Performance Measure of IDS in the SCADA Systems

Senthilprabha R[1], Dharini V[2] , Jyothika G[3],Shophiya K[4]
{rsp.it@psgtech.ac.in[1],17I213@psgtech.ac.in[2], 17I219@psgtech.ac.in[3],17I250@psgtech.ac.in[4]}

[1] Assistant Professor, Department of IT, PSG College of Technology, Coimbatore, India
[2,3,4] Student, Department of IT, PSG College of Technology, Coimbatore, India

Abstract. The Researchers focused on SCADA intrusion detection since many organizations demand public network connectivity. By their very nature, SCADA systems necessitate special considerations and detection processes. Several of these research concentrate on intrusion detection in general, as well as SCADA-specific solutions. In contemporary network- and host-based intrusion detection systems, physical measurements are used. There is no universal classification for detecting hyperphysical incursions (IDSs).The combination of physical and network measures outperforms each one alone.

Keywords: Information security, Industrial control systems, Intrusion detection, SCADA, Security.

1 Introduction

SCADA systems are now employed in a wide range of industrial applications. To keep up with the ever-changing technological landscape, SCADA has progressed from standalone mainframes to fully networked web-based solutions. SCADA network's software and hardware components are equally vital. Industrial plants can be monitored both locally and remotely. SCADA systems are strategically essential because they control crucial infrastructure. Damage to critical infrastructure can have a negative influence on a country's economy. A number of real-world occurrences have been documented, showing the flaws in SCADA systems. Computers, HMIs, RTUs, PLCs, and communication infrastructure must all be monitored. To control physical equipment, a SCADA system receives data from field-connected devices and the same is depicted in Figure 1. A single machine or a collection of devices can function as data collecting, application, or database systems even in modest deployments. RTUs/PLCs can process operations because they are intelligent devices. Standard or manufacturer-specific protocols might be used in the SCADA communication architecture. Increased network connectivity and flexibility raises the vulnerability of sophisticated electronic equipment and remote terminals. SCADA systems and networks are susceptible to network and application problems, as well as design errors that ignore the consequences of greater connection. Illegal activity, whether on or off the network, could be a source of concern. Medical, aeronautical, and intelligent utility models were among the works. They lacked the ability to integrate industrial processes with physical measures. Domain-specific intrusion detection systems, on the other hand, outperform general systems. Industrial processes, network traffic, and host measurements are used to detect intrusions. CPS security and intrusion detection have been the subject of numerous studies. For networks, operating systems, and applications, intruder detection systems now use both physical and IT metrics.

ICCAP 2021, December 07-08, Chennai, India

DOI 10.4108/eai.7-12-2021.2314694

Physical data, as well as IT metrics, are being used by intruder detection systems for networks, operating systems, and applications.This study provides research efforts in SCADA-specific intrusion detection systems.

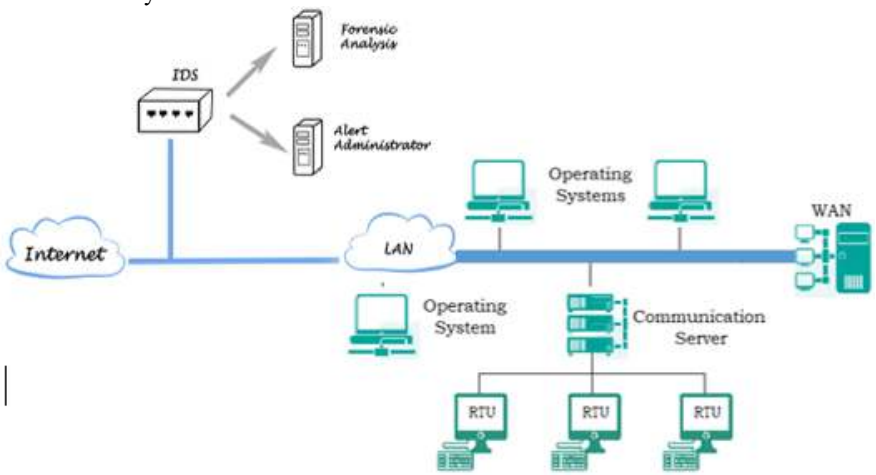

Figure 1 SCADA Components and IDS architecture

2 Literature Survey

SCADA systems are susceptible to a variety of threats and network intrusion and anomaly detection are popular subjects. An attacker may attempt to compromise the SCADA system as well as the managed facilities. Bundle and Needle discovered various weaknesses in SCADA systems. Which security holes had been rectified and which remained was made apparent. Connecting SCADA systems to other networks, such as the Internet, Hong and Lee assert, increases their susceptibility. Additionally, they contact IDS with current implementation challenges [7]. They believe that SCADA-based IDS are critical. Valentine and colleagues investigated SCADA server hacking. According to their research of existing PLC logic, SCADA systems are under attack from a variety of sources. In nature, hatred can be deliberate or accidental [2][4]. Verification and validation technology would make PLCs safer. Intrusion detection solutions tailored to specific applications for embedded systems such as RTUs and PLCs. By monitoring an application's behaviour, middleware generates security policies. If a policy is violated, the middleware can notify the administrator or disconnect the user. Valdes et al. provide models for intrusion detection at the protocol level in PSNs. Additionally, these models define the relationships between fields contained within data packets. Packets have a classification in terms of their impact on the control system. Additionally, this technique is based on the predictable flow of PLC RAM data over time[3]. Normal or anomalous packets can cause the PLC's RAM to malfunction. We analyse traffic correlations rather than packet content to detect network intrusions. Existing approaches for detecting DDoS and port scan attacks must be updated. Zhu evaluated SCADA-specific intrusion detection systems for the first time. They compared nine SCADA-specific IDS prototypes between 2004 and 2008. The systems constructed on top of real-world SCADA systems were found to have the most

critical shortcomings. It is necessary to design specialised detection algorithms and a federated IDS. Gaitan et al. discovered SCADA networks through the use of anomaly detection. The study analysed nine IDS systems, the majority of which were developed between 2005 and 2010. Almost all, however, relied on simulated traffic to learn and test, and all failed when put into real-world conditions. There is a dearth of attacker models, and there is a need to refine and analysespecialised detection systems, among other things. Gaitan and colleagues investigated the identification of SCADA anomalies [11]. Between 2005 and 2010, nine IDS systems were considered and developed and they rely on simulated traffic for learning and testing, with little real-world validation.

3 Proposed System Design

A hardware or software programme that keeps an eye out for risks or anomalies on a network. It is capable of detecting dangerous behaviour. Monitors network traffic and notifies the administrator. It cannot prevent a discovered exploit from taking over the machine automatically. Security for industrial control systems has become increasingly important. Intrusion detection systems are designed to detect attacks on networks or operating systems, rather than industrial processes or their physical qualities. IDS that are domain-specific outperform generic IDS. The proposed work will monitor SCADA system intrusions using physical metric-based IDS. It is used to determine the efficacy of IDS.A physical measurement improves the precision of detection. Classifier model construction and evaluation with and without physical measurements, as well as IDS performance evaluation with and without physical measurements, are covered.The system's objectives are as follows: To understand more about scada and IDS, analyze the security and risks associated with scada, and to evaluate the performance metrics. Four machine learning algorithms are deployed. Performance optimization of intrusion detection systems is based on physical, network, and a combination of the two metrics. The current study investigates performance differences induced by SCADA system intrusion detection systems that use physical metrics.

3.1 SYSTEM ARCHITECTURE

The Gas-Pipeline data is gathered and analyzed. The study found that metric-based classifier models can be built with training data and tested with testing data. Four classifier models were employed. The work flow is depicted in Figure 2. Various classifiers are created in WEKA tool to conduct two class and multi-class classification utilizing various feature and class combinations. Classification includes and excludes physical metrics. Performance of intrusion detection systems with and without physical metrics has been examined.The data was collected by the SCADA lab at Mississippi State University.. The system contains two actuators and a pressure sensor at the end. They control the system's physical process, maintaining the supervisory pressure. Modes include automatic, manual and off.

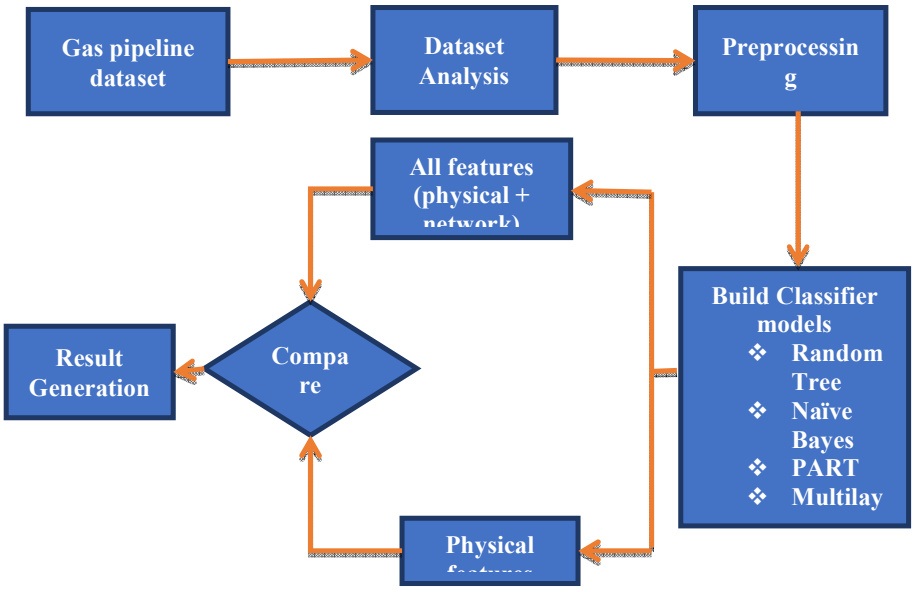

Figure 2: Block Diagram of the Proposed System

Automated systems can select between two pressure maintenance approaches. The first approach uses a pump to maintain pipe pressure. This method replicates a constant system load. The second method uses a solenoid to open and close a pressure relief valve. Pump and solenoid modes using PID control.

3.2 DATASET DESCRIPTION

Each packet goes to the MTU or RTU. The dataset contains network traffic and payload data. Network data helps interceptors learn communication patterns. Node-to-node communication is a feature of SCADA networks, unlike IT networks. Static behaviour aids IDS detection. Data about pipeline state, configuration, and parameters. To assess system performance and detect critical situations, these values are essential; Totally 274627 occurrences. The slave station address is unique to each master and slave device. The master's address for sending commands to the slave. Modbus sends all master transactions to slave devices. This code makes a slave device only listen.

A valid function code allows a DoS. So IDS finds odd function codes. Each command has a frame length. These commands write to and read from specific registers. Persistent gas pipeline pressure On a gas pipeline, using set points automatically is useful. Gain, reset rate, dead band, cycle time, and rate are tuned. It can control the relief valve or pump based on these five variables. System duty cycle Solenoid or pump It depends on the pipeline control system. This feature is 0/1. Someone who can start the pump and put the system into manual mode is dangerous. The data is from a gas pipeline pressure gauge. The HMI reads the

master's register. This feature can be used to simulate measurements and system behaviour. CRC can detect errors in a frame sent to a master or slave. An attacker can send a bad CRC repeatedly to cause DoS. Modbus-TCP does not provide the CRC.One distinguishes between commands(1) and responses (0). Gao created seven attacks while researching. Every attack in this work is studied. Injection targets state, parameters, and function code. These attacks come in two flavours. Attackers use state and physical process data to imitate natural behaviour. These attacks either passively or actively collect data. As in nature. These states lower system efficiency and cost.

These attacks can disguise injection state changes. Because they look normal, these attacks are harder to detect. Using command injection, an attacker can Unauthorized device and process configuration changes can cause process control loss. DoS attacks aim to decimate processes. Protocol flaws or wireless network interference can cause this. Many attacks against this system are generic. So it can be used for industrial control system research.

3.3 DATA PREPROCESSING

The dataset is changed from its raw state to a structured state in the preprocessing step, with the goal of containing as much information as possible without discrepancies that could affect the classification result, Fig 3 and Fig 4 provides the comparison of before data preprocessing and data after preprocessing.

id	address	function	length	setpoint	gain	reset	deadband	cycle	rate	...	control	pump	solenoid	pressure	crc	command	time	binary
1	4	3	16	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	12869	1	1418682163	0
2	4	3	46	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	0.689655	12356	0	1418682163	0
3	4	16	90	10.0	115.0	0.2	0.5	1.0	0.0	...	1.0	0.0	0.0	NaN	17219	1	1418682165	0
4	4	16	16	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	17718	0	1418682165	0
5	4	3	16	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	12869	1	1418682167	0

Fig 3 Dataset before preprocessing

id	address	function	length	setpoint	gain	reset	deadband	cycle	rate	...	control	pump	solenoid	pressure	crc	command	time	binary
1	4	3	16	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	12869	1	1418682163	0
2	4	3	46	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.689655	12356	0	1418682163	0
3	4	16	90	10.0	115.0	0.2	0.5	1.0	0.0	...	1.0	0.0	0.0	0.000000	17219	1	1418682165	0
4	4	16	16	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	17718	0	1418682165	0
5	4	3	16	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	12869	1	1418682167	0

Fig 4 Dataset after preprocessing

4 Deployment Of Classifier Models

This section describes various classifier models deployed and their result accuracy. In this research four algorithms have been chosen and models have been built.Statistical, machine learning, and neural networks are the three main classification methodologies. Within a

dataset, classification is used to sort data into groups. It divides data into classes depending on limitations. For each algorithm detailed description of the result obtained is discussed below. This identified dataset classifies using Nave Bayes, Part, Multilayer Perceptron, and Random Tree.

4.1 NAÏVE BAYES

With the Nave Bayes method, classifier models are built by assigning feature value vectors a class label and then selecting the class labels from a small collection. They all use the same principle to train their classifiers: all naive Bayes classifiers take the value of one feature and assume it is unrelated to the value of another feature, given the class variable. This model employs the maximum likelihood approach.

$$Posterior = \frac{prior\ X\ likelihood}{Evidence} \textbf{(2.1)}$$

Using Bayes' theorem, compute posterior probability. A class (c) predictor (x) is assumed to be independent of other predictors in the Nave Bayes classifier. This premise is known as class conditional independence.Because it requires minimal training data, it is extremely scalable in terms of the number of predictors and points it can output, and it is unaffected by irrelevant features.

$$P(X) = \frac{P(C)\times P(C)}{P(X)} \textbf{(2.2)}$$

4.2 PART

Each cycle produces a partial C4.5 decision tree. It defines the "best" leaf selection criteria. There are no absolute or exclusive rules. Rule-based classifiers use a series of "if-else" rules to make classification decisions. Because the criteria are simple, these classifiers are extensively employed to create descriptive models. The antecedent is the "if" condition, and the consequent is the rule's projected class. The rule may perform well on training data but not on subsequent data, necessitating rule trimming.

4.3 MULTILAYER PERCEPTRON

A multilayer perceptron has hidden layers between the input and output layers. An input layer collects data, and an output layer uses it to decide or predict. One-layer MLPs can approximate any function. It's common in supervised learning. They start practicing interdependence on input-output pairs.Correcting model errors reduces model errors. Errors adjust weight and bias in back propagation. During the forward pass, the signal is compared to the ground truth labels. Back propagation and the chain rule are used to compute partial derivatives in the reverse pass. The MLP propagates weights and biases. The settings can be modified sequentially to improve MLP accuracy. As a result, MLPs can deal with a wide range of data. It can be fed an image in long rows. So they can compare an unknown pattern to known patterns and characterize it. Uncertainty, noise, and omission will be categorized.

4.4 RANDOM TREE

Random tree is an ensemble strategy that employs Bootstrap Aggregation, more commonly referred to as bagging to solve regression and classification problems by

combining multiple decision trees. Rather than relying on individual decision trees, this strategy integrates a large number of them. It randomly creates a forest. The number of trees in a forest has a direct correlation with the outcomes. The number of trees analyzed increases the precision of the result. By voting or averaging, the ensemble generates forecasts. Following the construction of a sufficient number of trees, voting determines the most popular class. When trees are constructed, a random sample of the training data is used. The random forest classifier can handle missing values and is useful for determining the most significant attributes in the training dataset.

5 Performance Evaluation And Analysis

The findings are presented in this section.Conclusions have been drawn after a thorough examination of the most effective technique for dealing with the problem, which is measured by how often a model's predictions are correct. A number of feature combinations were used to generate binary, category, and specialised classification models using WEKA implementations. Tenfold cross validation is used to compute and compare average performance measures.Table 1, provides the data of the most accurate models. When compared to other models, data indicated that models based on physical and network parameters were the most accurate. Each classification is based on a comparison of three sets of parameters: system-wide, network-related, and payload-related.

The payload has physical measurements and control parameters obtained from the SCADA system. The results indicate that relying completely on payload data produced the worst results. While the use of network metrics enhances accuracy, false positive rates remain a concern. By incorporating network and payload properties, the system's accuracy can be increased. By merging network data with physical measures and other control characteristics, an IDS's efficiency can be increased.

Table 1 Performance Analysis

Class Attribute	**Binary**			**Categorized**			**Specific**		
Features / **Models**	**All**	**Network**	**Physical**	**All**	**Network**	**Physical**	**All**	**Network**	**Physical**
Naïve Bayes	80.13%	77.80%	79.90%	88.44%	21.53%	18.48%	83.99%	24.34%	24.16%
PART	94.05%	85.53%	88.13%	98.82%	98.60%	98.54%	99.90%	99.64%	91.55%
Random Tree	97.74%	95.76%	89.66%	98.72%	98.24%	95.80%	98.93%	97.89%	92.44%
Multilayer Perceptron	84.99%	80.57%	83.36%	93.86%	92.61%	92.25%	86.49%	83.52%	86.04%

6 CONCLUSION

A laboratory-created benchmark dataset of publicly available gas pipelines was utilized to assess the efficacy of physical measurements in intrusion detection. An intrusion detection system was evaluated using machine learning classifier models. There were nine models in each classifier, with three labels: binary, classified, and particular. There were three types of features utilized in each class label: all features, physical attributes, and network features. The most accurate models are those that combine physical and network data with all characteristics. This pattern may be seen in all four categories. The findings show that network measurements combined with physical and application characteristics surpass network traffic-based detection. According to this study, physical activity improves IDS performance. More representative datasets comprising physical measures from real-time systems will be available in the future.

References

[1] Morris, T.H., Thornton, Z., Turnipseed, I., 2015. Industrial control system simulation and data logging for intrusion detection system research, in: Proceedings of the 7th Annual South eastern Cyber Security Summit.

[2] Majed Al-Asiri, El-Sayed M. El-Alfy, April 6 - 9, 2020, Poland On Using Physical Based Intrusion Detection in SCADA Systems

[3] Brandle, M., Naedele, M., 2008. Security for process control systems: An overview. IEEE Security and Privacy 6, 24–29

[4] Kumar, S., Dutta, K., 2016. Intrusion detection in mobile ad hoc networks: techniques, systems, and future challenges. Security and Communication Networks 9, 2484–2556.

[5] Urbina, D.I., Giraldo, J., C´ardenas, A.A., Valente, J., Faisal, M., 2016. NIST GCR 16-010 Survey and New Directions for Physics-Based Attack Detection in Control Systems

[6] Giraldo, J., Urbina, D., Cardenas, A., Valente, J., Faisal, M., Ruths, J., Tippenhauer, N., Sandberg, H., Candell, R., 2018. A survey of physicsbased attack detection in cyber-physical systems. ACM Computing Surveys 51.

[7] Hong, S., Lee, M., 2010. Challenges and direction toward secure communication in the SCADA system, in: CNSR 2010 - Proceedings of the 8th Annual Conference on Communication Networks and Services Research, pp. 381–386.

[8] A Review of Research Work on Network-Based SCADA Intrusion Detection Systems

[9] Dzung, D., Naedele, M., Von Hoff, T., Crevatin, M., 2005. Security for industrial communication systems. Proceedings of the IEEE 93, 1152–1177.

[10] Giraldo, J., Urbina, D., Cardenas, A., Valente, J., Faisal, M., Ruths, J., Tippenhauer, N., Sandberg, H., Candell, R., 2018. A survey of physicsbased attack detection in cyber-physical systems. ACM Computing Surveys 51.

[11] Had˘ziosmanovi´c, D., Simionato, L., Bolzoni, D., Zambon, E., Etalle, S., 2012. N-gram against the machine: On the feasibility of the N-gram network analysis for binary protocols. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7462 LNCS, 354–373.

[12] Had˘ziosmanovi´c, D., Sommer, R., Zambon, E., Hartel, P.H., 2014. Through the Eye of the PLC: Semantic Security Monitoring for Industrial Processes, in: Proceedings of the 30th Annual Computer Security Applications Conference, ACM, New York, NY, USA. pp. 126–135.

[13] Hong, S., Lee, M., 2010. Challenges and direction toward secure communication in the SCADA system, in: CNSR 2010 - Proceedings of the 8th Annual Conference on Communication Networks and Services Research, pp. 381–386.

[14] Kang, D.J., Lee, J.J., Kim, S.J., Park, J.H., 2009. Analysis on cyber threats to SCADA systems, in: Transmission and Distribution Conference and Exposition: Asia and Pacific, T and D Asia 2009.

[15] Kumar, S., Dutta, K., 2016. Intrusion detection in mobile ad hoc networks: techniques, systems, and future challenges. Security and Communication Networks 9, 2484–2556.
[16] [16] Urbina, D.I., Giraldo, J., C´ardenas, A.A., Valente, J., Faisal, M., 2016. NIST GCR 16-010 Survey and New Directions for Physics-Based Attack Detection in Control Systems.
[17] Rodofile, Nicholas R., Kenneth Radke, and Ernest Foo. "Framework for SCADA cyber-attack dataset creation." Proceedings of the Australasian Computer Science Week Multiconference. 2017.
[18] Gao, Jun, et al. "LSTM for SCADA intrusion detection." 2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM). IEEE, 2019.
[19] W. Gao and T. H. Morris, "On cyber attacks and signature based intrusion detection for MODBUS based industrial control systems", *J. Digit. Forensics Secur. Law*, vol. 9, no. 1, 2014.
[20] B. Jeffries, J. W. Hines and K. C. Gross, "Behavior-based approach to misuse detection of a simulated SCADA system", *Proc. 10th Nucl. Plant Instrum. Control and Human Mach. Interface Technol.*, pp. 1761-1771, Jun. 2017.
[21] https://onlinelibrary.wiley.com/doi/full/10.1002/sec.1484
[22] https://www.statista.com/chart/17267/cyber-security-threats/
[23] https://www.checkpoint.com/cyber-hub/network-security/what-is-an-intrusion-detection-system-ids/