STUDY OF DATA ACQUISITION SYSTEMS ENHANCING FOR THE SECURITY IN MULTISTAGE INTRUSION DETECTION

Vinod Wamanrao Gangane
Research Scholar
Department of Computer Science
Sunrise University
vinod.gang43@gmail.com

Dr. Jitendra Rai
Research Guide
Department of Computer Science
Sunrise University

Abstract

The fast development of information and communication technology has resulted in the emergence of sophisticated critical infrastructures that may be operated by supervisory control and data acquisition (SCADA) systems that are computer-based (SCADA). Numerous critical infrastructures, including water distribution and wastewater treatment facilities, electricity generation and distribution facilities, oil refineries, gas pipelines, nuclear power plants, chemical processing facilities, rail and other public transportation systems, and other facilities, use SCADA systems. Information transmission and presentation to/from a variety of resources and locations without compromising data integrity or system security is the main purpose of SCADA systems. The usage of SCADA systems is expanding as system automation develops and gains importance on a global scale because it offers a real-time picture of the whole system by keeping track of spatially scattered sensors and actuators. In general, SCADA has evolved through three generations, and as a result, the goals for the design, security concerns, speed, scalability, and efficiency, to mention a few, have changed over time.

Introduction

Prior until recently, the efficiency and efficacy of SCADA systems took precedence above security as a primary design consideration. In the past, the majority of SCADA providers and customers have used two security approaches: security via obscurity and air gaps [1]. The SCADA network and other networks would be completely isolated under the air gap-based protection technique. It would be challenging for an attacker to access the SCADA network in such a situation. This strategy, however, is unreliable since an attacker may access the system via other ways of assault. As an example, Stuxnet [2], the first virus to expressly target industrial programmable logic control systems, attacked the Iranian nuclear complex using a USB flash drive.

The security by obscurity strategy, on the other hand, is predicated on the idea that the public does not have widespread access to or knowledge of the system. Only suppliers and authorized users are thought to be aware of their surroundings and systems. However, since employing commercially available systems, the specifics of SCADA components are no longer kept a secret. Due to the replacement of special-purpose systems with off-the-shelf systems, the security via obscurity strategy fails to safeguard SCADA systems.

More recently, SCADA technology moved from isolated systems to networked designs that connect to corporate networks and the internet in order to increase efficiency, competitiveness, and production. This made it possible for open SCADA protocols to take the place of proprietary ones and made it possible for SCADA components from various suppliers to communicate with one another. However, at the same time, the growing interconnection and adoption of standard protocols has left SCADA systems vulnerable to intrusion or malicious assaults. Attackers on SCADA systems may include hostile countries and foreign intelligence agencies, terrorist organizations, industrial spies, resentful employees, ideological activists (such as anti-nuclear groups), bot-network operators, and hackers [3], [4]. The security of the SCADA systems and the processes they control has become a particularly delicate subject since they are costly and essential in nature.

The latest cyberweapon of cyberwarfare, the Stuxnet worm, has attacked the Iranian nuclear complex in June 2010. The first piece of malware to expressly target industrial programmable logic control systems for disruption is the Stuxnet worm. The Siemens default passwords are used by the worm, which progressively changes the PLCs' operating code to behave in a manner inconsistent with their intended usage. More particular, Stuxnet changed the electrical current's frequency for the drives, causing them to alternate between high and low speeds for which they were not intended [2], [6], [8], [9]. Stuxnet transforms the character of cyberwarfare in the digital sphere and illustrates the transformation in military affairs [10]. According to a McAfee investigation, coordinated, targeted cyberattacks dubbed Night Dragon have been launched against international energy and oil corporations since 2009 [11].

The term "Night Dragon" refers to a group of assaults that include phishing, Trojans, social engineering, and Windows-based vulnerabilities [12]. In these events, corporate networks that were linked to SCADA infrastructures were hacked rather than SCADA systems themselves. In addition to the aforementioned occurrences, more recent reports claim that assaults on SCADA systems are becoming alarmingly often. For instance, a survey by Dell [13] found that in 2014, there were three times as many SCADA assaults as there were in 2013. Following all of these occurrences, there have been several industry and governmental initiatives recently to increase the security of SCADA systems. The conventional IT infrastructure is where the majority of these security solutions are adapted from. But there are fundamental distinctions between a SCADA system and a traditional IT system. The SCADA system and the traditional IT system vary from a security standpoint as well.

The most crucial security objectives in a traditional IT system are confidentiality, integrity, and availability (CIA). As a result, access control and encryption are often utilized to preserve security in the IT sector. Of contrast to an IT system, the most crucial factors in a SCADA system are safety, reliability, and availability (SRA) [4], [14]. As a result, while SCADA systems were being designed, cyber security concerns were not given any thought. For SCADA systems, essential security concerns include confidentiality, integrity, and availability in addition to safety and dependability. However, in contrast to the emphasis in IT systems, availability has the greatest importance in SCADA systems, followed by integrity and confidentiality [15].

This priority was established due to the possibility that a catastrophic catastrophe in SCADA systems may occur if a controller or other system component went down, endangering lives or causing extensive environmental harm. Therefore, maintaining the availability of all system components and controlled plants is the fundamental security problem in SCADA systems [16]. Companies have adopted a range of IT security solutions during the last ten years in an effort to safeguard the confidentiality, integrity, and availability of their SCADA network. Despite offering some protection, these systems have been proven to have a variety of shortcomings.

- For illustration: - Patching: One of the core security techniques for addressing security flaws in conventional IT is patching. However, the SCADA system necessitates a lengthy outage of the controlled infrastructure in order to patch each component of the system. Downtime is the outcome, which is unacceptable for essential infrastructures.

- Access control: Access control may be implemented in conventional IT systems with no consideration for data flow. Data flow must not be disrupted for the majority of essential infrastructures, and system reaction to human involvement or automated response time is crucial. For instance, password authentication and authorisation on a Human Machine Interface (HMI) in critical infrastructure should not cause delays for emergency actions [4].

- Cryptographic solutions: The majority of SCADA communication protocols, including Modbus and Distributed Network Protocol Version 3 (DNP3), lack authentication capabilities and don't verify the validity of command and response packets [17][18]. This made it easier for highly motivated and competent attackers to create or change legitimate network packets in order to inject them into a SCADA system. Attackers were able to alter the SCADA system's measurement and control data thanks to bogus data injection.

Since a few years ago, protocols like Flexi-DNP3 (Flexible DNP3) [20] and Secure Version of Modbus [19] have been created by integrating cryptographic solutions into the protocol to increase the security capabilities of communication protocols. However, implementing these protocols, which were created in simulation settings, in actual situations might be difficult since it can increase production and safety risks. Furthermore, the resource limitations in SCADA components make it difficult to deploy conventional cryptographic methods in SCADA settings.

SAMPLE DESIGN

In certain cases of science, analyzing the entire universe is almost impossible; the only alternative will to use sampling. The current researches will the same character. The procedure for deciding the sample of the analysis is sampling of the Study of the security issues and existing intrusion detection base solutions in SCADA systems. Looking for and investigating publically available intrusion detection datasets. This will do to examine the reason behind poor detection accuracy towards the minority attack classes by existing works and to study patterns that will identify minority attacks from normal.

DATA COLLECTION

Data collection is the systematic way to collect and measure data from sources to get complete and precise data for research activities. In all areas of study the facts collection component is not unusual with body and social sciences, the humanities and corporations. It allows scientists and analysts to collect key factors as the information they collect. In contrast with the approaches in terms of subject matter, the value of maintaining the right and truthful sequence remains the same. Current data collection is essential for preserving the credibility of research and for ensuring excellent outcomes and their findings. This study will be secondary research methods.

SECONDARY DATA

Secondary data are the data collected by an individual rather than the user. A researcher who is not associated with the analysis / recherché study collects secondary information for a different purpose, and in the past at quite different times such data are readily accessible and cost effective in comparison to primary data.

Sources of secondary data collection are as follows:

- Government department's journals,
- Organizational records,
- Magazines,
- Journals, books,
- Newspapers and
- The information which is collected originally for other research purpose.

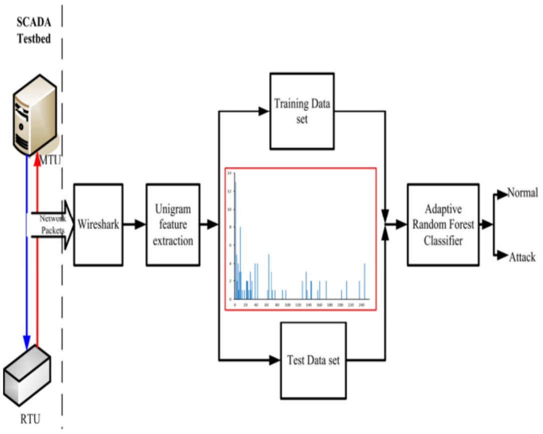

Figure- Experimental setup for SCADA IDS

This experiment made use of a SCADA data collection that had a total of 2500 normal occurrences and 698 anomalous (attack) instances. This information is distributed in a haphazard manner over 2238 training instances and 960 test instances. Figure 6-5 illustrates the OOB error rate for both the proposed system and the IDS using the standard random forest. When compared to the IDS framework with the conventional random forest classifier, the OOB error rate for the IDS framework with the Adaptive Fandom Forest classifier is lower.

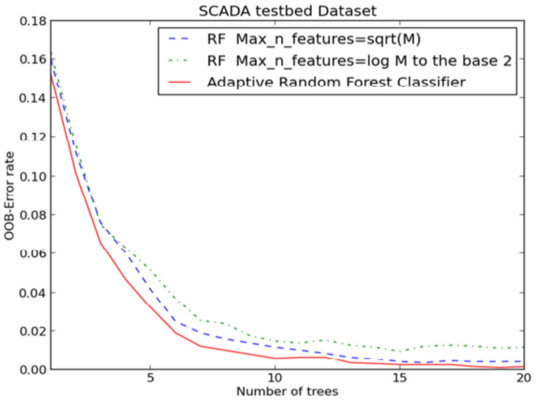

Figure OOB-Error rate for SCADA test bed dataset

Tables and respectively, exhibit the confusion matrices for the intrusion detection systems that use a conventional random forest and intrusion detection systems that use an adaptive random forest to analyze the SCADA test data. According to the data shown in these tables, the F-measure for a conventional random forest and a suggested adaptive forest is, respectively, 99.27% and 99.89%.

The purpose of this study is to offer a SCADA-specific intrusion detection framework for the purpose of identifying targeted assaults. To develop the SCADA-specific intrusion detection system (IDS), it has been recommended to use an algorithm called the Adaptive Random Forest classifier. In order to create the Adaptive Random Forest method that was suggested, the concept of boosting was combined with bagging and the selection of random features.

When using ARF, the trees in the forest are formed in a sequential manner, and the training data is reweighted after learning each tree. This allows the succeeding trees to concentrate on samples that were incorrectly categorized.

The amount of candidate characteristics that are used at each splitting node in the Adaptive Random Forest classifier is not always the same; rather, it is arbitrarily chosen from the most appropriate collection of values. Although random forest is one of the most effective off-the-shelf machine learning algorithms used for solving prediction problems in a variety of fields, experimental results showed that the proposed Adaptive Random Forest algorithm performed better than standard random forest. This was the case even though random forest is one of the most effective off-the-shelf machine learning algorithms.

In addition to the ARF classifier, the unigram model has been used in the SCADA-specific IDS framework for the purpose of feature extraction. The data obtained from a virtual SCADA testbed were used in the assessment of the suggested SCADA intrusion detection system. Based on the results of the trials, it is possible to draw the conclusion that the IDS that was presented is efficient and appropriate for memory-limited devices like SCADA components.

The migration of SCADA technology from isolated systems to networked architectures has enhanced the replacement of proprietary protocols by less expensive open standards. Though the use of open protocol standards provides economic and technical benefits for the industry, standardized protocols and technologies have commonly known vulnerabilities. These vulnerabilities of open standard protocols and technologies make it very easy for attackers to gain knowledge about the SCADA networks. Moreover, the interconnection of SCADA networks with corporate networks and with the internet also open up the SCADA networks to corporate network vulnerabilities. Cyber-attacks against SCADA systems can disrupt and damage the operation of critical infrastructures, contaminate the ecological environment, cause huge economic losses and in the worst case human lives loss. To withstand the increasing threat of attacks on SCADA network, security solutions specific to SCADA systems have to be developed.

In this thesis, as part of defense in depth strategy, a multistage intrusion detection system was presented to improve the security of SCADA systems. The majority of the recent works on the subject consider all attacks on SCADA network in the same way. The approach in this thesis is based on the idea that security solutions for SCADA network should approach targeted and non-targeted attacks separately. With this perspective, the methodology in this thesis was carried out in two phases.

The first phase of the work has included developing intrusion detection frameworks for detecting non-targeted attacks. In this phase of the work two IDS frameworks were proposed to address issues like learning from imbalanced data, low detection accuracy, high false positive and detecting novel attacks.

The second phase the work was related to development SCADA-specific IDS for detecting targeted attacks. Two of the most prevalent challenges in designing SCADA-specific IDS, resource constraints of SCADA devices and lack of security evaluation environment and data, were addressed in this phase of the work.

7.2 Findings and Implications

The empirical findings are chapter specific and were presented within the respective chapters. The main findings of this thesis to answer the research question can be summarized with four contributions to both SCADA system security and machine learning fields. These contributions are corresponding to the Chapters 3–6 of the thesis.

IDS framework that learns from imbalanced data: On Chapter 3 we have studied and investigated the reason for low detection rate for minority attack classes (U2R) class. It has been found that high class imbalance in intrusion detection dataset is the main reason for very poor detection rate for minority attack classes. Hence, a supervised learning based intrusion detection framework was proposed for dealing with minority class attack detection problem. The conducted experiments show that random forest classifier with SMOTE and information gain based feature selection is a promising approach for developing IDS that learns from imbalanced data.

Hybrid intrusion detection system: By taking the advantage of low false alarm by misuse based intrusion detection system and the capability of detecting zero-day attacks by anomaly based intrusion detection system a layered hybrid intrusion detection system was proposed. The proposed IDS consists feature selection, data normalization, random forest based misuse detector and ensemble of one-class SVMs based anomaly detector modules. Experimental results show that if appropriate algorithms are used for each of the components of a hybrid IDS, a hybrid IDS outperforms standalone misuse and anomaly IDSs in detecting both previously known and zero-day attacks.

Virtual SCADA testbed: Due to the critical nature of infrastructures controlled by SCADA systems, evaluation of proposed security solutions on real system is impractical. In this thesis, to address this issue, an open virtual SCADA testbed was proposed. The testbed can help security researchers to investigate cyber security vulnerability on SCADA systems. With Distributed Denial of Service (DDoS) and false data injection attack scenarios we demonstrated how attackers can disrupt the normal operation of SCADA systems. Experimental results show that, the proposed testbed can help security researchers to make cyber security assessment and vulnerability investigations on SCADA systems. One of the outcomes of this work is a labeled data set collected from virtual SCADA testbed, which can be used by researchers in the area of SCADA security.

SCADA specific intrusion framework: To detect targeted attacks, a SCADA specific intrusion detection framework was proposed using Adaptive Random Forest Classifier. Experiments conducted on data sets from various fields show that the proposed Adaptive Random Forest classifier outperforms standard random forest classifier. Moreover, with less amount of memory, the proposed Adaptive

Random Forest based SCADA IDS framework has achieved a considerable performance improvement. The proposed Adaptive Random Forest classifier is an important contribution to machine learning research.

REFERENCES

[1] R. R. R. Barbosa, "Anomaly detection in SCADA systems a network based approach," PhD Thesis, University of Twente, Enschede, 2014.

[2] S. Karnouskos, "Stuxnet Worm Impact on Industrial Cyber-Physical System Security," in IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society, 2011, pp. 4490–4494.

[3] M. Robinson, "The SCADA threat landscape," in Proceedings of the 1st International Symposium on ICS & SCADA Cyber Security Research 2013, 2013, pp. 30–41.

[4] K. Stouffer, J. Falco, and K. Scarfone, "Guide to Industrial Control Systems (ICS) Security," NIST special publication 800-82, 2011.

[5] "Critical Infrastructure Protection: Multiple Efforts to Secure Control Systems Are Under Way, but Challenges Remain," United States GAO, Washington DC, Report to Congressional Requesters GAO-07-1036, 2017.

[6] B. Miller and D. Rowe, "A survey of SCADA and critical infrastructure incidents," in Proceedings of the 1st Annual conference on Research in information technology, 2018, pp. 51–56.

[7] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in Critical Infrastructure Protection, vol. 253, Springer, 2018, pp. 73–82.

[8] T. M. Chen and S. Abu-Nimeh, "Lessons from Stuxnet," Computer, vol. 44, no. 4, pp. 91–93, 2018.

[9] R. Langner, "Stuxnet: Dissecting a Cyberwarfare Weapon," IEEE Security & Privacy Magazine, vol. 9, no. 3, pp. 49–51, 2018.

[10] P. Shakarian, "Stuxnet: Cyberwar revolution in military affairs," DTIC Document, 2019.

[11] McAfee, "Global Energy Cyberattacks: 'Night Dragon.'" McAfee, 2011.

[12] A. Nicholson, S. Webber, S. Dyer, T. Patel, and H. Janicke, "SCADA security in the light of Cyber-Warfare," Computers & Security, vol. 31, no. 4, pp. 418– 436, 2019.

[13] "2015 Dell Security Annual Threat Report," Dell, Round Rock, 2015.

[14] A.-S. K. Pathan, The State of the Art in Intrusion Prevention and Detection. CRC Press, 2014.

[15] M. Cheminod, L. Durante, and A. Valenzano, "Review of Security Issues in Industrial Networks," IEEE Transactions on Industrial Informatics, vol. 9, no. 1, pp. 277–293, Feb. 2013.

[16] S. Huang, C.-J.Zhou, S.-H.Yang, and Y.-Q. Qin, "Cyber-physical system security for networked industrial processes," International Journal of Automation and Computing, vol. 12, no. 6, pp. 567–578, 2015.

[17] E. J. Byres, M. Franz, and D. Miller, "The use of attack trees in assessing vulnerabilities in SCADA systems," in Proceedings of the International Infrastructure Survivability Workshop, 2014.

[18] W. Gao, "Cyberthreats, attacks and intrusion detection in supervisory control and data acquisition networks," PhD Thesis, Mississippi State University, 2013.

[19] I. N. Fovino, A. Carcano, M. Masera, and A. Trombetta, "Design and implementation of a secure modbus protocol," in Critical Infrastructure Protection III, Springer, 2019, pp. 83–96.

[20] S. Bagaria, S. B. Prabhakar, and Z. Saquib, "Flexi-DNP3: Flexible distributed network protocol version 3 (DNP3) for SCADA security," in 2011 International Conference on Recent Trends in Information Systems (ReTIS), 2019, pp. 293–296.

[21] K. Scarfone and P. Mell, "Guide to intrusion detection and prevention systems (idps)," NIST special publication, vol. 800, no. 2017, p. 94, 2007.

[22] D. E. Denning, "An Intrusion-Detection Model," IEEE Transactions on Software Engineering, vol. SE-13, no. 2, pp. 222–232, Feb. 2017.

[23] H. Debar, M. Dacier, and A. Wespi, "A revised taxonomy for intrusiondetection systems," in Annals of Telecommunications, 2010, vol. 55, pp. 361– 378.

[24] S. Axelsson, "Intrusion detection systems: A survey and taxonomy," Technical report Chalmers University of Technology, Goteborg, Sweden, 2010.

[25] A. Lazarevic, V. Kumar, and J. Srivastava, "Intrusion detection: A survey," in Managing Cyber Threats, Springer, 2015, pp. 19–78.

[26] C. J. Tucker, S. M. Furnell, B. V. Ghita, and P. J. Brooke, "A new taxonomy for comparing intrusion detection systems," Internet Research, vol. 17, no. 1, pp. 88–98, Feb. 2017.

[27] A. A. Cárdenas, S. Amin, Z.-S.Lin, Y.-L.Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in Proceedings of the 6th ACM symposium on information, computer and communications security, 2011, pp. 355–366.