IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. IX (Mar-Apr. 2014), PP 99-103
www.iosrjournals.org

Dendritic Cell Algorithm and Dempster Belief Theory Based Approach for Intrusion Detection System

Mr. Ved Prakash Sharma1, Dr. Rajdev Tiwari2, Mr. Rajesh KumarVarun3
[1]*M.Tech Student, Noida Institute of Engineering and Technology, Greater Noida, UP, India*
[2]*Director of MCA, Deptt. , Noida Institute of Engineering and Technology, Gr. Noida, UP, India.*
[3]*Asst. Pro.of IT Dept. Northern India Engineering College, Shastripartk, New Delhi, India.*

Abstract: *Everyone knows the services of the Internet. Internet services contain large scale of data. Security is one of the important issues during transmission of data over the network. To avoid misuse of data an efficient Network Intrusion Detection Systems (NIDS) is needed. To predict intrusion over the network a combined features of Dendritic Cell Algorithm (DCA) and Dempster Belief Theory (DBT) has been used called hybrid model of DCA and DBT. This algorithmic approach comes under the Artificial Immune Detection System. The presented system is about to analyze the network statistics under different parameters and identify the False Negative Rate and the Uncertainty Parameters. Based on these parameters analysis, the intrusion will be identified. The work is about to decrease the False Positive and False Negative rates so that the more effective intrusion detection will be performed.*

Keywords: *Dendritic Cell Algorithm, Dempster Belief Theory, Network Intrusion Detection System, KDD, False positive and False Negative.*

I. Introduction

Intrusion prevention measures, such as encryption and authentication, can be used in ad-hoc networks to reduce intrusions, but cannot eliminate them. For example, encryption and authentication cannot defend against compromised mobile nodes, which often carry the private keys. Integrity validation using redundant information (from different nodes), such as those being used in secure routing, also relies on the trustworthiness of other nodes, which could likewise be a weak link for sophisticated attacks. To secure mobile computing applications, we need to deploy intrusion detection and response techniques, and further research is necessary to adapt these techniques to the new environment, from their original applications in fixed wired network. Intrusion Detection attempts to detect computer attacks by examining various data records observed through processes on the same network. These attacks are split into two categories, host-based attacks [1, 2 ,3] and network-based attacks [4, 5,6]. Host-based attacks target a machine and try to gain access to privileged services or resources on that machine. Host-based detection usually uses routines that obtain system call data from an audit-process which tracks all system calls made on behalf of each user. Network-based attacks make it difficult for legitimate users to access various network services by purposely occupying or sabotaging network resources and services. This can be done by sending large amounts of network traffic, exploiting well known faults in networking services, overloading network hosts, etc. Network-based attack detection uses network traffic data (i.e., tcpdump) to look at traffic addressed to the machines being monitored. Intrusion detection systems are split into two groups, anomaly detection systems and misuse detection systems. Anomaly detection is the attempt to identify malicious traffic based on deviations from established normal network traffic patterns [7, 8]. Misuse detection is the ability to identify intrusions based on a known pattern for the malicious activity [9, 10]. These known patterns are referred to as signatures. Anomaly detection is capable of catching new attacks. However, new legitimate behavior can also be falsely identified as an attack, resulting in a false positive. The DCA is a population-based algorithm, designed for handling anomaly-based detection tasks. It is inspired by functions of natural DCs of the innate immune system, which form part of the body's first line of defense against invaders. DCs have the ability to combine a multitude of molecular information and to interpret this information for the T-cells of the adaptive immune system, to induce appropriate immune responses towards perceived threats. Therefore, DCs can be seen as detectors for different policing sites of the body as well as mediators for inducing a variety of immune responses [12].The Dempster-Belief theory (DBT) is a mathematical theory of evidence. It allows one to combine evidence from different sources and arrive at a degree of belief (represented by a belief function) that takes into account all the available evidence. The Dempster-Shafer theory, also known as the theory of belief functions, is a generalization of the Bayesian theory of subjective probability. Whereas the Bayesian theory requires probabilities for each question of interest, belief functions allow us to base degrees of belief for one question on probabilities for a related question. These degrees of belief may or may not have the mathematical properties of probabilities; how much they differ from probabilities will depend on how closely the two questions are related [13].

II. Kdd Data Set Description

KDD^{-21} data set can be used for building a network detector, a predictive model capable of distinguishing between intrusions and normal connections. The simulated attacks fall in one of the following four categories:

- Denial of Service Attack (DoS): is an attack in which the attacker makes some computing or memory Resource too busy or too full to handle legitimate requests, or denies legitimate users access to a Machine.
- User to Root Attack (U2R): is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system.
- Remote to Local Attack (R2L): occurs when an attacker who has the ability to send packets to A machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine.
- Probing Attack: is an attempt to gather information about a network of computers for the apparent purpose of circumventing security controls.

III. Related Work

Significant amount of work on intrusion detection has been carried by many groups. In 1980, the concept of intrusion detection began with Anderson’s paper [14]; he introduced a threat classification model that develops a security monitoring surveillance system based on detecting anomalies in user behavior. In 1986, Dr. Denning proposed several models for commercial IDS development based on statistics, Markov chains, time-series, etc [15]. In the early 1980’s, Stanford Research Institute (SRI) developed an Intrusion Detection Expert System(IDES) that monitors user behavior and detects suspicious events [16]. In 1988, a statistical anomaly-based IDS was proposed by Haystack [17], which used both user and group-based anomaly detection strategies. In 1996, Forrest et al. proposed an analogy between the human immune system and intrusion detection that involved analyzing a program’s system call sequences to build a normal profile [18]. In 2000, Valdes et al. [19] developed an anomaly based IDS that employed naïve Bayesian network to perform intrusion detecting on traffic bursts. In 2003, Kruegel et al. [20] proposed a multisensory fusion approach using Bayesian classifier for classification and suppression of false alarms that the outputs of different IDS sensors were aggregated to produce single alarm. In the same year, Shyu et al. [21] proposed an anomaly based intrusion detection scheme using principal components analysis (PCA), where PCA was applied to reduce the dimensionality of the audit data and arrive at a classifier that is a function of the principal components. In 2003, Yeung et al. [22] proposed an anomaly based intrusion detection using hidden Markov models that computes the sample likelihood of an observed sequence using the forward or backward algorithm for identifying anomalous. Lee et al. [23] proposed classification based anomaly detection using inductive rules to characterize sequences occurring in normal data. In 2000, Dickerson at al. [24] developed the Fuzzy Intrusion Recognition Engine (FIRE) using fuzzy logic that process the network data and generate fuzzy sets for every observed feature and then the fuzzy sets are used to detect network attacks. In 2003, Ramadas et al. [25] presented the anomalous network traffic detection with self organizing maps using DNS and HTTP services that the neurons are trained with normal network traffic then real time network data is fed to the trained neurons, if the distance of the incoming network traffic is more than a preset threshold then it raises an alarm. In 2004, Yi Hu proposed a data mining approach for detecting malicious transactions in a Database System. The experiment illustrates that the proposed method works effectively for detecting malicious transactions provided certain data dependencies exist in the database. In 2010, Rajeshwar Katipally proposed a data mining techniques to find the patterns of generated alerts by generating Association rules. A data mining techniques automatically discover patterns in multistage attack, visualize patterns, and predict intrusions. In 2011, Chung-Ming Ou proposed an immunity-based intrusion detection system. This IDS tries to determine some malicious attacks from system calls directly. In 2012, Kola Sujatha. P, proposed the incremental training algorithms for the network intrusion detection. In this work an improved incremental SVM algorithm, which uses the RBF kernel function has been developed, to select the positive kernel. A fuzzification is carried out to classify the attributes into sub attributes that enhances the rule creation. In 2102.Song Yuan proposed MMDCA which highlights the nature of of each antigen through the multiplier. Experiments show that the algorithm presented has considerable detection accuracy and stable detection performance.

IV. Proposed Resarch Methodology

In this presented work combined features of DCA and DBT as a model has been proposed to perform the intrusion detection on the network dataset. The presented model is defined in the following steps as follows:

Step 1. To process on the proposed mining operation we need some authenticated dataset so that the analysis can be performed. The analysis will be performed respective to the detection ratio. The work is about to obtain the high recognition ratio from the system. The KDD99 dataset is now the benchmark for training, testing and evaluating learning IDSs , so it is basic for IDS developers.

Step2. From original database is converted to binary format by admin. Extension database is created by extending these binary format data. Finally original binary format and extended data are integrated to form integrated database. Fuzzy logic will be used for selecting items to identify the most appropriate attributes that can be used identify the intrusion over the network.

Step3. DCA is basically a filtration algorithm that will perform the feature based analysis on dataset. It performs the grouping of communicating data and will group them respective to the anomaly coefficient analysis. It is defined as the prediction model where the probabilistic and analytical decision will be taken regarding the group creation.

Step 4. At the final stage DBT will be implemented to perform the final classification of error based on evidence theory. It will identify the evidence under different parameters and by collecting all these evidences, take the intelligent probabilistic decision about the intrusion.

4.1 Building Dendritic Cell Algorithms (DCA):

```
Algorithm()
{
    Define an empty DC pool with no cell defined internally.
    L=Length(AvailableDataItems)
    For i=1 to L
    {
        Select CurData=AvailableDataItems(i)
        If (Class(CurData)=Normal)
        {
            Obtain the Data Element Characteristics and Update the input signal concentration elements such as
                a. Mean value
                b. Standard Deviation
                c. Cell size
                d. Bare Nuclei
                e. Normal Nucleoli
        }
        Obtain the Difference Value between the Average and Current signal conceration values
        OutputCytokines=AverageCenentrationValues-CurrentDataConcentrationValue
        Perform Analysis on OutputCytokines to obtain the danger signal concentration
        Obtain SafeLimit=Median(DataSetConcentrationValues)
        If (OutputCytokines>SafeLimit)
        {
            Migrate Signal to Safe Signal Dataset
        }
        Else If(OutputCytokines<SafeLimit And OutputCytokines<PAMPLimit)
        {
            Migrate Signal to PAMP Signal Dataset
        }
        Else
        {
            Migrate Signal to Danger Signal Dataset
        }
        Obtain MeanDangerSignal from Dervied OutputCytokines
        Assign the weightage to Safe, PAMP and danger signal based on DC maturation table
        Obtain the OutputConcentration Values for csm, semi and mat signals
        If semi>mature
        {
            Set AntigenContext=Semi
        }
        Else
        {
            Set AntigenContext=Mature
        }
        For i=1 to length(AntigenContext)
{
    Identify the number of appearances as antigen as semi or mature
    If semi>mature
    {
        Set Antigen=Benign
    }
    Else
    {
        Set Antigen=Malignant
    }
}
}
```

4.2 Dempster Belief Theory (DBT):

- **Step l**: Choose a node ordering. Note that node ordering will make a difference in the topology of the generated tree. An optimal node ordering with respect to the Dempster Belief Theory iS NP-hard to find.
- **Step2:** Loop through the nodes in the ordering. For each node Xi, create a set Si of all its neighbors. Delete the node Xi from the moralized graph.
- **Step3:** Build a graph by letting each Si be a node. Connect the nodes with weighted undirected edges. The weight of an edge going from Si to Sj is |Si ƯSj||
- **Step4:** Let Dempster Belief Theory be the maximal-weight spanning tree of the cluster graph.

In this presented work DCA and DBT based approach is presented to predict the intrusion over the network. The presented work is a probabilistic model in which different kind of attacks over the network will be identified and intrusion detection will be performed.

In this prediction model, a weighted analysis is been performed on the dataset attribute and based on the dendritc cell algorithm the initial training of data is performed. Respective to this training the network over the nodes is constructed. The probabilistic relationships between the attributes are identified. The dataset is defined with a set of attributes called X=(X1,X2…Xn). Each attribute is defined with some discrete value represented by Val(X). When the training algorithm is applied on this attribute set, some weighted value is identified for each attribute. Based the weighted values, the relationship between the attributes is identified. This relationship and probabilistic weighted values collective helped to generate a graph over the attribute set. The dendritic c cell algorithm also deals with the attributes using the conditional probability analysis. When one cell attribute is compared with the outside values conditionally. The cell attributes are considered as the independent attributes and outside cell attributes are dependent attributes. Based on this relationship, the conditional probability is estimated for the dataset. The conditional probability between two attributes I and j is given by P(Xi/Xj). Once all the attributes are defined with probabilistic attributes, then the particular instance value for all attributes is given by Product(P(Xi/Xj), P(Xj/Xi)) where i >=1 and i<=n, j>=1 and j<=n Some examples of probabilistic decision is listed as under P(X1=attack)=0.2,P(X1=not)=0.8.

V. Result And Discussion

In order to evaluate the performance of proposed algorithm for network intrusion detection we get following result.

Table1: Obtained Result

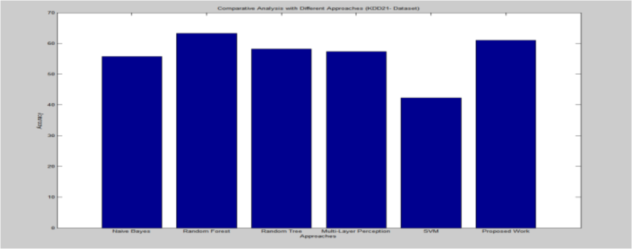
Class	Total Instances	Detected Successfully	Wrong Detected	Accuracy Ratio
Anomaly	11850	7235	4615	61.05485232
Normal	9698	5333	4365	54.99071974

In this section we discuss effective implementation of combined features of DCA and DBT. This proposed framework is done with KDD^{-21} set which is considered as a benchmark data set for intrusion detection system. in this paper since most of the anomaly detection systems work with binary labels, i.e., anomalous and normal, rather than identifying the detailed information of the attacks.

Table 2: Result comparison on KDD^{-21} data set

Technology	Naive Bayes	Random Forest	Random Tree	Multi/Layer Perception	Support vector machine	Purposed Technology
Accuracy rate	53.97	55.77	63.26	58.51	42.29	61.50

As can be seen in Table 2, the accuracy rate of my proposed is relatively high as compare to Naive Bayes, Random Forest, Random Tree, Multi/Layer Perception and Support Vector Machine. Graphical representation of the above table shown as below:



VI. Conclusion And Future Work

In this paper a new frame work has been proposed with the help of DCA and DBT for network based intrusion detection system. Our results show that the proposed method has maximum accuracy rate as compare to other methods to increase the rate of detection of intrusion and give better result on KDD-21 data set comparatively because we have initially perform high level of filtration to select an attribute after which we have analyzed their features using DCA based low level of filtration and finally used DBT based evidence theory for probability based decision for network intrusion detection system.

Reference

[1]. T.Subbulakshmi," Detection of DDoS Attacks using Enhanced Support Vector Machines with Real Time Generated Dataset", IEEE-ICoAC 2011 978-1-4673-0671-3/11©2011 IEEE.

[2]. Marinova-Boncheva," Applying a Data Mining Method for Intrusion Detection", International Conference on Computer Systems and Technologies - CompSysTech'07.

[3]. Neelam Sharma," Layered Approach for Intrusion Detection Using Naive Bayes Classifier", ICACCI'12, August 3-5, 2012, Chennai, T Nadu, India. ACM 978-1-4503-1196-0/12/08.

[4]. C.I. Ezeife," NeuDetect: A Neural Network Data Mining Wireless Network Intrusion Detection System", IDEAS10 2010, August 16-18, Montreal, QC [Canada]; Editor: Bipin C. DESAI; ACM 978-1-60558-900-8/10/08.

[5]. Wenke Lee," Mining in a Data-flow Environment: Experience in Network Intrusion Detection", KDD-99 San Diego CA USA 1999 l-581 13-143-7/99/08.

[6]. LTC Bruce D. Caulkins," A Dynamic Data Mining Technique for Intrusion Detection Systems".

[7]. K C Nalavade," Intrusion Prevention Systems: Data Mining Approach", International Conference and Workshop on Emerging Trends in Technology (ICWET 2010) – TCET, Mumbai, India ICWET'10, February 26–27, 2010, Mumbai, Maharashtra, India. ACM 978-1-60558-812-4.

[8]. C.I. Ezeife," WIDS: A Sensor-Based Online Mining Wireless Intrusion Detection System", ACM 978-1-60558-188-0/08/09.

[9]. Klaus Julisch," Mining Intrusion Detection Alarms for Actionable Knowledge", SIGKDD '02 Edmonton, Alberta, Canada ACM 1-58113-567-X/02/0007.

[10]. Guanhua Yan," Towards a Bayesian Network Game Framework for Evaluating DDoS Attacks and Defense", CCS'12, October 16–18, 2012, Raleigh, North Carolina, USA ACM 978-1-4503-1651-4/12/10.

[11]. Neelam Sharma,"Layered Approach for Intrusion Detection Using Naive Bayes Classifier", ICACCI'12, August 3-5, 2012, Chennai, T Nadu, India. ACM 978-1-4503-1196-0/12/08.

[12]. Al-Hammadi, Y., Aickelin, U., & Greensmith, J. (2008). DCA for Bot Detection.

[13]. G. Shafer, a Mathematical Theory of Evidence, Princeton, University Press, Princeton, Shafer, A Mathematical Theory of Evidence, Princeton, University Press, Princeton, NJ, 1976.

[14]. James P. Anderson, "Computer security threat monitoring and surveillance," Technical Report 98-17, James P. Anderson Co., Fort Washington, Pennsylvania, USA, April 1980.

[15]. Dorothy E. Denning, "An intrusion detection model," IEEE Transaction on Software Engineering, SE-13(2), 1987, pp. 222-232.

[16]. Dorothy E. Denning, and P.G. Neumann "Requirement and model for IDES- A real-timeintrusion detection system," Computer Science Laboratory, SRI International, Menlo Park, CA 94025-3493, Technical Report # 83F83-01-00, 1985.

[17]. S.E. Smaha, and Haystack, "An intrusion detection system," in Proc. of the IEEE Fourth Aerospace Computer Security Applications Conference, Orlando, FL, 1988, pp. 37-44.

[18]. S. Forrest, S.A. Hofmeyr, A. Somayaji, T.A. Longstaff, "A sense of self for Unix processes,"Proc. of the IEEE Symposium on Research in Security and Privacy, Oakland, CA, USA, 1996, pp. 120-128.

[19]. A. Valdes, K. Skinner, "Adaptive model-based monitoring for cyber attack detection," in Recent Advances in Intrusion Detection Toulouse, France, 2000, pp. 80-92.

[20]. C. Kruegel, D. Mutz, W. Robertson, F. Valeur, "Bayesian event classification for intrusion detection," in Proc. of the 19th Annual Computer Security Applications Conference, Las Vegas, NV, 2003.

[21]. M.L. Shyu, S.C. Chen, K. Sarinnapakorn, L. Chang, "A novel anomaly detection scheme based on principal component classifier," in Proc. of the IEEE Foundations and New Directions of Data Mining Workshop, Melbourne, FL, USA, 2003, pp. 172-179.

[22]. D. Y. Yeung, and Y. X. Ding, "Host-based intrusion detection using dynamic and static behavioral models," *Pattern Recognition*, 36, 2003, pp. 229-243.

[23]. W. Lee, S.J. Stolfo, "Data mining approaches for intrusion detection," In Proc. of the 7th USENIX Security Symposium (SECURITY-98), Berkeley, CA, USA, 1998, pp. 79-94.

[24]. J.E. Dickerson, J.A. Dickerson, "Fuzzy network profiling for intrusion detection," In Proc. of the 19th International Conference of the North American Fuzzy Information Processing Society (NAFIPS), Atlanta, GA, 2000, pp. 301-306.

[25]. M. Ramadas, S.O.B. Tjaden, "Detecting anomalous network traffic with self-organizing maps,"In Proc. of the 6th International Symposium on Recent Advances in Intrusion Detection, Pittsburgh, PA, USA, 2003, pp. 36-54.