to 0.75. For cross-layer
attack single layer defense in Section 2.4.3, \( \lambda_1 \) is 0.79 and \( \lambda_2 \) is 0.36. The threshold \( \rho \) in Section 2.3.5 is 10.

### 2.4.2 Results for Single Layer Attack & Defense

We study the attack/defense in PHY layer as described in Section 2.3.2. The results are shown in Fig. 19. In this figure, the x-axis is attack probability and y-axis is channel availability defined as the probability that the primary user detection result is idle when real channel status is idle. When attack probability is 0 (no attacker), channel availability is near 1 (considering possible false alarm made by honest nodes).

Fig. 20 shows the results for MAC layer attack/defense discussed in Section 2.3.3. The x-axis is attack probability and y-axis is average transmission probability of honest user. Note that the sum of transmission probability of all users is less than 1 due to the existence of collision.

Three curves are compared for single layer attack/defense: (1) \( N(=10) \) secondary users, two of them are attackers, no defense; (2) \( N(=10) \) secondary users, two of them are attackers, single layer defense; (3) \( N(=8) \) secondary users, no attackers. From these two figures we make following observation,

- When there is no defense, the channel availability in PHY layer (or transmission probability in MAC layer) decreases significantly as the attack probability increases.

- The performance in PHY layer is more sensitive to the increasing of attack probability than that in MAC layer.

- With PHY layer defense, the channel availability decreases first and then starts to increase and finally converges to the case that all remaining nodes are honest. As shown in Table 1, the optimal attack probability is 0.15 and
the maximum performance degradation is 11.73%.

- Similarly, with MAC layer defense, the optimal attack probability is 0.6 and the maximum performance degradation is 5.84%.

<table>
<thead>
<tr>
<th>Attack/Defense</th>
<th>Damage</th>
<th>Optimal $P_1$</th>
<th>Optimal $P_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHY</td>
<td>11.73%</td>
<td>0.15</td>
<td>N/A</td>
</tr>
<tr>
<td>MAC</td>
<td>5.84%</td>
<td>N/A</td>
<td>0.6</td>
</tr>
<tr>
<td>CASD</td>
<td>17.40%</td>
<td>0.15</td>
<td>0.6</td>
</tr>
<tr>
<td>CACD</td>
<td>9.73%</td>
<td>0</td>
<td>0.65</td>
</tr>
</tbody>
</table>

2.4.3 Cross-Layer Attack v.s. Single Layer Defense (CASD)

The results for cross-layer attack v.s. single layer defense is shown in Fig. 21. Here, the attack goes cross-layer but the defense is done independently in two layers. In this figure, the x-axis is PHY layer attack probability, y-axis is MAC layer attack probability, and z-axis is average transmission probability of honest users.

An interesting observation is that the best performance happens on the four corners of the figure, which indicates either no attack or always attack. In CASD, when a user is classified as malicious in one layer, the detection result is sent to other layers such that the malicious user cannot cause further damage in other layers. In this case, the optimal attack probabilities are $P_1 = 0.15$ and $P_2 = 0.6$, which means conducting the optimal attack in two layers simultaneously. The largest damage attackers can cause is 17.4%. That is, malicious users can bring down the performance by 17.4% without being detected.

2.4.4 Cross-Layer Attack v.s. Cross-Layer Defense (CACD)

Fig. 22 shows the average transmission probability of honest users when the proposed cross-layer defense is used to handle the cross-layer attack. The x-axis is
and the y-axis is $P_2$. Compared with the results of CASD (see Fig. 21), we can see that the average transmission rate achieves its best performance (i.e. as if there are no attackers) in much more area. In terms of the worst-case performance, the maximum performance degradation ratio is 9.73%, when the attackers choose $P_1 = 0$ and $P_2 = 0.65$. The worst-case performance degradation of CACD is only 56% of that of CASD. Obviously, the proposed cross-layer defense scheme can effectively handle cross-layer attacks.

The worst-case performances of all tested schemes are summarized in Table 1. It is clearly seen that cross-layer attack is stronger than attacks in single layers when there is no coordination among defense schemes in different layers. Furthermore, the proposed cross-layer trust framework can defeat cross-layer attack effectively.

### 2.5 Conclusion

When the attackers start to coordinate their actions in different layers, their attack strength is enhanced even if there are abnormal detection mechanisms in each layer. In this paper, we raised the concern about cross-layer attacks and utilized cognitive radio networks as the platform to demonstrate such attacks. A specific attack, time domain coordination of RFSD and SBW, was studied in details. More importantly, we designed a cross-layer trust defense scheme by developing (1) abnormal detection schemes in PHY and MAC layers and (2) the cross-layer trust manager. The proposed defense demonstrated excellent performance against this cross-layer attack. Finally, we would like to point out that the concept of cross-layer attack and the proposed defense framework can be applied to other attacks in cognitive radio networks, and to other types of wireless networks.
List of References


In submission to IEEE Transactions on Information Forensics and Security, 2011. The earlier conference paper was accepted by IEEE Global Communication Conference in July 2011.

Topology based Modeling of Cascading Failures in Power Systems - Part I: Foundations, Models, and Assessment Metrics

Wenkai Wang, Haibo He, and Yan Sun

Department of Electrical, Computer, and Biomedical Engineering
University of Rhode Island, Kingston, RI 02881

Corresponding Author: Yan Sun
Electrical, Computer, and Biomedical Engineering
University of Rhode Island
4 East Alumni Ave.
Kingston, RI 02881, USA
Phone:+1-401-874-5803
Email address:yansun@ele.uri.edu
Abstract

With the continuous growing energy demand and environmental concerns, it has recently attracted significant attention of academia, industry, and governments in the development of a smart electric power grid to provide affordable, reliable, efficient, and secure supply of electricity. Among many enabling technologies toward such a smart grid, security has been widely identified as one of the key components for such a complex system. In Part I of this two-part series study, we present a comprehensive analysis of the foundations, system models, and assessment metrics for power system cascading failures. Our objective in this part is to understand the limitations of traditional largest-load based attack strategies and provide critical insights to understand the power grid behavior subject to complex attacks. This part will layout important foundations to study the topology-based modeling of power system cascading failures for Part II, which discusses two representative models, the non-recoverable model and recoverable model in detail.

3.1 Introduction

The power grid is one of the largest interconnected systems on earth. Taking the U.S. power grid as an example, it has more than 9,200 electric generating units, 1,000,000 megawatts of generating capacity, and 300,000 miles of transmission lines [1]. Such a grid has been progressively developed for over a century, and now is facing many issues such as aging, inefficiency, congestion, and incapability of meeting the future energy and security needs of the society [2]. Furthermore, the driving forces toward “green energy” for environmental concerns have presented ever
greater challenges to the power and energy community [3, 4, 5]. The development of such a future power grid means we are facing an extremely large-scale interconnected network with different types of heterogenous energy resources, which could suffer from physical failures and cyber failures caused by unintentional errors or malicious attacks. In fact, according to the DOE definition [1], “security” has been identified as one of the key aspects of the smart grid, together with seven other characteristics including “intelligent, efficient, accommodating, motivating, opportunistic, quality-focused, and environment friendly.”

In 2003 there was a major blackout in North America. It started in Ohio, then spread to large portions of the Midwest and Northeast United States, including metropolitan areas such as New York City, and finally affected Ontario in Canada, with a total impacted area of about 50 million people and 61,800 megawatts (MW) of electric load [6]. Two recent stories appeared in major national medias also attracted many discussions in academia, industry, and government agencies. The first one is a discussion regarding cyber-attack penetration to the U.S. electrical power grid at different levels (reported by Wall Street Journal [7]), and the second one is an academic research of the U.S. power grid under cascade-based attacks (see New York Times [8]). The phenomena of this kind of failure propagation is called cascading failure in the complex network literatures [9, 10, 11, 12]. Generally speaking, in complex networks like the power grid, if one component (e.g., substation or transmission line) fails, it may trigger the failure of other components. The failure will propagate further and eventually cause a large scale system malfunction. Although there are extensive efforts and large initiatives across the world to investigate different aspects of smart grid technology, such as the Grid 2030 of the U.S. Department of Energy (DOE) [2], the European Technology Platform (ETP) SmartGrids 2020 [13], the IEEE Smart Grid Initiative [14], to name a
few, limited understanding has been achieved for the power grid security given the complexity in terms of network topology, modeling and simulation, visualization, and intrinsic fundamental principles governing power flow in such a network [4, 5].

The catastrophic consequence of cascading failure in power grid demonstrates the importance of understanding this extremely complicated behavior. In this paper we specifically focus on how complex attacks rather than random failures can cause cascading failures in the power grid. In particular, we aim to find appropriate models and metrics to provide important foundations in understanding the power grid attacks (Part I), and then further develop specific attack strategies and analyze grid behavior based on the Western North American power grid benchmark (Part II).

To study cascading failures, it is important to model the behaviors of network upon node failures. This is described by the load redistribution models, which specify how the load carried by the failed node will be redistributed to other remaining nodes. In the current literature, researchers often use two types of load redistribution model. The first type is non-recoverable model [15, 16], in which load of failed node is redistributed to its one-hop neighbors. The second type is recoverable model [11, 17, 18, 19]. In this model, when a node is attacked, the paths that previously traverse this node may have to be changed. The load of failed node is redistributed to nodes on the new paths. Due to the load redistribution, the remaining nodes in the network may be overloaded. In the first model, the overloaded nodes cannot function anymore, which causes more load redistribution. In the second model, overloaded nodes are still functional but their efficiency is reduced.

The rest of the paper is organized as follows. Section 3.2 introduces the network model, attack model, and load redistribution models in power grid. Limi-
lations of traditional load based attack strategy is analyzed in Section 3.3. Afterwards, existing and new assessment metrics to under power grid cascading failures under both non-recoverable model and recoverable model are discussed in Section 4.2. Finally, a conclusion is given in Section 4.5.

3.2 System Models

The typical power grid operations include power generation, transmission, and distribution. In principle the power produced by generators must be equal to the power consumed by users plus the loss during transmission at any time. As power grid usually involves multiple service providers and the user energy demand is constantly changing, the system model based on power flow analysis is very complicated especially when network size is large. For this reason, in this paper we study the cascading failures in power grid using topology based models. We first describe network model in Section 3.2.1, then state the attack model in Section 3.2.2, followed by two load redistribution models in Section 3.2.3.

3.2.1 Network Model

In topology based models, the power grid network is viewed as a graph with substations being its the nodes and transmission lines being its edges. In general, there are three types of substations: generators, transmission substations, and distribution substations. Every substation has to carry certain load when electricity flows through it.

In the current literature [11, 15, 17, 20], there are two ways to measure the load. The first one is the degree model [15, 20]. In this model, load is calculated from the degree of each node and its neighboring nodes. Denote \( k_i \) the degree of node \( i \), the initial load \( L_i \) of node \( i \) is computed as,

\[
L_i = [k_i (\sum_{m \in \Gamma_i} k_m)]^\alpha,
\]

(55)
where $\Gamma_i$ is the set of neighboring nodes of node $i$ and $\alpha$ is a tunable parameter that can control the distribution of initial load.

The second one is *betweenness* model [11, 17]. When power is delivered from generators to distribution substations, it has to pass through several substations. For each generator and each distribution substation, there is a routing path between them. In this model, load is evaluated as the betweenness of a node, which is the number of paths that traverse this node. In the betweenness model, the load of a node is just the betweenness of this node. For a node that there is no path passing through, its initial load is set to 1.

For each node, *capacity* is defined as the maximum amount of load that the node can tolerate. It is generally assumed that the capacity of a node, $C_i$, is proportional to its initial load $L_i$ [15, 17, 20],

$$C_i = T \times L_i,$$

where $T$ ($T \geq 1$) is *system tolerance*. Higher $T$ means higher capacity of each node, and higher system tolerance. Generally speaking, the cascading failures are less likely to occur when the system tolerance is high.

### 3.2.2 Attack Model

The power grid networks are vulnerable to different types of attacks. For instance, **cyber attack** [21, 22, 23, 24] which is common in traditional wireless networks or computer networks can also be exploited in power grid networks. Attackers can launch cyber attack in the power grid against protection (e.g., generator protection, line protection, etc), supervisory control and data acquisition (SCADA), electricity management system, and others [25, 26, 27, 28, 29]. In 2001 the California Independent System Operator that manages the electricity supply of California was penetrated by cyber attacks, which “got close” to disrupt the
power flow according to the Los Angeles Times [30].

There are many other wireless networks existing in power grid besides the SCADA system. For example, in the newly emerged smart grid networks, advanced metering infrastructure (AMI), which is believed to have the capability to save costs for both energy suppliers and consumers, is largely deployed. In some implementations the AMI nodes are connected to form multi-hop wireless mesh networks that provide reliable multi-path communications. These wireless networks, mostly composed by low cost and low power-usage devices, are vulnerable to cyber attacks that occur to most current wireless networks [31].

Besides cyber attacks, another important type of attack is physical attack. In physical attack, the targets are the physical components of power grid such as the substations and transmission lines. Although nature forces such as storms and hurricanes are still the most common causes of physical failures nowadays, deliberate attacks on the power grid which are more threatening has been an increasing concern of the power grid studies. Attackers such as criminals, terrorists, or even foreign army forces can destruct transmission lines, power plants, substations and other physical structures to cause large scale power failure.

In this paper, we mainly study attacks aiming to knock down substations. Such attacks can be either physical or cyber. Attacker can knock down one or multiple substations, which are referred to as victim nodes. It is obvious that the damage of taking down different substations can be very different. Some may have very limited impact, but some may cause large scale system blackout. In this paper, an attack strategy refer to a specific way for selecting victim nodes, whose failure can cause severe cascading failures. On the other hand, from the defense point of view, the victim nodes discovered by the attack strategies are the weakest components of the power grid. Studying attack strategies can reveal
system vulnerabilities and help operators to protect the weak components.

3.2.3 Load Redistribution Models

When a substation is taken down due to either physical or cyber attacks, the load carried by this substation has to be redistributed to other nodes. A node, which receives redistributed load from the failed node, may be overloaded when the summation of its original load and the redistributed load exceeds its capacity. In the current literature, there are mainly two types of load redistribution models that describe the behaviors of the overloaded nodes [11, 15, 17, 20] .

In the first type [15, 20], when a node is overloaded, it completely fails to operate. That is, it can neither carry its original load nor the extra load from other failed nodes. This type of model is referred to as the non-recovery model. In particular, we will study the non-recovery model in 3.2.1 that has the following features. The load is defined from the degrees of network nodes as in equation (55). When node $i$ is attacked or overloaded, its load is redistributed to its one-hop neighbors. The extra load redistributed to each neighbor is proportional to the initial load of the neighbor. Specifically, the additional load redistributed to neighbor node $j$, denoted by $\Delta_{ij}$, is

$$
\Delta_{ij} = L_i \frac{L_j}{\sum_{m \in \Gamma_i} L_m} 
$$

(57)

In this model when a node is attacked, its one-hop neighbors are affected first. Then if a neighbor node, say node $j$, is overloaded, node $j$’s initial load plus the load received from node $i$ has to be redistributed to node $j$’s neighbors excluding node $i$. Node $j$’s neighbors, which receive extra load from node $j$ and possibly from other overloaded nodes, may be overloaded too. We see that node failures can propagate until no new nodes are overloaded. This model is simple since it assumes that load is only redistributed to one-hop neighbor and overloaded nodes.
do not function at all.

In this paper we also investigate a more complicated model \cite{11_17}, in which the load is the betweenness defined in Section 3.2.1. When node \( i \) is taken down, all paths that previously traverse node \( i \) have to be replaced by new paths that do not go through node \( i \). This will increase the load of all nodes along the new paths. Some of the nodes on the new paths may be overloaded due to the load increase. When a node is overloaded, the model assumes that the node is still in function but its capability in power transmission is degraded. Specifically, an initial efficiency value is associated with every edge in the power grid. When a node is overloaded, the efficiency values of all edges that the node is connected to decrease. This model is also referred to as the \textbf{recoverable model}.

When the efficiency values of edges change due to overloading, the efficiency values of paths, computed as the harmonic composition of the efficiency value of all edges along the path, may change too. Then the previously most efficient path between a generator and a distribution substation may not be most efficient any more. In other words, the paths between generators and distribution substations are changed, which result in load/betweenness change of nodes. This load change could cause more overloading, which results in change in edge efficiency, new path selection, and load change of nodes. This recursive process continues until all paths (i.e. the most efficient paths between each generator and each distribution substation) are stable.

As a summary, the main difference between the two models is the way cascading failures propagate. In the first model, one-hop neighbors of the victim nodes are affected first, then two-hop neighbors and so on. In the second model, all nodes on the new paths are affected. These affected nodes might be far away from the victim nodes. The two models make different assumptions and may apply to dif-
different scenarios. In this paper we first study the non-recoverable model in Section 4.2, then investigate the recoverable model in Section 4.3.

3.3 Limitations of Traditional Largest-Load Based Attack Strategy

The goal of attack is to maximize the severity of cascading failures given specific attack resources. In this work, we assume that the attackers know the entire network topology. As discussed earlier, the attack strategy refers to the method of victim nodes selection. In traditional attack strategy [11, 15, 17], the attacker always chooses the victim nodes according to load. For single-node attack, it will choose the node with the largest load as victim node. For multi-node attack, it will choose the top $K$ largest load nodes as victim nodes if it has the capability of taking down $K$ nodes.

This strategy is based on the fact that the failure of nodes with large load leads to a large redistributed load to other nodes. Then it is highly possible that some nodes are overloaded because of the large redistributed load. This type of attack strategies have been widely used when studying vulnerabilities of the power grid and defense solutions. However, can it represent the strongest attack?

In our work, we have discovered that the traditional attack strategy has a severe limitation. That is, it hardly considers about power grid topology while the topology plays a critical role in cascading failures.

For the non-recoverable model, the load of victim node will be redistributed to its one-hop neighbors. Note that although some nodes carry high load, they probably have a lot of neighbors to share their load if they are taken down. On the other hand, a node with low load may have few neighbors to share load. The initial load and the number of neighbors all play important roles in the cascading failure. However, the traditional load based attack strategy only considers the load
Figure 23. Load redistribution example under non-recoverable model. The notation “X: y + z” means node X has initial load y, and receives additional load z after the victim node is taken down.

but neglects the number of neighbors.

To illustrate the above discussion, we show two representative nodes from the Western North America power grid in Fig. 23. Node A, the center node in the left plot, carries a high load 1414. When A is taken down, its load is redistributed to its 14 neighbors according to equation (57). For instance, node C will carry 101.5 extra load. Since C’s initial load is 486, the extra load is about $101.5/486 = 20.8\%$ of its initial load. As long as the network tolerance ($T$) is greater than 1.208, node C will not be overloaded. For all of the A’s neighbors, the ratio of the additional load over the initial load of a neighbor is less than 30%. It means that if the system tolerance value $T$ is greater than 1.3, there will be no cascading failure when node A is taken down. On the contrary, node O, shown in the right plot, only carries 20 initial load. When node O is down, the additional load redistributed to its neighbor P is 11.42, which is 71% of node P’s initial load. Thus, cascading failure will occur as long as $T$ is less than 1.71. In this example, attacking the node with lower load (i.e. node O) is actually more effective than attacking the node with higher load (i.e. node A).
In the recoverable model, we have found that the load based strategy is very reasonable when there is only one victim node. The details can be seen in the detailed simulation study and analysis in the Part II of this series. However, for multiple-node attack, the load based strategy is not effective.

**Fig. 24** illustrates an example. Suppose node B and C are generators, node K and L are distribution substations. In the recoverable model, it assumes that a path has to be established between every generator and distribution substation pair. Initially when all nodes are operating normally, this path will be the shortest path between each pair. For example, the path between node B and K would be $B \rightarrow A \rightarrow G \rightarrow K$. Recall that in the recoverable model, load is betweenness defined as the number of paths that traverse a node. The load of node A and G is 4, since there are 4 shortest paths going through them (between A and K, between B and K, between A and L, and between B and L). For other nodes, there are no paths passing through them, and their initial load is 1.

According to the load based strategy, if attackers can take down two nodes, node A and G will be selected as victim nodes. When node A and G are taken down, the edges that they are connected to are down too. In this case, system will try to find alternative routes between generator and distribution substation...
pairs. For example, the new path between B and K would be $B \rightarrow O \rightarrow N \rightarrow M \rightarrow K$. However, if the attacker takes down A and O, there will be no alternative path between B and K. The scenario is similar for other generator and distribution substation pairs. Obviously, taking down nodes A and O, which is more harmful than taking down A and G, is a stronger attack strategy than the load based strategy.

The above analysis clearly demonstrates the existence of attack strategies that are stronger than the traditional load based strategies under both load redistribution models.

### 3.4 Assessment Metrics

In this section, we present in detail of the existing and proposed new assessment metrics for power system cascading failures under both non-recoverable models and recoverable models, which will provide the criteria for performance evaluation discussed in Part II of this series.

#### 3.4.1 Assessment Metrics for Non-Recoverable Model

**Percentage-of-failure:** In the non-recoverable model some nodes fail to operate after the cascading failure. We can use the percentage-of-failure, referred as $\lambda$, as a metric to evaluate the damage,

$$\lambda = 1 - \frac{N'}{N},$$

where $N$ is the number of nodes before the cascading failure and $N'$ is the number of survived nodes after the cascading failure.

Obviously, $\lambda$ depends on system tolerance and the attack strategy. Given a fixed system tolerance value $T$, we can run simulation to obtain $\lambda(T, i)$ which represents the percentage-of-failure when the system tolerance is $T$ and node $i$ is
chosen as the victim node. The attacker can use $\lambda(T, i)$ as a metric to choose the victim nodes. That is,

\begin{center}
\textit{Strategy $\lambda$: selecting the victim node as the node with the largest $\lambda(T, i)$ value.}
\end{center}

We conduct simulation using Western North American power grid data. Given a specific system tolerance value $T$, through simulation, we obtain the percentage-of-failure values for all nodes. In Fig. 25, the sorted percentage-of-failure for all nodes under three fixed system tolerance value ($T = 1.1, 1.3,$ and $1.5$) is shown. We can observe that the strategy $\lambda$ has several limitations.

- First, $\lambda$ depends on $T$. Different $T$ will yield different $\lambda$ for each node.

- Second, when the $T$ value is fixed, the nodes can be roughly divided into two groups. In one group, the nodes have $\lambda = 0.002$. That is, when the victim node is chosen from this group, there is no cascading failure. No other nodes are taken down except the victim node itself. In the other group, $\lambda$ is close to 1. That is, when the victim node is chosen from this
group, most of the network nodes will be overloaded and the network will experience severe cascading failure. This is observed in Fig. 25, and also clearly demonstrated in Fig. 26 a) that shows the histogram of $\lambda$ for $T = 1.3$.

From this observation, we see that the $\lambda$ values cannot well address the difference among the nodes within each group.

- Third, the percentage-of-failure based strategy requires obtaining $\lambda$ for every node. This computation cost can be large when the network size is large. Due to these limitations, percentage-of-failure (i.e. $\lambda$) is not a good metric for the attacker to select victim nodes.

**Required Redundancy (RED)**  
To overcome the first and second limitations of the percentage-of-failure metric, we develop a new metric called Required Redundancy (RED). The definition of RED value of node $i$ is the *minimal required system tolerance value such that cascading failure does not occur when node $i$ is taken down.*
Using Fig. 27, we illustrate how the RED values are obtained. Fig. 27 shows the percentage-of-failure as a function of the system tolerance value when node 1 and 100 is taken down, respectively. One curve in the figure is $\lambda(T, 1)$ v.s. $T$ and the other is $\lambda(T, 100)$ v.s. $T$. For node 1, we see the sharp drop in the percentage-of-failure value when $T$ is around 1.25. That is, if node 1 is chosen as the victim node, the cascading failure will not occur as long as $T$ is greater than 1.25. Recall that $\lambda = 0.002$ means all nodes function well except the victim node. Thus the RED value of node 1 is 1.25. Similarly, the RED value of node 100 is 1.125.

By simulation we obtain the RED values of all nodes. We have three observations. First, the RED metric overcomes the first limitation of the percentage-of-failure metric since RED does not depend on $T$. Second, the RED metric overcomes the second limitation of the percentage-of-failure metric because RED has a much better distribution. The histogram of RED is shown in Fig. 26 b). It is seen that nodes can have very different RED value. It is possible for the RED metric to describe the difference among network nodes. Third, there is no close form formula for RED. To obtain RED for a node, we have to run simulations for
different $T$ values assuming this node is the victim node. The computation cost is huge for a large network. Therefore, RED still has the third limitation (i.e. high computation cost).

**Risk if Failure (RIF)** To overcome the limitation of RED, we propose another new metric called Risk if Failure (RIF). The computation of RIF only requires local network information and it can replace RED.

We define $RIF_i$, the RIF value of node $i$, as the ratio between the initial load of node $i$ and the total initial load of $i$’s neighbors,

$$RIF_i = \frac{L_i}{\sum_{m \in \Gamma_i} L_m}. \quad (59)$$

From equation (59), we see that $RIF_i$ only depends on local information: the initial load of node $i$ and its neighbors. Compared with RED, the computation of RIF is extremely simple.

Next, we discuss why RIF can be used to replace RED in selecting victim node. By equation (59) we calculate the RIF values of all nodes in the Western North American power grid network. The histogram of RIF in shown in Fig. 26 c). We can see that Fig. 26 c) is very similar to Fig. 26 b), which indicates that there may be a strong correlation between RED and RIF.

Fig. 28 demonstrates such a correlation. The x-axis is RIF and the y-axis is RED. Each node is represented by a blue dot in Fig. 28. The position of the dot is determined by this node's RED and RIF values. All of these blue dots, describing the RED values of all nodes versus their corresponding RIF values, are referred to as the original data set.

Besides the original data set, we also construct the worst-case and best-case data set. If several nodes have the same RIF value, we find the node that has the largest (or smallest) RED value, and put the corresponding RIF and RED pair into the worst-case (or best-case) data set.
In Fig. 28 there are three lines. The upper line is the least square fitting for the worst-case data set. The middle one is the least square fitting for the original data set. The lower curve is the least square fitting for the best-data set. From Fig. 28 we made two observations,

- RED and RIF have a strong linear relationship in the original data set and the worst-case data set.
- If we know the RIF value of a node, we can use it to estimate the upper bound of the node's RED value. With a high probability, the estimated upper bound is close to the true RED value.

In conclusion, we can use RIF, which is very easy to compute, to replace RED as a metric to find critical node in terms of causing cascading failure. The higher the $RIF_i$ value is, the more likely the cascading failure occurs if node $i$ is taken down. In other words, $RIF_i$ roughly describes the risk of the network if node $i$ is the victim node. The RIF metric can be very useful for attackers to design attack strategies.
**Initial Load as a Metric**  In Section 3.3 we present some discussion on the traditional load based attack strategy. In this subsection, we examine whether load can be a good metric to select victim node based on real data.

The histogram of initial load for all nodes is shown in Fig. 29 a). We can see that most nodes have very low initial load, while a few nodes have high initial load. Fig. 29 b) shows the RED value of each node versus its initial load. If we view this plot from right to left, we can see that although some nodes have very high initial load, their RED values are low. It means that even if these high load nodes are attacked, the system is still relatively safe. On the contrary, some nodes have low initial load but their RED values are high. These observations verify our discussion regarding the load based attack strategy in Section 3.3.

In conclusion, we have investigated four metrics in describing the critical level of a node in terms of causing cascading failures.

- Initial load: used in the traditional attack strategy and is not effective;
- Percentage-of-failure ($\lambda$): a direct measure but has three limitations;
- Required redundancy (RED): a good metric but has high computation cost;
- Risk if failure (RIF): a good metric and only needs simple computation.

3.4.2 Assessment Metrics for Recoverable Model

In the recoverable model, since overloaded nodes are still in function, we cannot use the percentage-of-failure or Required Redundancy (RED) metrics defined in Section 3.4.1. Since the load of failed nodes is redistributed to a broader range of nodes instead of just one-hop neighbors, the Risk if Failure (RIF) metric, which is computed from local information, is not a suitable metric either.

**Edge Efficiency:** To measure the damage of cascading failures under recoverable model, we adopt a metric called network efficiency as in [17, 19]. Each edge in the power grid network is associated with a value called edge efficiency. The edge efficiency value ranges from 0 to 1. Higher edge efficiency value means better power delivery performance of the edge. We denote $e_{ij}$ the edge efficiency of node $i$ and node $j$. Initially, if there is a direct link between node $i$ and $j$, $e_{ij}$ is 1. Otherwise $e_{i,j}$ is 0.

When a node is overloaded, the efficiency values of all edges that the node is connected to decrease. In particular, if node $i$ is overloaded at time $t + 1$, the edge efficiency value $e_{ij}$ is changed as follows [17].

$$e_{ij}(t + 1) = \begin{cases} 
e_{ij}(0) \frac{C_i}{L_i(t)} & \text{if } L_i(t) > C_i \\ e_{ij}(0) & \text{if } L_i(t) \leq C_i \end{cases},$$

where $e_{ij}(t)$ is the efficiency of link between node $i$ and $j$ at time $t$ and $L_i(t)$ is load of node $i$ at time $t$. Compared to the initial efficiency $e_{ij}(0)$, the efficiency $e_{ij}(t + 1)$ decreases proportionally to the overload extent $L_i(t)/C_i$.

**Path Efficiency:** With the definition of edge efficiency, we can define the efficiency value of a path according to the description in [17]. The path efficiency
value $\varepsilon_{ij}$ between two nodes $i$ and $j$ is defined as the harmonic composition of the efficiency value of all edges along the path,

$$
\varepsilon_{ij} = \frac{1}{\sum_{k=1}^{\tau} 1/x_k},
$$

(61)

where $\tau$ is the number of edges on the path and $x_k$ is the efficiency value of each edge on the path. For example, if there are three edges with efficiency value $x_1 = x_2 = x_3 = 0.5$ on a path between $i$ and $j$, then the efficiency value of the path is $[\sum_{k=1}^{3} 1/x_k]^{-1} = 1/6$. When there are multiple paths between node $i$ and $j$, the path with largest efficiency value will be chosen as the power transmission route.

**Network Efficiency:** Suppose there are $N_g$ generators and $N_d$ distribution substations in a power grid network. The average network efficiency of the power grid is defined as [17],

$$
E = \frac{1}{N_g N_d} \sum_{i \in G} \sum_{j \in D} \varepsilon_{ij},
$$

(62)

where $G$ and $D$ are the generators set and distribution substations set, respectively. Here, $\varepsilon_{ij}$ is the efficiency value of the most efficient path between $i$ and $j$.

**Damage of Attack:** When a node is taken down by the attacker, the paths that previously traverse this node have to be rerouted. This will change the network efficiency as follows.

- Due to the rerouting, the nodes on the new routes need to carry extra load, which is previously carried by the nodes on the paths passing through the victim node.

- This may cause the overloading problem, which changes the efficiency of all edges connected to the overloaded nodes (see equation (60)).

- The changes in the edge efficiency will lead to the changes in the path efficiency, which may make some paths no longer be the most efficient paths.
• As a consequence, the most efficient paths between a generator and a distribution substation may need to be updated, which causes the load change of the nodes.

• Continuously, the load change may cause the overloading problem that result in load change again. This recursive process will continue until the average network efficiency stabilizes.

We use average network efficiency before and after the cascading failure to measure the damage. It is defined as [17],

$$\rho = \frac{E(G_0) - E(G_f)}{E(G_0)},$$

(63)

where $E(G_0)$ is the average network efficiency before the attack and $E(G_f)$ is the stabilized average network efficiency after the attack.

In summary, in this section we presented in detail of the existing and proposed new assessment metrics to understand the power system cascading failure under complex attack scenarios. In Part II of this series, we will show the limitations and advantages of each of these metrics for the non-recoverable models and recoverable models, which will be investigated based on the Western North American power grid benchmark.

### 3.5 Conclusion

In this Part I of the two-part series, we focus on the foundations, models, and assessment metrics to understand power system cascading failures under complex attacks. In particular, we provide a systematic discussion of the two representative network models, the non-recoverable model and recoverable models, and analyze their characteristics under different attack scenarios. In order to characterize the complex behavior of power grid under cascading failure, we also propose some new
assessment metrics such as RED and RIF to better represent and analyze power grid behavior under attacks. This part provides critical insights to understand the limitations of traditional largest load-based attacks, and layouts important foundations to develop advanced attack strategies and understand complex grid behavior as we will discuss in the Part II.

List of References


W. Wang, Q. Cai, Y. Sun, and H. He, “Risk-aware attacks and catastrophic cascading failure in u.s. power grid,” in *Proceedings IEEE Global Communications Conference (GlobeCom’11)*, 2011.


Manuscript-IV

In submission to IEEE Transactions on Information Forensics and Security, 2011. The earlier conference paper was accepted by IEEE Global Communication Conference in July 2011.

Topology based Modeling of Cascading Failures in Power Systems - Part II: Attack Strategies and Simulation Analysis

Wenkai Wang, Haibo He, and Yan Sun
Department of Electrical, Computer, and Biomedical Engineering
University of Rhode Island, Kingston, RI 02881

Corresponding Author: Yan Sun
Electrical, Computer, and Biomedical Engineering
University of Rhode Island
4 East Alumni Ave.
Kingston, RI 02881, USA
Phone:+1-401-874-5803
Email address:yansun@ele.uri.edu
MANUSCRIPT 4

Topology based Modeling of Cascading Failures in Power Systems - Part II: Attack Strategies and Simulation Analysis

Abstract

This paper is the second part of a two-part study addressing topology-based modeling of cascading failures in power systems. Part I presents a comprehensive analysis of the foundations, system models, and assessment metrics to understand this problem. In this Part II, we study specific attack strategies and analyze their simulation results based on the Western North American power grid benchmark under two representative topology based models. Our goal is to analyze the power grid behavior and find effective attack strategies when the attacker can take down one or multiple nodes. The first model we investigated is the non-recoverable model, in which overloaded nodes fail to operate, and the second network model is recoverable model, in which overloaded nodes are still in function but their performance in power delivery is reduced. In both network models, the proposed attack strategies, which represent novel ways for joint consideration of load and topology, are much more destructive than the traditional load based strategies.

4.1 Introduction

In the first part of this two-part study, we have presented a systematic discussion regarding the foundations, models, and assessment metrics to understand power system cascading failures under complex attacks. In this part, our major focus is to develop advanced attack strategies and analyze the grid behavior when two representative models, the non-recoverable model and recoverable model, are considered.

Due to the importance of power grid security, there are many efforts in the
community to study the cascading failures in power grid from different aspects. The first type of approaches consider the case when only one single node in the power grid can be knocked down. Load based analysis has been an important consideration in these approaches [1, 2, 3, 4, 5, 6]. For instance, in [4] it found that the breakdown of a single node is sufficient to collapse the efficiency of the entire system if the node is among the ones with top largest load. Interesting work has presented that it might not be the best strategy to always knock down the node with the largest load. For instance, in [1, 6] it was claimed that under different parameter settings, attack on the nodes with top lowest load is in fact could be more harmful than the ones with top highest load. While this counter-intuitive finding is interesting, it is not clear whether this observation is an actual reflection of the physical characteristics of the power grid [5]. In another cascading failure analysis, a more comprehensive power grid was studied [2]. Simulation results suggested that the loss of a single substation can result in up to 25% loss of transmission efficiency.

In addition to single-node attack, some approaches also aim to investigate the grid behavior when attackers have the capability to take down multiple nodes. For instance, in [7], attack-induced cascading failure subject to multiple-node attacks that consider the physical quantity in the electrical grid was presented. Simulated cascades triggered by random failures and by intentional attacks suggested that global cascade could only happen if the network exhibits a highly heterogeneous distribution of load [7]. In [8], attack and error tolerance of complex networks were analyzed, and it was concluded that scale-free networks actually have a surprisingly high degree of tolerance against random failures.

In understanding the attacks and cascading failures of power grid, it has been widely recognized that topological structure has a key impact [9, 10, 11, 12, 13, 14].
Most recently, the cascading failure of interdependent networks of power grid and 
communication network was presented in [13]. This work suggested a failure of 
a very small fraction of nodes in one network may lead to catastrophic failure 
of several interdependent networks. In [12], three topological based mitigation 
strategies were presented and analyzed of power grid. In [10], a comparative study 
of different topological structure including random, preferential-attachment, and 
small-world structure was analyzed. It was concluded that power grid responses 
substantially differently under these abstract models.

While many of the existing work focus on finding strong or even optimal attack 
strategies, people also looked at the problem from the defense point of view. For 
instance, in [15], the important question regarding how to design networks of finite 
capacity that are safe against cascading failures was discussed, and a theoretical 
upper bound for the network capacity was derived. Game theory based defense 
strategies were presented in [16, 17], in which the interaction between attackers 
and defenders was envisaged as a game. Another approach to defend against 
deliberate attacks was expanding the transmission network to mitigate the impact 
of cascading failures [18, 19].

Note that most of the work on power grid security assumes the attack happens 
on nodes, a few studies have investigated the system behavior when the attack 
happens on the grid links [7, 20, 21]. In [7], range-based attacks on link was 
presented. It was concluded that many scale-free networks are in fact more sensitive 
to attacks on short-range than on long-range links. The assessment of long-term 
effect of the n-1 criterion on cascading line failure was analyzed in [20]. And in [21], 
mathematical models were built to identify critical system components, including 
transmission lines, generators, and transformers in power grid.

While the existing research results have provided many useful suggestions and
insights about cascading failures in power grids, the fully understanding of its behaviors under different types of attacks still remains a grand challenge. We still do not know whether load is the most effective metric for finding strong attacks, although it is used in almost all existing works. We also have very limited knowledge on effective multiple-node attacks, in which attackers take down multiple nodes simultaneously. Like risk analysis and vulnerability evaluation [22, 23, 24, 25] in most security studies, in this paper, we aim to investigate the vulnerabilities of power grid network and propose specific single-node and multi-node attack strategies. In particular, we study one representative non-recoverable model proposed in [1], and one representative recoverable model proposed in [2]. Simulations are performed based on Western North American power grid network data [26]. The simulation results as well as analysis suggest the following important discoveries.

• Traditional attack strategies are mostly based on load, which might not be the best metric in developing strong attack strategies. For example, in single-node attacks, the RIF based strategy is much stronger than the load based strategy under the non-recoverable model.

• Load is particularly not good when attackers can attack multiple nodes. In the recoverable model, our proposed clustering based multi-node attack strategy can reduce 30% more network efficiency than the load based strategy.

• Network topology plays a critical role in cascading failures. Jointly investigating network topology and load distribution is a promising direction for finding strong attacks against power grid. For the non-recovery model, an effective method is to investigate the load and the risk of a neighborhood, instead of individual node. For the recoverable model, an effective method is to attack multiple nodes that are sufficiently apart in the topology.
Although we focus on attacks in this paper, we generate insightful results for defense as well. The proposed attack strategies identify critical nodes in the power grid. From the attack point of view, taking down these critical nodes will lead to severe cascading failures. From the defense point of view, protecting these critical nodes will lead to more robust grid than simply protecting the nodes with the top largest load.

The rest of the paper is organized as follows. Section 4.2 discusses the proposed attack strategies under non-recoverable models and Section 4.3 discusses recoverable models. Detailed simulation results and analysis are presented in Section 4.4 based on the Western North American power grid benchmark. Finally, conclusions and discussions on future research are presented in Section 4.5.

4.2 Attack Strategies under Non-recoverable Model

In this section, we study attack strategies assuming the non-recoverable network model. Based on the metrics as we discussed in Part I (Section IV) of this series, we discuss the single-node attack strategies in Section 4.2.1 and the multi-node attack strategies in Section 4.2.2.

4.2.1 Single-node Attack Strategies

We denote single-node attack strategy by $AS^s$. Attackers will try to find the optimal single-node attack strategy, denoted by $AS^s_{opt}$, such that taking down of the selected victim node can cause the most severe cascading failure. In Part I, we showed that Required Redundancy (RED) is a good metric to describe the critical level of nodes. Using RED as a measure, the optimal single-node attack strategy is

- $AS^s_{opt}$: Among all $N$ network nodes, choose the victim node as the node with the largest RED value.
However, we know that the computation cost of obtaining RED for all nodes is high. Instead, as analyzed in the part I, RIF is simple in calculating and has the good property of RED in differentiating nodes. Therefore, we define a RIF based attack strategy as

- $AS_{rif:K}^s$: Among the top $K$ nodes with the highest RIF value, choose the node that yields the largest RED value as victim node.

Since the calculation of RIF is easy, the computation complexity of $AS_{rif:K}^s$ is about $K/N$ of the computation complexity of $AS_{opt}^s$.

Furthermore, we will compare $AS_{rif:K}^s$ with two other strategies,

- $AS_{rand:K}^s$: From $K$ randomly selected nodes, choose the node that yields largest RED value as the victim node.

- $AS_{load:K}^s$: Among the top $K$ nodes with the largest load, choose the node that yields largest RED value as victim node.

Note that the three attack strategies ($AS_{rif:K}^s$, $AS_{rand:K}^s$, $AS_{load:K}^s$) have roughly the same computation complexity. In other words, if the attacker can compute RED values for $K$ nodes, the attacker can use one of the three attack strategies. The comparison results will be shown in Section 4.4.1.

### 4.2.2 Multi-node Attack Strategies

The optimal multi-node attack strategy, denoted by $AS_{opt}^m$, is to find $M$ victim nodes such that their simultaneous failure yields the largest percentage-of-failure value, given a specific system tolerance $T$ value. It can be easily seen that $AS_{opt}^m$ requires to compute the percentage-of-failure values for $\frac{N!}{M!(N-M)!}$ possible combinations of victim nodes. The computation complexity can be prohibitively high.
To develop practical multi-node attack strategies, we obtain the following insights from the previous results for single-node attack strategy.

Let $\text{RED}_{\text{max}}$ denote the maximum RED value of all nodes. From part I analysis, we can see that $\text{RED}_{\text{max}} = 1.8$ for the Western North American power grid. Recall that RED is defined as the minimum system tolerance value such that there is no cascading failure when a node taken down. According to the definition, when $T = \text{RED}_{\text{max}}$, there is no cascading failure no matter which single node is taken down. When $T \geq \text{RED}_{\text{max}}$, only multiple-node attacks can cause cascading failures. We consider three scenarios when $T \geq \text{RED}_{\text{max}}$.

In the first scenario, shown in Fig. 30 a), two victim nodes (e.g. A and B) are one-hop neighbors. When A is down, the load of A is redistributed to A’s neighbors except B. In this case, the A’s remaining neighbors will receive higher additional load, compared to the case that only node A is taken down in the single node attack. Cascading failure may occur in this scenario.
In the second scenario, shown in Fig. 30 b), two victim nodes (e.g. F and G) are two-hop neighbors. Then, the common neighbor of F and G has to carry the additional load from both F and G, when both F and G are knocked down. The additional load received by the common neighbor (i.e. N) will be higher, compared to that in the single node attack case. Cascading failure may occur in this scenario.

In the third scenarios, shown in Fig. 30 c), two victim nodes are far away from each other (i.e. more than two hops away). We can envision that only those two nodes will break down and the rest nodes will not be affected, when $T \geq RED_{\text{max}}$.

**Procedure 3** RIF based M-node attack strategy

1: Compute the RIF values for all nodes
2: Pick the node $i$ which has the highest RIF value
3: Put all of the $i$’s one-hop and two-hop neighbors in set $S$
4: for each unique selection of $M - 1$ nodes from $S$ do
5: Compute the percentage-of-failure when node $i$ and these $M - 1$ nodes are taking down;
6: end for
7: Among all computed percentage-of-failure values, find the maximum one; The corresponding set of nodes is chosen as the victim nodes.

The above scenarios can be easily extended to the case when the number of victim nodes is greater than two. Based on these insights, we propose to choose victim nodes within a neighborhood, instead of from the entire network. In particular, a RIF based multi-node attack strategy, denoted by $AS_{\text{ rif }, M}^m$, is described in Procedure 3.

Similarly, we propose a load based multi-node attack strategy $AS_{\text{ load }, M}^m$, by replacing the word “RIF” in line 1 and line 2 in Procedure 3 with “load”. Note that the $AS_{\text{ load }, M}^m$ is not the traditional load based strategy. In the traditional load based strategy, the attacker simply picks the top K largest load nodes as the victim node. The traditional strategy will not work in many cases. For example, in the Western North America power grid network, the top 1 and top 2 largest
Figure 31. Topology of top 20 largest nodes and their one-hop neighbors
load nodes are 24 hops away when load is calculated using degree based model.
When $T \geq 1.8$, taking down these two nodes will not cause cascading failures.
Therefore, $AS_{\text{load},M}^m$, a load based strategy based on our investigation on multiple-
node attacks, is a stronger attack than the traditional load based strategy.

In the proposed multi-node attack strategies, we only need to test $\Psi \choose M$ pairs of victim nodes, where $\Psi$ is the number of one-hop and two-hop neighbors of the node with the highest RIF or load. As $\Psi$ is much smaller than the total number of nodes $N$, the proposed strategy requires much less computation than the optimal strategy. Their performance will be demonstrated in Section 4.4.1.

4.3 Attack Strategies under Recoverable Model

In Section 4.2, based on the non-recoverable model, we have discovered that: 1) load based strategy is not an effective attack strategy; 2) in the multiple-node attack strategies, the concept of neighborhood plays an important role. In this section, we will extend our study to the recoverable model that is more complicated. In particular, we first present single-node attack strategies in Section 4.3.1 and then discuss multi-node attack strategies in Section 4.3.2.

4.3.1 Single-node Attack Strategies

Under the recoverable model, the traditional load based strategy, denoted by $AS_{l}^s$, is
- $AS^*_i$: choosing the victim node as the node with the largest load.

This is a strong single-node attack strategy. Recall that load is defined as the number of paths that traverse the node in the recoverable model, when the victim node has the largest load, a large number of paths need to be rerouted, which may cause more nodes on the new paths to be overloaded. As a result, the network efficiency can be greatly reduced.

The load based single-node attack strategy is effective but not optimal. We conduct simulations using the network topology of the Western North American power grid network, and search for the optimal strategy. We find that

- When attacking the node with the largest load (i.e. $AS^*_i$), the network efficiency decreases from 0.0594 to 0.035.

- When attacking the node with the second largest load, the network efficiency decreases from 0.0594 to 0.033.

In this experiment, choosing the victim node as the one with the second largest load is a stronger strategy than $AS^*_i$.

Although $AS^*_i$ is not optimal, we still adopt it as the single-node attack strategy for two reasons. First, the performance of $AS^*_i$ ($\rho = 0.41$) and the performance of the optimal strategy ($\rho = 0.44$) is very close. Second, the computation complexity of $AS^*_i$ is extremely low. It is important to point out that obtaining the network efficiency value is computational expensive when the network size is large. Before the network efficiency stabilizes, in each iteration, one needs to find the most efficient paths between all generator and distribution substation pairs. When network size is large, simulating all possible single-node attacks and find the optimal one requires very high computation complexity. On the other hand, in $AS^*_i$, there is no need to compute network efficiency at all. Therefore, in this paper, we adopt
the existing load based single node attack strategy for the recoverable model, and focus on developing multiple-node attack strategies.

4.3.2 Multi-node Attack Strategies

When the attacker can take down $M$ victim nodes, the load based strategy is

- $AS_l^m$: choosing the victim node as the top $M$ largest load nodes.

This strategy, however, is not an effective attack strategy. We conduct experiments based on the Western North American power grid topology and made several important observations.

First, in Fig. 31, we show the topology of top 20 largest load nodes and their one-hop neighbors. It is seen that the nodes with large load can be very close to each other. For example, the top 4 largest load nodes (node 4220, 2544, 4165, and 2529) are within 3 hops of each other.

Second, when the attacker chooses two victim modes, according to $AS_l^m$, the victim nodes should be node 4220 and node 2544. In this case, the average network efficiency will be reduced from 0.0594 to 0.031. The damage of the attack is $\rho = 0.478$, which is only slightly higher than the damage of the single node attack ($\rho = 0.41$, see Section 4.3.1). In other words, taking down one additional node does not make the attack much stronger. The reason is that node 4220 and node 2544 are neighbors. Many paths that pass node 4220 also pass node 2544. No matter just node 4220 is taking down or both are taking down, similar paths need to be detoured away from the neighborhood of node 4220.

Third, we argue that an effective multiple-node attack strategy should consider not only the load but also the distance between victim nodes. For example, if the attacker takes down node 4220 (rank 1) and node 70 (rank 12), the stabilized average network efficiency is 0.026 and the damage of the attack is $\rho = 0.56$. This
is a much stronger attack than $AS_i^m$, partially because the two victim nodes are 5 hops away and are in different neighborhoods (or regions) of the network.

Inspired by the above investigation, we propose to jointly consider the load and the regions of the network, and design a strategy for selecting $M$ victim nodes as

- $AS_c^m$: Choosing the victim nodes that have large load and belong to different regions of the network according to the three-step procedure in Procedure 2.

**Procedure 4** Multi-node attack strategy $AS_c^m$ under recoverable model

1: Select top $R$ nodes with largest load from all nodes
2: According to the distance (measured by the number of hops) among these $R$ nodes, divide these $R$ nodes into $M$ clusters.
3: **for** each cluster **do**
4: Select node with the largest load within the cluster as one victim node
5: **end for**
6: Totally $M$ victim nodes are chosen. The attacker takes down these $M$ victim nodes simultaneously.

In the first step, top $R$ ($R > M$) largest load nodes are selected and put into a candidate pool for victim nodes. As mentioned in the previous section, load plays an important role for cascading failures under the recoverable model. We want to select victim nodes that have high load, but not necessarily the highest load. In the second step, we run a clustering algorithm based on the distance among the nodes in the candidate pool, and divide all candidates into $M$ groups. The aim of this step is to identify critical regions of the power grid. In the third step, the node with the highest load in each cluster is selected as one victim node, and totally $M$ victim nodes are chosen.

Obviously, $AS_c^m$ is not an optimal strategy. However, it represents a big step forward from the traditional load based attack strategy. In $AS_c^m$, the victim nodes carry high load and represent different regions in power grid. When these victim
nodes are attacked (meaning many critical nodes in different regions are taken down), the alternative paths can be much longer than the original paths between the generators and the substations. As a consequence, the network efficiency can be greatly reduced. The detailed simulation results will be shown in Section 4.4.2.

In addition, the computation complexity of $AS_c^m$ is only slightly higher than that of $AS_l^m$, due to the clustering algorithm. It does not require the computation of the network efficiency. Recall that there is no close form representation of the network efficiency, and searching for the optimal attack strategy is prohibitively expensive in the recoverable model.

In the proposed multi-node attack strategy, we need to choose the parameter $R$, the size of the candidate pool. When $R$ is too large, the candidate pool will include many nodes whose load is not very high. As a result, some of the selected victim nodes may just carry relatively low load. Taking down these nodes will not cause severe cascading failure in their regions. For example, when $R = 2000$ and $M = 4$, the stabilized average network efficiency is 0.028. When we reduce $R$ to 200, the network efficiency can be as low as 0.02. (In our network topology, there are total 4941 nodes.) On the other hand, if $N$ is too small, the initial candidate set cannot represent different regions in the network. For instance, clustering 20 top largest load nodes into 10 groups ($R = 20$ and $M = 10$) usually does not make significant difference from simply choosing top 10 largest load nodes. We have conducted simulations for different $R$ values and found that the performance of the proposed strategy is not sensitive to the selection of $R$ as long as $R$ is between 50 and 1000. For the simulation results in Section 4.4.2, we choose $R = 100$.

Besides the proposed strategy and the load based strategy, another intuitive multiple-node attack strategy is

- $AS_d^m$: Choosing the victim nodes that have the top $M$ largest degree.
The philosophy of this attack strategy is, taking down the nodes with larger degree will affect more edges. The proposed attack strategy will be compared with $AS_t^m$ and $AS_d^m$ through simulations in Section 4.4.2.

Finally, we summarize the attack strategies discussed in Section 4.2 and Section 4.3 in Fig. 32.

### 4.4  Simulation Results

We choose the Western North American power grid network [26] as the benchmark in simulation. The network is composed of 4941 substations and 6594 transmission lines. We built a simulator using Matlab to simulate the load redistribution process and the consequences of different attacks. The simulation results for non-recoverable model and recoverable model are shown in Section 4.4.1 and 4.4.2, respectively.

#### 4.4.1  Simulation Results for Non-recoverable Model

In Section 4.2.1 we described three **single-node attack** strategies: the traditional load based strategy $AS_{\text{load}:K}^s$, the random strategy $AS_{\text{rand}:K}^s$, and the proposed RIF based strategy $AS_{\text{rif}:K}^s$. For $AS_{\text{rand}:K}^s$, we run the simulation for 1000
times for each parameter setting. Then we obtain the average RED value or percentage-of-failure value.

We simulate the three attack strategies and compare their performance in Fig. 33. The x-axis is $K$, the number of candidate victim nodes. It changes from 1 to 60. The y-axis is the RED value of the selected victim node. Recall that RED is the minimal required system redundancy such that cascading failure will not occur if the victim node is taken down. The higher RED value is, the stronger the attack is. The number of candidate victim nodes $K$ will affect the computation cost of simulation. The higher the $K$ value is, the more computation is needed to find the victim node. From Fig. 33, we have made the following observations,

- When $K = 1$, attackers choose the node with largest load in $AS_{load:1}^s$, choose the node with largest RIF in $AS_{rif:1}^s$, or randomly choose a node in $AS_{rand:1}^s$. The RED values for $AS_{load:1}^s$, $AS_{rand:1}^s$ and $AS_{rif:1}^s$ are 1.225, 1.22 and 1.8, respectively. We can see that the performance of the load based attack is similar to that of the random attack. And they are both much less effective than the RIF based attack. Actually, the RIF based strategy already achieves the maximum RED value for $K = 1$. Although this result is for this specific power grid network, it shows the obvious advantage of the RIF based attack strategy.

- As $K$ increases, the RED value for $AS_{rand:K}^s$ increases gradually. This is because the search for the victim node is conducted among a larger set of nodes. The RED for $AS_{rif:K}^s$ remains the same because it already achieves the maximum RED value when $K = 1$. The RED for $AS_{load:K}^s$ increases with $K$ and achieves the maximum value when $K = 44$.

In Fig. 34, we demonstrate the percentage-of-failure of the three attack strategies when system tolerance $T = 1.6$. We can see that when $K = 1$, $AS_{rand:1}^s$ and
$AS_{load:1}^s$ do not cause cascading failure because their percentage-of-failure values are almost 0, whereas $AS_{rif:1}^s$ causes a severe cascading failure since the percentage-of-failure is 96%. As $K$ increases, the percentage-of-failure for $AS_{rand:K}^s$ and $AS_{load:K}^s$ increases. $AS_{load:K}^s$ can cause cascading failures when $K \geq 5$, while its percentage-of-failure is always lower than that of $AS_{rif:K}^s$.

In conclusion, the RIF based strategy is a very effective attack strategy for single-node attack. Compared to the load based strategy and random strategy, it has very low computation complexity and yields strong attack. Actually, under the non-recoverable model, simply picking the node with highest RIF value would yield a very strong attack.

For multi-node attack, we simulated four attack scenarios based on the strategies described in Section 4.2.2: 1) load based two-node attack $AS_{load:2}^m$; 2) RIF based two-node attack $AS_{rif:2}^m$; 3) load based three-node attack $AS_{load:3}^m$; 4) RIF based three-node attack $AS_{rif:3}^m$. The simulation results are shown in Fig. 35. The x-axis is system tolerance value $T$ and the y-axis is the percentage-of-failure under the four scenarios.
Obviously, for all strategies, the percentage-of-failure reduces as $T$ increases. In other words, when the nodes can handle less additional load (i.e. smaller $T$), the number of failed nodes increases (i.e. higher percentage-of-failure). From the percentage-of-failure curves, we can obtain the RED values. Let us gradually reduce the $T$ value and see when the cascading failures occur in all attack strategies.

- When $T$ is greater than 2.225, the percentage-of-failure for all strategies is almost 0. There are no cascading failures in this case.

- As $T$ reduces to a little below 2.225, the proposed $AS_{rif:3}^m$ causes a cascading failure in which about 22% of nodes are taken down (see the sharp change in the percentage-of-failure value). Thus, the RED value of $AS_{rif:3}^m$ is 2.225.

- As $T$ reduces to below 2.075, the proposed $AS_{rif:2}^m$ causes a cascading failure in which about 62% of nodes are taken down. Thus, the RED value of $AS_{rif:2}^m$ is 2.075.

- Similarly, we see that the RED value for $AS_{load:3}^m$ is 2.0, and for $AS_{load:2}^m$ is 1.5.
Obviously, $A_{rif}^{m:3}$ is much stronger than $A_{load}^{m:3}$, and $A_{rif}^{m:2}$ is much stronger than $A_{load}^{m:2}$. We observe that the RIF based attack strategy is stronger than the load based attack strategy in multi-node attacks. In addition, given a system tolerance value, the RIF based strategy can cause cascading failures by taking down the less number of substations. For example, when $T = 2.05$, the RIF based attack will need 2 victim nodes, and the load based attack will need at least 4 victim nodes, in order to cause cascading failures.

### 4.4.2 Simulation Results for Recoverable Model

In the recoverable model, load is calculated as betweenness. To compute betweenness, we need to find the most efficient paths for all generator and distribution substation pairs. However, the Western North American power grid data contains only general substations information and does not specify whether a substation is generator, transmission substation, or distribution substation. In our simulation, we identify distribution substations using a method similar to that in [9]. That is, substations that have only one transmission line connected to them have high potential to be distribution substations. There are 1226 nodes that have...
only one connection in the Western North American power grid. Therefore, we randomly select 800 nodes from the 1226 node set as distribution substations and randomly select 600 nodes from the remaining nodes as generators. By doing so, we can create one snapshot of the power grid topology. Obviously, different topology snapshot can be generated when we select different generators or distribution substations. In this paper, we generate two topology snapshots for testing purpose.

For the proposed multi-node attack, we choose the size of the candidate set be 100 (i.e. $R = 100$), and vary the number of victim nodes ($M$) from 1 to 10. In the step 2 of Procedure 2, we use K-means [27] clustering algorithm to classify the $R$ nodes into $M$ groups. It is well known that there is randomness in the clustering results, especially when $M$ is large. To reduce this randomness, we plan to run the simulation 1000 times for each $M$ value. Recall that for each selection of the victim nodes, the computation of the network efficiency requires finding the most efficiency paths between each generator and distribution substation pair in one iteration round and there are multiple rounds. Repeating the simulation for 1000 times is computationally prohibitive.
To solve this problem, we use an approximation method as follow. First, for each $M$ value, we run the clustering algorithm 1000 times and record all possible selections of the victim nodes. Here, we have not computed the network efficiency yet. Let $S_j$ denote one selection of the victim nodes, and assume $S_j$ appears $n_j$ times. Second, we rank $S_j$ according to $n_j$ from high to low. For example, in one network topology snapshot, when the number of victim nodes is two, the most frequently appeared victim nodes selection (denoted by $S_0$) appears $n_0 = 554$ times in 1000 tests. Third, we find the $r$ value such that $\sum_{j \geq r} n_j < 0.01 \times 1000$, and remove the victim nodes selection $S_j, \forall j \geq r$. In other words, we only keep frequently occurred selections that contribute to more than 99% of possible victim selections. Fourth, we compute the network efficiency for each $S_j, \forall j < r$. Assume the network efficiency is $NE_j$ when the victim nodes are selected as $S_j$. Then, the final result is $NE_{avg} = \frac{\sum_{i:\forall j < r} NE_i \cdot n_j}{\sum_{i:\forall j < r} n_j}$. From the above description, we can see that $NE_{avg}$ is a good approximation of the average network efficiency when the simulation is performed 1000 times. To obtain $NE_{avg}$, we need to compute network efficiency for $(r-1)$ times. With this approximation, the complexity of the simulation is greatly reduced. For example, in our experiments when the number of victim nodes $M$ is 5, $r$ is 9 for both topology snapshots.

In Fig. 36, we show how the network efficiency changes during the cascading failures. The x-axis is iteration round. It may take several iteration rounds for the system to choose the most efficient paths before the final network efficiency stabilizes. The y-axis is average network efficiency. In the simulation, we set the system tolerance value $T$ to be 1.2. The attack strategy is the load based $AS_l^m$. The three curves represent the consequence of taking down one, two, or three victim nodes, respectively.

For example, when there is one victim node, the network efficiency drops
quickly in the first few rounds because more nodes experience the overloading problem after re-routing. After a few rounds, the network efficiency starts to fluctuate around 0.035. This is because the system always tries to find more efficient paths as long as there are overloaded nodes. To compute the final network efficiency, we record network efficiency values in 8 rounds after the network efficiency starts to fluctuate. Then, we take the average of these 8 values as the final network efficiency.

For the multi-node attacks, we simulate three attack strategies: the degree based strategy $AS_d^m$, the load based strategy $AS_l^m$, and the proposed clustering based strategy $AS_c^m$. Fig. 37 shows the results for one network topology snapshot, and Fig. 38 shows the results for another topology snapshot. In these two figures, the x-axis is the number of victim nodes and y-axis is the stabilized average network efficiency. We make the following observations.

- $AS_d^m$ is the weakest attack strategy. This is because a node has high degree does not necessarily on the most efficient paths between a large number of generator and distribution substation pairs. If a node is not on the most
The proposed $AS^m_{mc}$ is stronger than the load based $AS^m_l$, because $AS^m_{mc}$ can reduce the network efficiency more. For example, when there are three victim nodes ($M = 3$), under topology snapshot 1, the proposed attack reduces the network efficiency from 0.0594 to 0.02, whereas the load based strategy reduces the network efficiency from 0.0594 to 0.029. The average network efficiency after the proposed attack is 31% lower than that of the load based strategy. Under topology snapshot 2, when $M = 3$, the average network efficiency after the proposed attack is 33% lower than that of the load based strategy. This verifies our analysis in Section 4.3.2 that $AS^m_l$ is not the most effective attack strategy for multi-node attacks since it does not consider topology. The proposed $AS^m_{mc}$ is a stronger attack strategy because it can take down critical nodes in different regions.

- The results under two topology snapshots are similar. This indicates that the
proposed attack strategy could also cause severe damage to other network topologies, and be a severe threat to other types of power grid networks.

### 4.5 Conclusion and Future Work

In this paper, we studied the cascading failures in power grid networks based on topology models. In particular, we investigated the attack strategies that describe how to choose victim nodes in order to cause severe cascading failures under two load redistribution models. Our study showed that the traditional load based attack strategies often do not yield the strongest attacks. Under the non-recoverable model, RIF based attack strategy has low computation complexity and can cause severe cascading failures. This strategy is stronger than the load based strategy in both single-node attacks and multi-node attacks. Under the recoverable model, although the load based attack strategy is good in single-node attacks, it is not the strongest attack strategy in multi-node attacks. We proposed an attack strategy that jointly considers load and distance among victim nodes. The proposed attacks can significantly reduce network efficiency compared to the load based attacks. The investigation on attack strategies not only enabled us to identify critical nodes in power grid, but also yielded in-depth understandings of the power grid networks.

In this series of two parts, our investigation on cascading failures focuses on attacks against network nodes using topology based models. In future, we will extend the work in following directions. First, we can evaluate cascading failure using power flow analysis, which may capture more fundamental understandings of power grid under complex attacks. Second, we plan to investigate the cascading failures when links in power grid are attacked. Furthermore, we can study the combination of node attack and link attack, which is expected to be more harmful than node attack alone or link attack alone. Third, the current study is based
on western United States power grid data. It will be interesting to simulate more power grid benchmarks to discover their common vulnerabilities. Fourth, this work mainly focuses on developing attack strategies. It is critical to propose effective solutions to defend against the various attacks to protect the power grid.

List of References


“Common cyber security vulnerabilities observed in control system assessments by the inl nstb program,” National SCADA Test Bed (NSTP), 2008.


Wang, W., Cai, Q., Sun, Y., and He, H., “Risk-aware attacks and catastrophic cascading failure in u.s. power grid,” in *Proceedings IEEE Global Communications Conference (GlobeCom’11)*, 2011.


