Smart Grid - A Network based Intrusion Detection System

R. Bala Sri Swetha, M.E, Biometrics and Cyber Security, PSG College of Technology, Coimbatore

ABSTRACT

Smart grid is nothing but the electrical power grid which is fully automated. It is the emerging technology in power grid that includes both Electrical Engineering and Information & Communication Engineering. As the Information and Communication Technology comes into the grid, it is subjected to both physical and cyber attacks. In the present three layer architecture of the smart grid network, each layer is subjected to cyber attack from the outsiders at different format. To provide effective security an Intrusion Detection System is must for the Network. In the proposed model the Intrusion Detection system is distributed among the three layers of the Network. If an attack is detected an alarm will be given at the corresponding layer. If a detection decision cannot be made at certain layer, then it is left to the upper layer which has a wider scope. This distributed network based Intrusion Detection system provides full protection against existing and future sophisticated security attacks.

Keywords

Intrusion Detection System, Smart Grid

1. INTRODUCTION

Smart Grids has been a solution for the development of Electric Power Grid which was not using any Communication development that happened last year. On merging communication network and the power electrical grid it is possible to gather more data about the consumers consumption and the stage of grid to enhance the current state of the grid and allows to use optimization technique to find the better quality and efficiency power generation, transmission and distribution technique^[2]. To achieve these goals Smart Grid should be the most secured network just like the security in common Communication Networks. This is a big challenge because Power Grid is an integral part of the society. Before implementing security frameworks in Smart Grids, it should be proposed deployed and tested^[5]. In the proposed system Signature based Intrusion Detection System is presented to avoid intrusion and the malfunctioning of a Smart Grid.

1.1 Layers of Smart Grid

Smart grid network architecture consists of three layers.

- Home area network
- Neighbourhood area network
- Wide area network

Home area network is a devoted network which is at the bottom layer of the Smart Grid. It is the network totally which connects all devices in the home. To monitor and control these appliances separate software applications can be deployed. In Smart Grid sector it is the most upcoming area to serve all the utilities and vendors. Home area network consist K. Goklia Meena, Assistant Professor, (S.G) Department of Information Technology, PSG College of Technology, Coimbatore

of smart meter and metering module.

Middle layer is Neighborhood Area Network (NAN) which connects the bottom layer through a range of small networks called community networks. This network transfers the data collected from all the surrounding devices which again sends these collected data to the next layer. It sends and receives data in duplex format.

Top layer is Wide Area Network which connects the main parts of the network like power plants, substations, and all other grid devices. This network connects a number of smart grid networks and forms an integrated smart grid network. Since it has an interconnected environment it can support various applications and solve their requirements. The applications include Supervisory Control And Data Acquisition (SCADA).

2. LITERATURE SURVEY

2.1 Security Measures

The other security techniques for smart grid are defending the grid against attacks launched by an adversary outside the grid. If the adversary attacks the grid through some compromised devices, these techniques will lose their effectiveness. Intrusion detection mechanisms are necessary for identifying attacked devices ^[4]. While existing network intrusion detection techniques may be applied to the smart grid network directly or with minor modification, below is a generic IDS framework and a DoS attack detection technique, both designed particularly for smart grid communications. Zhang et al. proposed a hierarchical IDS framework, where an IDS module is installed distributedly along the network hierarchy, that is, on control centers, community gateways and smart meters^[1]. The IDS module at the bottom layer accepts raw input from smart meters and the module at a higher layer accepts input only from the IDS module at the immediate lower layer. If an attack is detected by an IDS module, an alarm will be invoked at the corresponding layer. If a decision while detecting the intrusion cannot be made at certain layer. it will be left for the upper layer to make, since the upper layer has a wider scope of knowledge ^[6]. Each IDS module has two components: a classifier (for attack classification) and a recorder (for logging and accuracy evaluation). As in convention, either machine learning techniques or artificial immune systems may be applied for realizing the classifier. Zhang *et al.* suggested applying Support Vector Machine or clonal selection to build the classifier ^[4]. In either case, the classifier needs to be trained before put in use. Considering the difference of attacks likely happening at different layers, the training data will have a different concentration of attack types at each layer for a tradeoff of accuracy and time.

Authentication is performed between grid devices to ensure authenticity. It is a computation- intensive procedure generating noticeable delay and can become the target of DoS attack. Fadlullah et al. proposed a predication based DoS attack defense mechanism^[9]. They assumed that some compromised grid devices launch DoS attack distributedly by frequently sending false data or authentication requests along the network hierarchy. Unusual activities such as device failures and authentication failures are monitored at every layer and reported to the control center. The control center models each malicious event as a Gaussian process, realized by a collection of random variables representing event features such as number of defective devices, malicious authentication ratio, and so on ^[8]. It uses the collected reports as observations to form the prior beliefs of the Gaussian process. Using the prior beliefs and observations, it computes the posterior probability distributions of the Gaussian process through Gaussian process regression. The optimal parameters of the Gaussian process are obtained by maximizing the log likelihood of the training data with respect to the parameters. Then it is able to predict whether a DoS attack is going to happen, and to send early warning and instructions to the respective device so that the later can take appropriate actions (e.g. drop the data or authentication request) in advance to mitigate the forthcoming DoS attack.

3. PROBLEM STATEMENT

Smart grid is a compound system that is combination of various systems, networks and processes. It also includes a number of technologies like information technology and communication with electrical grid ^[3]. Smart grid has two infrastructures, power infrastructure and communication infrastructure. Communication infrastructure controls the power infrastructure. So success of smart grid depends on the security of the communication infrastructure.

3.1 Proposed Method

In the proposed method a network based distributed intrusion detection system can be used. Here the intrusion detection system going to be used is Signature based Intrusion detection system and the classifier used is SVM.

3.1.1 Signature Based IDS

A signature based IDS will monitor packets on the network and compare them against a database of signatures or attributes from known malicious threats. This is similar to the way most antivirus software detects malware. In the intrusion detection system detection engine is major portion of the system where the actual work happens. Network based detection engine processes a stream of time sequential TCP/IP packets to detect predetermined sequences and patterns. These patterns are known as signatures. The engines themselves are available in a variety of implementation like speed and configurability.

3.1.1.1 Network Signatures

Most network signatures are based on the contents of packets and are known as packet content signatures. Patterns are also detectable in the headers and in flow of the traffic, relieving the need to see into the packet. These signatures are known as traffic analysis signature.

3.1.2 Support Vector Machine

Support Vector Machine (SVM) is a powerful, state-of-the-art algorithm based on linear and nonlinear regression. Oracle Data Mining implements SVM for binary and multiclass classification. SVM has strong regularization properties. Regularization refers to the generalization of the model to new data. SVM is a kind of large-margin classifier. It is a vector space based machine learning method where the goal is to find a decision boundary between two classes that is maximally far from any point in the training data.

3.2 Objective

The proposed system is mainly to secure smart grid by placing IDS at each stage of the network hierarchy. The proposed system should detect all possible attacks like DOS attack, damaging the integrity of configuration, routing and communication traffic, illegitimate network operation and man-in-the-middle attack.

In the proposed system signature based Intrusion detection system is used and it is classified using SVM. A comparative study on SVM and other classifiers like decision tree, K nearest neighbour is also given.

3.3 Scope

A signature based IDS is based on the network signatures. These network signatures will examine the packet contents of each network packet so these are known as packet content signatures. These are compared to the available database of network signatures to find the threat or malware.

SVM is a classifier that is used to classify a particular value into one of the classes. It can be two class model or multi class model. Initially it was used for solving classification problem alone but latest researches provide way for it to solve regression problems also.

4. SYSTEM DESIGN

The system design of Smart grid security consist of creating a network, creating communication between nodes, acquiring dataset, conversion of dataset to readable format, data transfer through nodes building IDS.

As discussed in chapter 1, the security has to be provided in the three levels of Smart Grid such as,

- Communication between the residential networks (HAN)
- Communication between the community networks and bottom layer (NAN)
- Communication between the regional network and middle layer (WAN)

4.1 Architecture of Smart Grid



Fig No: 1 Architecture of the Smart Grid

4.1.1 Home Area Network (HAN)

A Home Area Network can be any individual user of Smart Grid. These networks consist of advanced meters called smart meters deployed in the Smart Grid architecture and Advanced Metering Infrastructure (AMI) for enabling automated twoway communication between the utility meter and the utility provider.

4.1.1.1 Smart Meter

A smart meter is usually an electronic device that records consumption of electric energy in intervals of an hour or less and communicates that information at least daily back to the utility for monitoring and billing purposes. Three-phase electricity meters are used in residential and commercial metering applications.

4.1.1.2 Metering Module

Metering module measures and transmits-fine-grained electric power usage information and information on the quality of electricity to the utility which can use this information for generating customer bills and also automatically control the consumption of electricity through delivery of load control messages to the smart meters.

4.1.2 Neighbourhood Area Network (NAN)

A NAN is a network that groups one or more HAN Networks and counts with interfaces to communicate with the higher layer in the Smart Grid. This is next level of metering and controlling network which collects metering and service information from the multiple HANs that is geographically near each other. Through a NAN, the utility provider is able to monitor how much power is being distributed to a particular neighbourhood by the corresponding distribution substation.

4.1.2.1 Central Access Control

The target of central access control is to ensure the sensor network services to be available to authorized users on time, even in presence of an internal or external attack. To reach this target, additional communication among nodes system may be adopted for successful delivery of every message to its recipient.

4.1.2.2 Smart Data Collector

Smart Data Collector is a communication gateway that coordinates communication within the NAN. It operates as the intermediary data concentrators, collecting and filtering data from groups of mesh-enabled meters, and economically sharing wide area network resources making communication more affordable while ensuring high performance.

4.1.3 Wide Area Network (WAN)

The WAN layer provides broadband wired and wireless communication between the NAN, substations, other distributed grid devices, and the utility. This layer should have similar characteristics to a backbone network, aggregating information from the users and transporting it to the control centers of the Smart Grid.

4.1.3.1 Energy Distribution System

Energy distribution system consists of a range of smaller-scale and modular devices designed to provide electricity, and sometimes also thermal energy, in locations close to consumers. They include fossil and renewable energy technologies, energy storage devices and combined heat and power systems.

4.1.3.2 SCADA Controller

Supervisory Control and Data Acquisition (SCADA) systems are basically Process Control Systems (PCS) that are used for

monitoring, gathering, and analyzing real-time environmental data from a simple office building or a complex nuclear power plant.

4.2 Architecture of the Intrusion Detection System



Fig 2 Architecture of Intrusion Detection System

4.2.1 Data Collection

Intrusion detection dataset is collected from the website http://nsl.cs.unb.ca/NSL-KDD/ . Dataset for both train and test database is obtained separately in text format. The Knowledge Discovery and Data (KDD99) intrusion detection datasets are based on the 1998 Defense Advanced Research Projects Agency (DARPA) initiative, which provides designers of intrusion detection systems (IDS) with a benchmark on which to evaluate different methodologies. To do so, a simulation is made of a factitious military network consisting of three 'target' machines running various operating systems and services. Additional three machines are then used to spoof different IP addresses to generate traffic. Finally, there is a sniffer that records all network traffic using the TCP dump format. The total simulated period is seven weeks. Normal connections are created to profile that expected in a military network and attacks fall into one of four categories: User to Root; Remote to Local; Denial of Service; and Probe.

- Denial of Service : Attacker tries to prevent legitimate users from using a service.
- Remote to Local : Attacker does not have an account on the victim machine, hence tries to gain access.
- User to Root : Attacker has local access to the victim machine and tries to gain super user privileges.

• Probe :Attacker tries to gain information about the target host.

4.2.2 Pre-processing

The content is received from the database text document has duration, protocol type, service provided, flag, source bytes, destination bytes, land, number of failed login, server count, serror rate, same server rate, different server rate, server different host rate and many more attributes. These attributes are separated and processed.

4.2.3 Capturing the packets

The protocols that are considered in KDD dataset are TCP, UDP, and ICMP that are explained below:

TCP stands for "Transmission Control Protocol". TCP is an important protocol of the Internet Protocol Suite at the Transport Layer which is the fourth layer of the Open System Interconnection (OSI) model. It is a reliable connectionoriented protocol which implies that data sent from one side is sure to reach the destination in the same order. TCP splits the data into labelled packets and sends them across the network. TCP is used for many protocols such as HTTP and Email Transfer.

UDP stands for "User Datagram Protocol". It is similar in behavior to TCP except that it is unreliable and connectionless protocol. As the data travels over unreliable media, the data may not reach in the same order, packets may be missing and duplication of packets is possible. This protocol is a transaction-oriented protocol which is useful in situations where delivery of data in certain time is more important than loosing few packets over the network. It is useful in situations where error checking and correction is possible in application level.

ICMP stands for "Internet Control Message Protocol". ICMP is basically used for communication between two connected computers. The main purpose of ICMP is to send messages over networked computers. The ICMP redirect the messages and it is used by routers to provide the up-to-date routing information to hosts, which initially have minimal routing information. When a host receives an ICMP redirect message, it will modify its routing table according to the message.

4.2.4 Categorising packets

The packets belonging to each protocol is categorised separately and all the attributes are checked for creating signatures.

4.2.5 Pattern matching

The test data obtained is matched with the signatures obtained from the train dataset values. These are done using the classifiers.

4.2.6 Different Intrusion Detection classifiers

In this design there is a comparative study on the classifiers mainly used for intrusion detection system normally. The classifiers are,

- Decision tree
- K nearest neighbour
- Support vector machine (SVM)

4.2.6.1 Decision Tree

The decision tree classifiers organized a series of test questions and conditions in a tree structure. Once the decision

tree has been constructed, classifying a test record is straightforward. Starting from the root node, apply the test condition to the record and follow the appropriate branch based on the outcome of the test. It then leads us either to another internal node, for which a new test condition is applied, or to a leaf node. When the leaf node is reached, the class label associated with the leaf node is then assigned to the record it traces the path in the decision tree to predict the class label of the test record, and the path terminates at a leaf node labeled NO.

4.2.6.2 K nearest Neighbour Classifier

In pattern recognition, the k-Nearest Neighbour algorithm (or k-NN for short) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour. In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbours.

k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms. Both for classification and regression, it can be useful to weight the contributions of the neighbours, so that the nearer neighbours contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbour a weight of 1/d, where *d* is the distance to the neighbour.

4.2.6.3 Support Vector Machine

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze patterns, data and recognize used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

5. IMPLEMENTATION

5.1 Hardware Requirements

The main hardware requirements are one processor with virtualized CPU RAM space of 4GB and disk space should be above 6GB. Regarding the software it needs a network simulator tool NS2, to run on Fedora 14 platform.

5.2 Simulation Methodologies

5.2.1 Creating a network

Network connection is created using NS2 simulator. The sample network is created with one component of each in a particular network.

5.2.2 Identifying the type of connection

After processing the dataset the type of protocol used is identified and categorised into separate tables.

5.2.3 Using train set to create signatures 20 % of KDD training dataset is used to create signatures separately for each and every protocol.

5.2.4 *Identification of attacks using Decision tree* The algorithm for decision tree classifier is mentioned below.

ID3(D, Attributes, Target)

Step 1: Create a node t for the tree.

Step 2: If all examples in D are positive, return the singlenode tree t with label "+".

Step 3: If all examples in D are negative, return the singlenode tree t, with label "–".

Step 4: Label t with the most common value of Target in D.

Step 5: If Attributes is empty, return the single-node tree t.

Otherwise:

Step 6: Let A* be the attribute from Attributes that best classifies examples in D. Assign t the decision attribute A*.

Step 7: For each possible value "a" in A* do:

Add a new tree branch below t, corresponding to the test $A^* = "a"$. Let D_a be the subset of D that has value "a" for A^* . If D_a is empty: Then add a leaf node with label of the most common value of Target in D. Else add the subtree ID3(D_a , Attributes \ { A^* }, Target). Step 8: Return t

5.2.5 Identification of attacks using K nearest neighbour

The algorithm of k-means clustering given by is written below:

Step 1: Take a space that contains K points and represent it as the objects that are being clustered.

These points refer to the centroids belonging to the initial group.

Step 2: Place the object to a group that has the closest centroid.

Step 3: After assigning the objects, recalculate the positions of the K centroids.

Step 4: Repeat Steps 2 and 3 until the centroids no longer move.

5.2.6 Identification of attacks using SVM

Machine learning is about learning structure from data. Here the dataset values are going to be classified based on SVM

The mapping: $X \rightarrow Y$,

where, $x \in X$ is input object that is to be classified.

 $y \in Y$ is a class name to which the objects are classified.

Input and output sets X, Y training set (x_1, y_1) (x_m, y_m) where, $x \in X$

y ∈ Y

Classifier: $y=f(x, \alpha)$

where, α is the parameter.

For example,

Example 1: A model of intrusion detection from a dataset, a set of destination bytes d_{n_i} then $f(x_i, \alpha) = \text{Euclid}(x_i, x_t)$ (x_i, y_i) is feasible if y=y' else it is not feasible.

6. RESULTS

The results obtained from the signature based intrusion detection system are discussed below:

6.1 Implementation of HAN

The Fig $\overline{6.1}$ shows simulated Home area network with single service and metering module.



Fig No: 3 HAN Network implementation

6.2 Implementation of NAN

The Fig 6.2 shows Negihbourhood area network with two home area networks with single Central access control and smart area controller.



Fig No: 4 NAN Network implementation

6.3 Implementation of WAN

The Fig 6.3 shows wide area network with two neighbourhood area networks each containing further two home area networks.





6.4 Implementation of Ideal Grid

The packet flow in the connected ideal grid is shown in Fig 6.4



Fig No: 6 Ideal Grid Network implementation

6.5 Signature Found In Each Protocol Category

The obtained input values of the dataset are categorized into three different protocols. The attacks in these protocols are classified and signature for each of these protocols is obtained and tabulated in Table 6.1.

Table 1 Protocol Category

Protocol type	Signature name
UDP	Normal, tear_drop, satan, nmap, rootkit
ТСР	Normal, Neptune, guess_password, land, portsweep, buffer_overflow, pihf, warezmaster, ipsweep, multihop, warezclient, perl, back, ftp_wire,

	loadmodule, satan, spy, imap, rootkit
ICMP	Normal, portsweep, ipsweep, smurf, satan, pod

6.6 Signatures Found for each Attack using Train Data Base

The signatures obtained are classified based on each attack and are listed below in the Table 6.2.

Denial of service attacks	User to root attacks	Remote to local attacks	Probes
Apache2	Anypw	Dictionary	Insidesniffer
Arpposion	Casesen	Ftpwrite	Ipsweep
Back	Eject	Guest	Is_domain
Crashiis	Ffbconfig	Httptunnel	Mscan
Dosnuke	Fdformat	Imap	Ntinfoscan
Land	Loadmodu l	Named	Nmap
Mailbomb	Ntfsdos	Ncftp	Queso
SYN Flood	Perl	Netbus	Resetscan
Ping of Death	Ps	Netcat	Saint
Process table	Sechole	Phf	Satan
Selfping	Xterm	Ppmacro	
Smurf	Yaga	Sendmail	
Sshprocesstable		Sshtrojan	
Syslogd		Xlock	
Tcprest		Xsnoop	
Teardrop			
Udpstrom			

Table 2 Attack Category

6.7 Protocol Vs Attack Category

The training dataset has a maximum of Smurf and Neptune attacks respectively and the normal records accounted to 19% of the total records. Smurf has the maximum frequency with 57% of all the attacks; Neptune is the second most frequent with 21% of the total attacks. The percentage of attacks according to their protocol obtained using SVM classifier is given in the Table 6.3. The Fig 6.5 gives the graphical representation of the percentage of attacks versus their protocol.

Table 3 Protocol	Vs	Attack	Category
------------------	----	--------	----------

	ТСР	UDP	ICMP
DOS	50.42	34	62.05
U2R	9.03	7	0
R2L	11.42	0	0
PROBE	29.32	60	39.01



Fig 7 Protocol Vs Attacks

6.8 Classifier Based Results

The results shown in Table 6.3 are from SVM classifier. SVM classifier is the best classifier among the three classifiers. Their detection rate, FRR and FAR is given in the Table 6.4.

6.8.1 Detection rate of each classifier

The detection rate, false rejection rate and false acceptance rate of each classifier is given below.

 Table 4 Classifier based results

	DECISION TREE (%)	K NN (%)	SVM (%)
Detection rate	95.5	88.9	97.5
FRR	1.2	4.1	2.9
FAR	3.4	5.2	1.3



Fig 8 Detection rate of different classifiers



Fig 9 False rejection and acceptance rate of each classifier

7. CONCLUSION AND FUTURE WORK

Smart Grid is the revolutionized version of the present electric power grid that includes communication and information technologies. Smart Grid enables both the vendors and the customers to transfer and manage energy usage data through the sensor networks. But the inclusion of sensor network will create new security challenges relating to the data theft, data integrity and connectivity. This creates a great disaster to both provider and the consumer.

This project proposes a novel method for providing cyber security to the Smart Grid by placing a network based intrusion detection system in distributed arena. It provides architecture for IDS in Smart Grid which is more reliable, dynamic and considers the real time nature of traffic for each component in Smart Grid. The results indicate that SVM is the best classifier that can be used for intrusion detection system of Smart Grid. It detects at a rate of 97.5% with low FRR of 2.9% and low FAR of 1.3%.

However this system is helpful only in detecting the intrusions. In future this project can be enhanced for intrusion prevention and also for forensic analysis in Smart Grid.

8. REFERENCES

[1] A.H. Mohsenian Rad and A. Leon Garcia, "Distributed

Internet-Based Load Altering Attacks Against Smart Power Grids," IEEE Transaction Smart Grid, Volume: 2, No. 4, pages: 667–74, 2011.

- [2] Agustin Zaballos, Alex Vallejo, and Josep M. Selga, "Heterogeneous Communication Architecture for the Smart Grid", IEEE Network, 2011.
- [3] Ata Arvani and Vittal S. Rao, "Detection and Protection against Intrusions on Smart Grid Systems", IJCSDF Volume: 3, No. 1, pages: 38-48, 2012.
- [4] Daojing He, Chun chen, Jiajun Bu, Sammy Chan, Yan Zhang, Mohsen Guizani, "Secure Service Provision in Smart Grid Communication", IEEE Communication magazine, 2012.
- [5] Emmanouil Vasilomanolakis, Mathias Fischer, Max Muhlhauser, "Collaborative Intrusion Detection in Smart Energy Grids", 1st International Symposium for ICS and SCADA Cyber Security Research 2013.
- [6] Xu Li, Inria Lille, Xiaohui Liang, Rongxing Lu, Xemin Shen, Xiadong Lin, Haojin zhu, "Securing Smart Grid: Cyber Attacks, Countermeasures, and challenges", IEEE Communication magazine, 2012.
- [7] Y. Yan et al., "A Survey on Cyber Security for Smart Grid Communications," IEEE Communication Surveys & Tutorials, Volume 14, No. 4, pages: 127-153, 2012.
- [8] Yichi Zhang, Lingfeng Wang, Weiqing Sun, Robert C. Green II and Mansoor Alam, "Distributed Intrusion Detection System in a Multi- Layered Network Architecture of Smart Grids", IEEE Transaction on Smart Grid, Volume: 2, No.4, pages: 52-75, 2011.
- [9] Z. M. Fadlullah *et al.*, "An Early Warning System against Malicious Activities for Smart Grid Communications," IEEE Network, Volume: 25, No. 5, pages: 50–55, 2011.
- [10] Y. Mo *et al.*, "Cyber-Physical Security of A Smart Grid Infrastructure," Proceedings IEEE, Volume: 100, No. 1, pages: 195–209, 2012.