International Journal of Computer Networks and Communications Security

VOL. 2, NO. 5, MAY 2014, 168–172
Available online at: www.ijcncs.org
ISSN 2308-9830

A Fuzzy Model for Network Intrusion Detection

S.Sethuramalingam[1] and Dr.E.R.Naganathan[2]

[1] Associate Professor and Head, Department of CS, Aditanar College, Tiruchendur

[2] Professor and Head, Department of CSE, Hindustan Univesity, Chennai

E-mail: [1]seesay@rediffmail.com, [2]ern_jo@yahoo.com

ABSTRACT

The network intrusion becomes ever growing problem. The complexity present in the collected network data set is absence of clear boundary between anomaly connection and normal connection. However fuzzy logic can well address this problem. In earlier works, combining fuzzy logic and data mining to develop fuzzy rules are explored to address this problem. In this paper, a new fuzzy model is developed to detect anomaly connections. The developed model is tested with NSLKDD data set. The model gives better result.

Keywords: *Network intrusion ,anomaly detection, fuzzy model, 10-fold cross validation.*

1 INTRODUCTION

As defined in [1], intrusion detection is "the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions, defined as attempts to compromise the confidentiality, integrity, availability, or to bypass the security mechanisms of a computer or network".

In a computer network, there are two main intrusion detection systems - Anomaly intrusion detection system and misuse intrusion detection system. The first one is based on the profiles of normal behaviour of users or applications and checks whether the system is being used in a different manner. The second one collects attack signatures, compares behaviour with the collected attack signatures and signals intrusion when there is a match [2].

System with characteristics such as impreciseness, vagueness and ambiguity make the system more complex. If these characteristics can be represented correctly then the understanding of complexity will be less. Fuzzy logic is a useful tool to represent the ambiguity present in the data set. On other hand the network intrusion data set is ambiguous and does not have clear boundary between anomaly and normal connections. In this work, fuzzy logic is proposed to represent the soft boundary present in the data set. The fuzzy rule based system is designed to address this problem [1 3 4]. Earlier the fuzzy rules are created from the knowledge of domain expert.

Today, with more and more computers getting connected to public accessible networks (e.g., the Internet), it is impossible for any computer system to be claimed immune to network intrusions. Since there is no perfect solution to prevent intrusions from happening, it is very important to be able to detect them at the first moment of occurrence and take actions to minimize the possible damage. Before data mining techniques are introduced into this field, intrusion detection was heavily dependent on a manually maintained knowledge base which contained signatures of all known attacks. Features of monitored network traffic were extracted and then compared with these attack signatures. Whenever a match was found, an intrusion was claimed to be detected and it was reported to the system administrator. Due to the difficulty and expense to manually maintain the knowledge base to reflect the ever changing situations, it was not feasible to continue working in this traditional way. Therefore now systems are developed to learn from the collected data. The remaining part of paper is organized as section 2 presents Related work section 3 presents the Proposed algorithm, section 4

discusses Experiments and results and section 5 discusses conclusion

2 RELATED WORK

In [5] authors collected profile of the network and constructed a intrusion detection algorithm. In [6] the authors proposed to develop dynamic fuzzy boundary from labeled data. In this work, a dynamic fuzzy boundary is proposed to detect anomaly connections. The boundary is developed using SVM and Fuzzy Logic. In [7] authors proposed hybrid model based on fuzzy logic and neural network. In [8] authors developed an algorithm using Artificial Neural Networks with fuzzy clustering. In[9][10] the Mamdani fuzzy model is applied for zooming function for digital camera and Permeability detection of skin respectively. In [11] authors discussed about designing fuzzy controller to detect anomaly based connections. However exploring fuzzy model to detect network intrusion is very few. In this paper, a fuzzy model is proposed to detect anomaly.

3 PROPOSED ALGORITHM

The new version of KDD data set NSL-KDD is publicly available for researchers through the website [12] [13] . Although, the data set still suffers from some of the problems discussed [14] and may not be a perfect representative of existing real networks, because of the lack of public data sets for network-based IDSs, the authors believe that it still can be applied as an effective benchmark data set to help researchers compare different intrusion detection methods.

The NSL-KDD [13] data set has 41 conditional attributes and one decision attribute. The value of the decision attribute is either anomaly or normal. The features service ,flags ,src_bytes, dst_bytes and dst_host_serror_rate i.e. 5 attributes out of the 41 attributes are used in the algorithm. A collection of numeric data is standardized by subtracting a measure of central location such as mean and divided by some measure of spread such as standard deviation [14].

3.1 Fuzzy Model

The figure 1 shows the Proposed fuzzy model which consists of

i. a fuzzifier (encoder);
ii. an inference engine (processor); and
iii. a defuzzifier (decoder).

i. Fuzzifier

A fuzzifier has the function of converting (or encoding) input categorical or numeric data (crisp values) into fuzzy values. Because these values propagate through a model and ultimately determine the output, fuzzification is the most crucial procedure in fuzzy modeling. Fuzzification of input data always relates to a fuzzy proposition and is carried out by means of a membership function which can be derived either from a priori knowledge of a system or by using input data

Let the connection record contains n attributes. $A_1,A_2,A_3,\ldots\ldots A_n$ be fuzzy set for anomaly class attributes 1,2 … and n respectively. $N_1,N_2,N_3,\ldots\ldots N_n$ be fuzzy set for normal class attributes 1,2 … and n respectively. The membership values for the attributes are $\mu_{A1}(x_{i1})$, $\mu_{A2}(x_{i2})$, $\ldots\mu_{Aj}(x_{ij})\ldots\ldots\ldots$ $\mu_{An}(x_{mn})$ for anomaly class. The membership values for the attributes are $\mu_{N1}(x_{i1})$, $\mu_{N2}(x_{i2})$, $\ldots..\mu_{Nj}(xij)\ldots\ldots\ldots$ $\mu_{Nn}(x_{mn})$ for normal class. For each attribute fuzzy membership function is defined as

$\mu_{A_j}(x)$ =gaussmf(x_{ij},[trn_amean,trn_astd]) where trn_amean and trn_astd are the mean and standard deviation of anomaly class respectively.

$\mu_{N_j}(x)$=gaussmf(x_{ij},[trn_nmean,trn_nstd]) where trn_nmean and trn_nstd are the mean and standard deviation of normal class respectively.

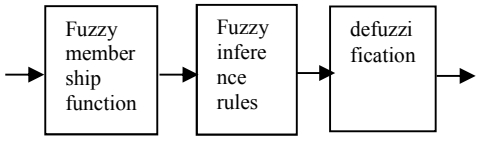

Figure 1: Architecture of a typical fuzzy model.

ii. Inference engine

An inference engine is the mind of a fuzzy model. Its function is to filter out informational noise and create a synthesized fuzzy set from the individual fuzzy sets transmitted by the fuzzifier. In the proposed model, The product of membership value for the i^{th} record for anomaly connection is computed by the following equation

$$ay_i = \prod_{j=1}^{n} \mu_{A_j}(x_{ij})$$

The product of membership value for the i^{th} record for normal connection is computed by the following equation

$$ny_i = \prod_{j=1}^{n} \mu_{N_j}(x_{ij})$$

now these values are mapped to the following function

$$f(x) \leftarrow (x-x')/\sigma$$

where x' and σ are mean and standard deviation respectively from this equation the value of input to the function is computed

$x = \sigma * f(x) + x$

using the equation the value for x is computed for anomaly connection as well as for normal connection.

iii. Defuzzifier

A defuzzifier transforms the synthesized fuzzy set back to a crisp set, which expresses the result of modeling. It can be a mathematical function or a subjectively- or objectively-defined threshold fuzzy value. Hellendoorn and Thomas (1993) describe a number of criteria that an ideal defuzzification procedure should satisfy. The most important criterion is that a small change in inputs of a fuzzy model should not cause a significant change in output. These x_i values are computed and compared and the given record is assigned to a class which has maximum x_i value.

Algorithm fuzzy_compos(trn_amean, trn_astd, trn_nmean, trn_nstd, tstdataset)

Tstdataset: testing data set has m records and n attributes

Trn_amean: mean of anomaly class records in the training data set

Trn_astd: standard deviation of anomaly class records for the training data set

Trn_nmean: mean of normal class records in the testing data set

Trn_nstd: standard deviation of normal class records in the testing data set

```
for each connection record in the testing data set
 py1←1
 py2←2
    for each attribute in the connection record
        y(i,j)←gausmf(x(i,j),[trn_amean,trn_astd])
        y1(i,j)←gausmf(x(i,j),[trn_nmean,trn_nstd])
        py1←py1*y(i,j)
        py2←py2*y1(I,j)
   end
        by1(i)←py1;
        by2(i)←py2;
  end
   for each connection record in the testing data set
   f1(i) ←(by1(i)*trn_astd)+trn_amean;
      f2(i) ←(by2(i)*trn_nstd)+trn_nmean;
     if (f1(i)>f2(i))
         if  i <= anolimit
           tp←tp+1;
        else
          fn←fn+1;
          end
          if  i> anolimit
            tn←tn+1;
          else
            fp←fp+1;
          end
       end
    end
```

A confusion matrix as shown in the Table 1 is typically used to evaluate the performance of the algorithm.

Table. 1. Standard metrics for evaluation of intrusions

Confusion Matrix (standard metrics)		Predicted connection label	
		Normal	Intrusion (Anomaly)
Actual connection Label	normal	True Negative (TN)	False Alarm (FP)
	Intrusion (anomaly)	False Negative (FN)	Correctly detected (TP)

From Table 1, recall and precision may be defined as follows

$$Precision = TP/(TP+FP)$$

$$Recall = TP/(TP+FN)$$

4 RESULTS AND DISCUSSION

In the proposed work, there are only two fuzzy values (f1(i) and f2(i)) are used to detect whether the given connection belongs to anomaly or normal for the training data set, the value of the parameters of fuzzy membership functions, mean and standard deviation are computed. Using these parameters, the fuzzy membership for each attribute in the connection record is computed for testing data set. For each connection record there are two set of member functions are assigned corresponding to the two classes. Product of membership value for each record is computed by multiplying their attribute membership value. These product values are mapped to a output function. Now the values of output function are compared. The given connection record is assigned a class for which the function has maximum value.

In 10-fold cross-validation, the original sample is randomly partitioned into 10 subsamples. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data. The cross-validation process is then repeated 10 times (the *folds*), with each of the 10 subsamples

used exactly once as the validation data. The 10 results from the folds then can be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once [15].

The algorithm is compared with another algorithm that uses correlation to match testing data with training data set in the inference process. The class which has maximum correlation is assigned to that record. Table 2 and Table 3 show the result of the algorithms respectively.

Table. 2. Algorithm using mapping to the function

Run	anomaly	normal	total	fp	fn	tp	tn	Precision = tp/(tp+fp)	Recall= tp/(tp+fn)
1	470	529	999	40	80	390	489	0.9069	0.8297
2	470	529	999	47	103	367	482	0.8864	0.7808
3	470	529	999	55	92	378	474	0.8729	0.8042
4	470	529	999	52	88	382	477	0.8808	0.8127
5	470	529	999	50	96	374	479	0.8820	0.7957
6	470	529	999	48	97	373	481	0.8859	0.7936
7	470	529	999	42	83	387	487	0.9020	0.8234
8	470	529	999	48	81	389	481	0.8901	0.8276
9	470	529	999	60	85	385	469	0.8651	0.8191
10	470	529	999	43	92	378	486	0.8978	0.8042
	average							0.8868	0.8091

Table. 3. Algorithm using mapping to the correlation

Run	anomaly	normal	total	fp	fn	tp	tn	Precision = tp/(tp+fp)	Recall = tp/(tp+fn)
1	470	529	999	36	74	396	493	0.9166	0.8425
2	470	529	999	57	102	368	472	0.8658	0.7829
3	470	529	999	55	83	387	474	0.8755	0.8234
4	470	529	999	57	86	384	472	0.8707	0.8170
5	470	529	999	51	93	377	478	0.8808	0.8021
6	470	529	999	56	88	382	473	0.8721	0.8127
7	470	529	999	48	81	389	481	0.8901	0.8276
8	470	529	999	68	72	398	461	0.8540	0.8468
9	470	529	999	64	77	393	465	0.8599	0.8361
10	470	529	999	44	96	374	485	0.8947	0.7957
	average							0.8777	0.8187

5 CONCLUSION

In this paper, a new fuzzy model is proposed to detect anomaly connection. A second algorithm is developed based on correlation in the inference stage and to test the proposed algorithm. The precision and recall values are almost same. It infers that the proposed algorithm works in expected manner. The impact of the algorithm is studied with different membership function for the input as well as mapping the output.

6 REFERENCES

[1] Wang, L. and J.M. Mendel, "Generating Fuzzy Rules from Numerical Data, with Applications", Technical Report USC-SIPI-169, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089, 1991.

[2] Fayyad.U.M, Piatetsky-Shapiro.G, and Smyth.P, "From data mining to knowledge discovery in databases". AI Magazine, vol. 17, no. 3, pp. 37.54, 1996

[3] Kosko, B. "Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence", Prentice Hall, Englewood Cliffs, NJ, 1992.

[4] [4]. Sudkamp. T. and R.J. Hammell II, "Interpolation,Completion, and Learning FLKZY Rules", IEEE Transactions on Systems, Man, and Cybernetics, 1994, 24, 2, pp.332-342.

[5] John E. Dickerson and Julie A. Dickerson "Fuzzy Network Profiling for Intrusion Detection", Electrical and Computer Engineering Department Iowa State University Ames, Iowa, 50011

[6] J.T. Yao S.L. Zhao L. V. Saxton "A study on fuzzy intrusion detection" ,Department of Computer Science, University of Regina, Regina Saskatchewan, Canada S4S 0A2

[7] Muna Mhammad T. Jawhar, Monica Mehrotra "Design Network Intrusion Detection System using hybrid Fuzzy-Neural Network", Faculty of Natural Science Department of computer science Jamia Millia Islamia ,New Delhi, 110025, India

[8] Gang Wang , Jinxing Hao b, Jian Mab, Lihua Huang "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering", School of Management, Fudan University, Shanghai 200433, PR China and Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

[9] I. Elamvazuthi, P. Vasant "The Application of Mamdani Fuzzy Model for Auto Zoom Function of a Digital Camera" Universitiy Technologi PETRONASTronoh, Malaysia J.Webb University of Technology Swinnburne, Sarawak Campus,Kuching, Sarawak, Malaysia

[10] Deepak R. Keshwani, David D. Jones,George E. Meyer and Rhonda M. Brand "IL Rule-based Mamdani-type fuzzy modeling of skin Permeability",Feinberg School of Medicine, Evanston, Biological Systems Engineering 1-1-2008

[11] Farzaneh Geramiraz, Amir Saman Memaripour, and Maghsoud Abbaspour "Adaptive Anomaly-Based Intrusion Detection System Using Fuzzy Controller" Computer Engineering Department, Faculty of Electrical and Computer Engineering, Shahid Beheshti University, G. C.,Evin, Tehran, Iran. International Journal of Network Security, Vol.14, No.6, PP.352-361, Nov. 2012

[12] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani . "A Detailed Analysis of the KDD CUP 99 Data Set Proceeding of the IEEE Symposium on Computational Intelligence in Security and Defence Applications" (CISDA 09)

[13] http://nsl.cs.unb.ca/NSL-KDD/

[14] Hai Jin Jianhua Sun, Han Chen, Zongfen Han "A Fuzzy Data Mining Based Intrusion Detection Model", Cluster and Grid Computing Lab. Huazhong University of Science and Technolory, Wuhan 430074 China.. Proc. Of the 10th IEEE International Workshop on Feature Trends of Distributed Computing Systems 2004 IEEE

[15] Picard, Richard; Cook, Dennis (1984). "Cross-Validation of Regression Models". Journal of the American Statistical Association 79 (387): 575–583.