International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 4, Number 7 (2014), pp. 747-752

http://www. irphouse.com

FPGA Based Distributed Network Intrusion Detection in Smart Grids Using Naives Bayes Classifier

N.Purnima[1] and Omprakash P[2]

[1,2]Electronics and Communication Engineering,
Velammal College of Engineering and Technology,
Madurai-625009, TamilNadu, India.

ABSTRACT

The communication architecture has to be reliable, scalable and extendable to future smart grid services and applications. Most of these changes will occur as an Internet-like communications network is superimposed on top of the Smart Grid using wireless mesh network technologies with the 802.15.4, 802.11, and Wimax standards. Each of these will expose the Smart Grid to cyber security threats. In order to address this issue, this work proposes a distributed network intrusion detection system for smart grids (SGDIDS) by developing and deploying an intelligent module in multiple layers of the smart grid. The Wireless networks in communication environment will be exposed to many threats, so that SGDIDS will find attacks using Naives Bayes Classifier. Naives Bayes Algorithm is trained using data that is relevant to their level and also improve detection. This paper proposes a FPGA based network intrusion detection in communication network of Smart Grid to detect and classify malicious data and possible cyber attacks.

Keywords- Smart Grids; Wimax; Naives Bayes Classifier; Cyber security attacks; FPGA.

1. INTRODUCTION

When the legacy power infrastructure is augmented by a communication infrastructure, it becomes a smart grid. This additional communication infrastructure facilitates the exchange of state and control information among different components of the power infrastructure. As a result, the power grid can operate more reliably and efficiently. Although deploying the smart grid enjoys enormous social, environmental and technical benefits, the incorporation of information and communication technologies into the power infrastructure will introduce many security challenges.

For example, it is estimated that the data to be collected by the smart grid will be an order of magnitude more than that of existing electrical power systems. This increase in data collection can possibly introduce security and privacy risks. Moreover, the smart grid will be collecting new types of information that were not recorded in the past, and this can lead to more privacy issues. All the essential parts of the smart grid will be its communication networks.

2. SMART GRID

A smart grid is a modernized electric grid that uses information and communication technology to gather and act on information, such as information about the behaviours of suppliers and consumers, in an automated fashion to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity. A Smart Meter is usually an electric meter 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. Smart meters enable two-way communication between the meter and the central system. Unlike home energy monitors, smart meters can gather data for remote reporting. Such an advanced metering infrastructure (AMI) differs from traditional Automatic Meter Reading (AMR) in that it enables two-way communications with the meter. The Communication network are supported by Wireless mesh network technology. So these networks are exposed to many cyber security attacks. In order to reduce those attacks we are proposing a method to find Smart Grid Intrusion Detection using FPGA.

3. INTRUSION DETECTION USING NAÏVE BAYES CLASSIFIER

In simple terms, a naive Bayes classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. An advantage of Naive Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

Naïve Bayes Classifier is satisfies by the variables in Smart Grid dataset and degree of class overlapping is small (i.e. potential linear decision boundary). Naïve Bayes classifier would be expected to achieve good results in smart Grid. For some dataset, with optimization using wrapper feature selection, NBC may defeat other classifier. And it is used in SGDIDS because of its high speed. The Naïve Bayes theorem proposes to find the class whether a normal packets or attacked packets with the numbers of data given then compute the probability of each class, and pick the most likely class. Bayes theorem proposes the equation in more tractable form.

Naïve Bayes theorem is simply,

$$P(C|D) = \frac{P(D|C) \cdot P(C)}{P(D)}$$

In terms of SGDIDS,

$$P(class|data_n) = \frac{P(data_n|class).P(class)}{P(data_n)}$$

This equation is simplified by removing $P(data_n)$ to rank $P(class|data_n)$ for each value of class. Since $P(data_n)$ will be the same every time and it does not depend on class. The equation is simplified as,

$$P(class|data_n) \propto P(data_n|class).P(class)$$

The prior probabilities, $P(class)$, can be calculated as described which shows whether it is an attack packets or normal packets..

That leaves$P(data_n|class)$. We want to eliminate the massive, probably very sparse, joint probability$P(data_1, data_2, \ldots\ldots, data_n|class)$. If each data are independent, then

$$P(data_1, data_2, \ldots\ldots, data_n|class) = \prod_i P(data_i|class)$$

Even if they are not actually independent, we can assume they are (that's the "naïve" part of naïve Bayes). The probability values of the data are found by the type of class i.e. may be a normal class or affected class (attacked class) according number of data in the problem.

In our original example, the features are continuous. In that case, you need to find some way of assigning $P(data_n = value|class)$ for each class. You might consider fitting then to a known probability distribution (e.g., a Gaussian). During training, you would find the mean and variance for each class along each feature dimension. To classify a point, we have to find

$$P(data_n = value|class)$$

by plugging in the appropriate mean and variance for each class. Other distributions might be more appropriate, depending on the particulars of the data, but a Gaussian would be a decent starting point.

4. TRAINING THE CLASSIFIER

According to the Naïve Bayes the probability values are found according to the data in the particular problem. The probability values vary for different types of the problem. The maximum amounts of data are known for each substation of the Smart Grids. So in Naïve Bayes training set we are fixing the values according to the substation. The values of the data differ for each substation. After finding the values for each substation we are given those values as the training data set to the training algorithm of the Naïve Bayes Classifier. The incoming packets may belong to any type of class i.e. it may be a attacked packet or may be a normal incoming packets. Form the training dataset the Bayes classifier will test the incoming packets. If the incoming packets are same as the training dataset then they are assumed as normal

incoming packets. So that source address and destination address should be same. If the incoming packets are different from the training dataset then they are assumed as attacked incoming packets. So that source address and destination address are different.

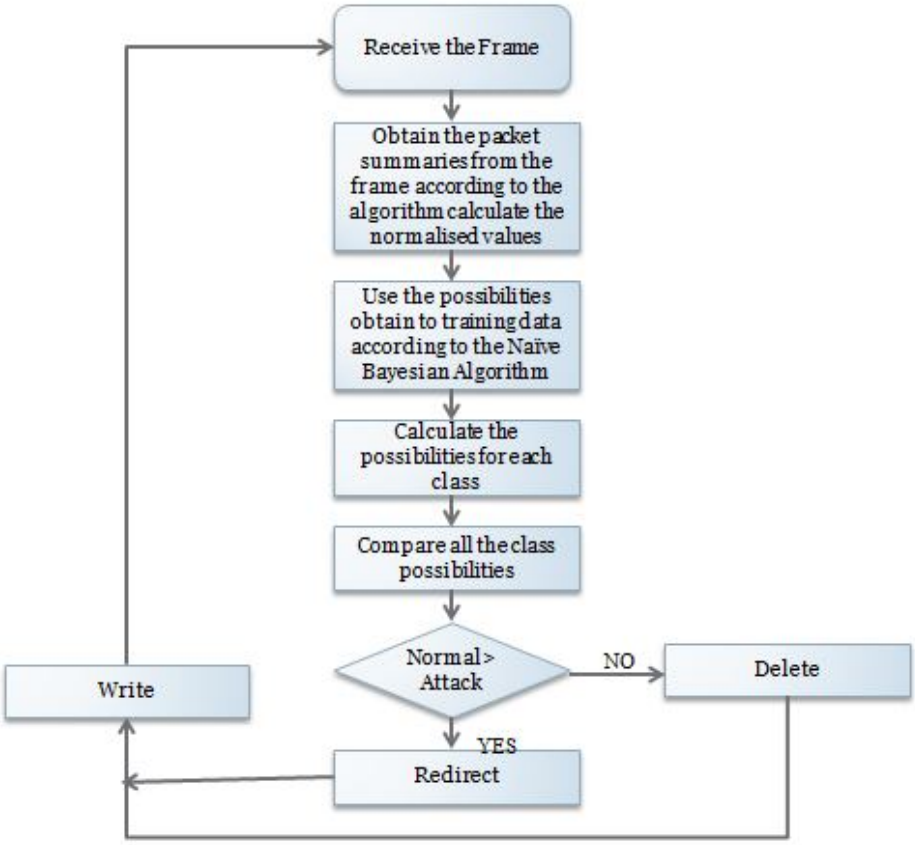

Figure 1. *Flow Scheme of the IDS using Naïve Bayes Algorithm*

Finally the attacked packets are found by this probability classification of the data according to the Naïve Bayes Classifier which differ for each substation in the Smart Grids.

5. VHDL BASED NETWORK INTRUSION DETECTION

Smart Grid Communication Network IEEE 802.16 standard in which Smart Grid Network Intrusion Detection System is modelled and implemented using the VHDL Very high speed integrated circuit- Hardware Description Language. Evaluation of the SGDIDS (Smart Grid Network Intrusion Detection System) is done through simulation. Simulation is done for the testing dataset of Naives Bayes Classifier. Smart Grids communication networks uses many protocols in the Transport layer such as Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) of the Internet Protocol Suite are commonly categorized as layer-45 protocols within

OSI. Smart Grid uses Internet Protocol version 4 (IPv4) in this paper. Network based Intrusion Detection in Smart Grids identifies threats in data which are supported by many protocols of OSI. In our paper for example we took the protocols such as Transmission Control Protocol and IPv4 in which packet normalization is done first to train the dataset using Naïve Bayes Classifier.

4.1 Normal Packets

The output waveform obtained in the Questasim for the normal packets is represented in the figure 2. Once the testing starts, the testing dataset will compare the packets with the trained dataset. If the incoming packet matches the packets of the trained dataset of the Naïve Bayes Algorithm, then output will display as normal packets.

4.2 Attacked Packets

These are the waveform for the attacked packets. Once the testing starts, the testing dataset will compare the packets with the trained dataset. If the incoming packet does not matches the packets of the trained dataset of the Naïve Bayes Algorithm, then output will display as attacked packets. Output waveform for the attacked packets is obtained in Questasim software of VHDL and which is represented in below figure 3.

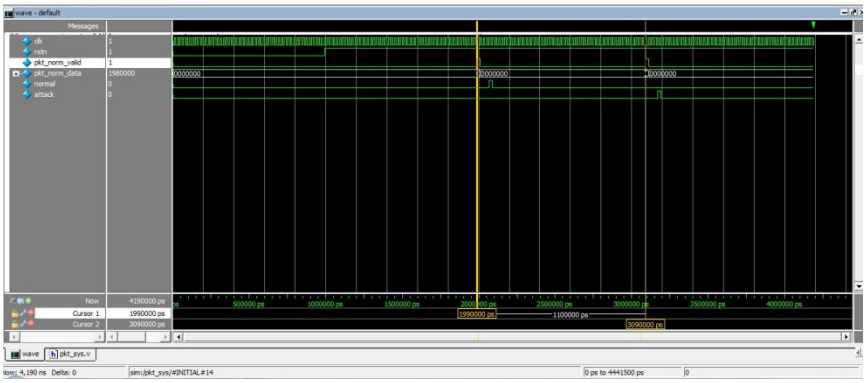

Figure 2. *Representations of Normal Packets*

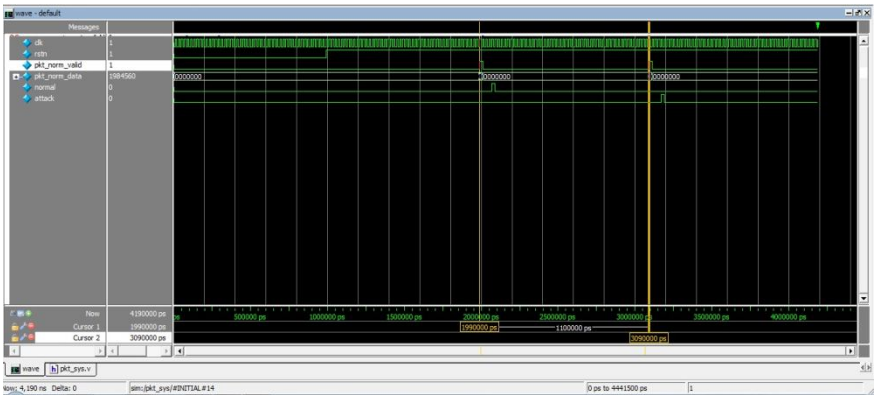

Figure 3. *Representation of attacked packets*

6. CONCLUSION

The Wireless mesh networks in communication environments of Smart Grids are secured from many threats, since SGDIDS found the attacks using Naives Bayes Classifier. The percentage of the attacked packets is found by this probability classification of the data according to the Naïve Bayes Classifier which differ for each substation in the Smart Grids. Hardware behaviour will directly reflect because of using VHDL based Smart Grids Network intrusion Detection (SGDIDS). The FPGA based intrusion detection produces faster and accurate method for finding the attacked packets in the Wimax communication network of Smart Grid.

7. REFERENCES

[1] S. X. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," *Appl. Soft Comput.*, vol. 10, no. 1, pp. 1–35, 2010.

[2] [Mentor] Mentor Graphics Company

[3] Ruiz-Llata, G. Guarnizo, and M. Y?benes- Calvino. FPGA implementation of a support vector machine for classification and regression. *WCCI 2010 IEEE World Congress on Computational Intelligence, IJCNN*, july 2010

[4] [IEC61850] International Electrotechnical Comission (IEC) 61850.J. Cho, B. Benson, S. Cheamanukul, and R. Kastner. Increased performance of FPGA-based color classification system. *Annual IEEE Symposium on Field-Programmable Custom Computing Machines*, pages 29–32, 2010.M.

[5] D. Kim and J. Park, "Network based intrusion detection with support vector machines," in *Proc. ICOIN*, 2003, vol. 2662, Lecture Notes in Computer Science, pp. 747–756.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*,vol. 20, pp. 273–297, 1995 [Online].Available: http://www.springerlink.com/index/K238JX04HM87J80G.pdf

[7] S. R. Gunn, "Support vector machines for classification and regression,"Faculty Eng., Sci., Math. School Electron. Comput. Sci., Tech.Rep., May1998[Online].Available:http://pubs.rsc.org/en/Content/ArticlePDF/2010/AN/B918972F/ 2009-12-23