Incorporating Hidden Markov Model into Anomaly Detection Technique for Network Intrusion Detection

J. Chandrakanta Badajena
College of Engineering and Technology
Bhubaneswar,Odisha

Chinmayee Rout
Ajay Binay Institute of Technology
Cuttack, Odisha

ABSTRACT

Now-a-days to increase the computation efficiency distributed systems are used in which the computing resources are shared among several systems. Such openness of distributed system leads to increase in potential attacks on the hardware and software by exploration of system vulnerability. This paper presents implementation of Intrusion Detection System (IDS) to model the behavior of users using Hidden Markov Model (HMM). This model attempts to detect intrusive attack efficiently. The IDS is an identification system which can be characterized by probabilities of false acceptance and false rejection. False acceptance means that the IDS allow intruders to continue their activity. False rejection means that the IDS stops the activity of a legitimate user. IDS can be developed by adoption of an appropriate mathematical model that allows us to generate user profiles efficiently and facilitates an effective and accurate decision-making process for intrusion detection. Due to the nondeterministic nature of user behavior, the decision about intrusive or nonintrusive behavior must take into account all evidence for and against the claim. So the probabilistic approach is to be implemented to model user profile to detect attack.

INDEX TERMS- Intrusion detection System, Anomaly detection technique, Hidden Markov Model, KDD Cup 1999 data set.

1. INTRODUCTION

In today's communication system computer and information security is a major concern as these are vulnerable to potential attackers. Because to increase the potential of advanced computer communication, distributed systems are used which combines all computer resources connected over a network to form a virtual super computer. This communication facility provides sharing of all resources among users independent of their location. Such facility provided by distributed computing environment leads to attack on the data flow over the network which affects integrity and availability of information greatly.

The attacker can also flood the network with so much traffic which slows down the legitimate traffic or flood a server or crash a server. Such an attack is called Denial-Of-Service (DOS)[5][7][11] attack which is an attempt to prevent legitimate users from accessing a computing resources. To detect such type of attack on data flow over network and prevent the damage to critical communication infrastructure the Intrusion Detection System (IDS) is adopted.

Intrusion Detection System is a defense system which assumes that the attacker gained an authorized access. It tries to identify the attackers by scanning the behavior of active users over the network and computing system. If a user exhibits a different characteristic then the normal user profiles, then it is identified as an attacker.

The IDS is an identification system which can be characterized based on false acceptance and false rejection probability. False acceptance means the IDS allow the intruders to continue their activity whereas false rejection is termed as probability to stop the activity of a legitimate user. Generally, an IDS analyses information patterns of network and host activity. The IDS logs the network event and looks after the existing system logs. Then it analyses the event logs to determine if any suspicious activity is going on. The IDS event analyzer uses knowledge of previous attacks and system vulnerabilities to identify intrusion.

The IDS uses two techniques according to the type of information used for intrusion detection: misuse detection [13] and anomaly detection [13].

Misuse detection uses knowledge about attacks. It attempts to model the attacks on a system as specific patterns and systematically scans the network and system events for each occurrence of the patterns. The advantage of this technique is that, the known attacks can be detected efficiently with low false positive error and it is economical enough as it requires scanning of known attack patterns. The disadvantage of this technique is that it suffers from detecting the new kind of generated attacks.

Anomaly detection technique is able to detect novel or newly generated and unknown attack, because it attempts to detect intrusions that have a significant deviation from normal behavior of a legitimate user. But drawback of anomaly detection technique is that the nonintrusive behavior failing outside the normal range maybe identified as an intrusion which in turn results high false positive error. Also a large amount of data and audit trail is to analyzed to model normal behavior.

2. LITERATURE REVIEW

A Probabilistic techniques[10] for intrusion detection based on Computer Audit Data proposed by Nong Ye, Xiangyang Li, Qiang chen, Sayed Masum Emran and Ming ming Xu, which presents a series of studies on probabilistic properties of activity data in an information system for detecting intrusion into the information system. It implements decision tree, Hotelling T^2 test and markov chain.

A model of Incorporation of soft computing techniques[12] into a probabilistic Intrusion Detection System proposed by Sung-Bae Cho, which presents a novel intrusion detection system that model normal behaviors using Hidden Markov Model (HMM) which is incorporated with neural network and fuzzy logic based on several models with different measures , fuzzy logic makes

the final decision of whether current behavior is abnormal or not. Fuzzy rules that utilize the models based on the measures of system call and file access.

3. RELATED WORKS

Previous to this proposed model there are several techniques implemented to detect an intrusion in a network. There are several techniques proposed for detection of anomalous behavior implementing IDS. There are a number of methods for constructing IDS models. Also it is possible to have IDS's deployment at different points in a working environment; like firewalls and application servers.

It is also possible to have different ways of detecting intrusions; using anomaly and misuse detection techniques. Besides these, different methods of modeling a specific IDS primarily based on the data source used in normal or attack pattern construction. Examples of such data sources can be user commands or request, packets exchanged, and system calls raised by applications.

Due to the nondeterministic nature of user behavior generally the probabilistic approach is the appropriate technique to be incorporated with anomaly detection technique for modeling IDS to model the user profile.

Anomaly detection[12][13] requires reference model, modeling techniques and recognition technique that determine whether current activity deviates from normal behaviors or not. The reference model requires collecting audit data from user, system and network activity. An IDS extracts several observations and applying modeling technique reduces them in order to profile the user's normal behavior. After modeling, IDS can determine whether current behavior is normal or not.

4. PROPOSED MODEL

4.1 Data Mining

Data mining (DM), also called Knowledge-Discovery and Data Mining, is the process of automatically searching large volumes of data for patterns using certain association rules.

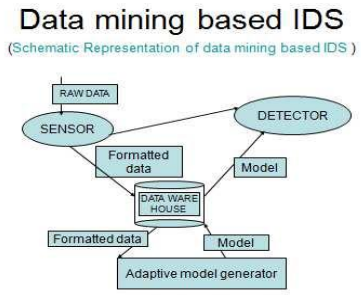

Figure 1: Datamining based IDS

It uses many computational techniques from statistics, information retrieval, machine learning and pattern recognition.

Here are a few specific things that data mining might contribute to an intrusion detection project:

• Remove normal activity from alarm data to allow analysts to focus on real attacks

• Identify false alarm generators and "bad" sensor signatures

• Find anomalous activity that uncovers a real attack

•Identify long, ongoing patterns (different IP address, same activity)

To accomplish the above tasks data miners employ one or more of the following techniques:

• Data summarization with statistics

• Visualization i.e. presentation of a graphical summary of the data

• Clustering of the data into natural categories

• Association rule discovery: defining normal activity and enabling the discovery of anomalies

• Classification: prediction of the category to which a particular record belongs

4.2 Data Mining and IDS

Data mining techniques can be characterized by their different model functionality and representation, preference criteria, and algorithms used. The main function of the model that we are interested in is classification, as normal, or malicious, or as a particular type of attack [5][6][7][11]. Additionally, data mining systems provide the means to easily perform data summarization and visualization, aiding the security analyst in identifying areas of concern.

Common representations for data mining techniques include rules, decision trees, linear and non-linear functions (including neural networks), instance-based examples, and probability models.

4.3 Data Mining and Real Time IDSs

In this paper, the authors explore the use of information-theoretic measures, i.e., entropy, conditional entropy, relative entropy, information gain, and information cost to capture intrinsic characteristics of normal data and use such measures to guide the process of building and evaluating anomaly detection models. They also develop efficient approaches that use statistics on packet header values for network anomaly detection. A serious limitation of their approaches (as well as with most existing IDSs) is that they only do intrusion detection at the network or system level.

This paper represents the data mining modeling of IDS, implementing Hidden Markov Model (HMM)[12][15] and KDD Cup 1999 data set[6].

4.4 KDD cup 1999 feature

Attacks that can be detected by KDD Cup 1999 fall into four main categories:

• DOS: denial-of-service, e.g. syn flood;

• R2L: Stands for Remote-to-Local specify unauthorized access from a remote machine, e.g. guessing password;

• U2R: Stands for User-to-Remote; unauthorized access to local super user (root) privileges, e.g., various ``buffer overflow'' attacks;

• probing: surveillance and other probing, e.g., port scanning.

4.5 Introduction to HMM

Hidden Markov Model (HMM)[12][15] based applications are common in various areas such as speech recognition , but the incorporation of HMM's for anomaly detection is in its initial stage.

The main strategy of our paper is to build an anomaly detection system, a predictive model capable of discriminating between normal and abnormal behavior of network traffic. Here we have to propose a model for detection of TCP network traffic as an attack or normal using HMM.

4.6 Behavior modeling with HMM

A HMM is a random, probabilistic and statistical process with an underlying probabilistic process that is not observable, bur can be only observed through another set of stochastic or random process that produce the sequence of observed symbols. This model can be thought of as an transition diagram with N nodes called state and edges representing transitions between those states. Each state node contains initial state distribution and observation probability at which a given symbol is to be observed. An edge maintains a transition probability with which a state transition from one state to another state is made.

Given an input sequence $O=O_1,O_2 ,\ldots\ldots,O_T$, HMM can model it with its own probability parameters using Markov process though state transition process cannot be seen outside. Once a model is built, the probability with which a given sequence is generated from the model can be evaluated.

A model $\lambda=(A, B, \pi)$ using its characteristic parameters as follows:

T: length of the observation sequences in the stochastic process

N: number of states in the model

M: number of observation symbols

Q: $\{q_1, q_2,\ldots, q_N\}$, set of states

V: $\{V_1, V2, \ldots., V_M\}$,discrete set of possible symbol observations

A: $\{a_{ij} \mid a_{ij} = Pr(q_j \text{ at } t+1)\}$,state transition probability distribution

B: $\{b_j(k) \mid b_j(k) = Pr(V_k \text{ at } t \mid q_j \text{ at } t)\}$, observation symbol probability distribution

π: $\{\pi_i \mid \pi_i = Pr(q_i \text{ at } t=1)\}$, initial state distribution

The probability with which the sequence is generated from the model can be calculated by summing the probabilities of all the possible state transitions.

Efficient methods used for observation modeling are:

1) Anomaly recognition

2) Normal behavior modeling

1) Anomaly recognition: Anomaly recognition matches current behavior against the normal behavior models and calculates the probability with which it is generated out of each model. Forward-backward procedure or Viterbi algorithm can be used for this purpose. Each probability is passed to the determination module that decides whether it is normal or not with a threshold.

Forward-back procedure calculates the probability $Pr(O \mid \lambda)$, with which input sequence O is generated with model λ using forward and backward variables.

Forward variable α denotes the probability at which a partial sequence $O=O_1,O_2 ,\ldots\ldots,O_T$ is observed and stays at state q_i

$$\alpha_t(i) = Pr(O_1,O_2 ,\ldots\ldots,O_t, i_t=q_i|\lambda)$$

According to this definition, $\alpha_t(i)$ is the probability with which all the symbols in input sequence are observed in order and the final stae is i. Summing up $\alpha_t(i)$ for all i yields the value $Pr(O \mid \lambda)$.

Back ward variable $\beta_t(i)$ can be calculated as $\beta_t(i) = Pr(O_{t+1}, O_{t+2} ,\ldots\ldots,O_T, i_t=q_i|\lambda)$

2) Normal behavior modeling: Determining HMM parameters is to adjust $\lambda=(A, B, \pi)$ to maximize the probability $Pr(O \mid \lambda)$. Because no analytical solution known for it, an iterative method called Baum-Welch reestimation is used.

This requires two variables:

$$\varepsilon_t(i,j) = Pr(i_t=q_i, i_{t+1}=q_j| O, \lambda)$$

$$= [\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)] / Pr(O \mid \lambda).$$

$\gamma_t(i)$ is the probability with which it stays at state q_i at time t:

$$\gamma_t(i) = \sum_{j=1}^{N} \varepsilon_t(i,j)$$

Summing up the two variables over time t, respectively the probability can be obtained with which state i transits to state j and the expectation that it says at state i.

Based upon the previous values calculated, a new model $\lambda =(A, B, \pi)$ can be adjusted using the following equations.

π_i = expected frequency (number of items) in state S_i at time (t=1)

$$= \gamma_t(i)$$

$\tilde{a}_{ij}$ = (expected number of transitions from state S_i to state S_j) / (expected number of transitions from state S_i)

$$= \sum_{t=1}^{T-1} \varepsilon_t(i,j) \;/\; \sum_{t=1}^{T-1} \gamma_t(i)$$

b_j = (expected number of times in state j and observing symbol v_k) / (expected number of times in state j)

$$= \sum_{t=1}^{T-1} \gamma_t(i) \;/\; \sum_{t=1}^{T-1} \gamma_t(i)$$

S.T $O_t=v_k$

After λ is adjusted from sequence O, $Pr(O| \lambda)$ is compared against $Pr(O| \lambda)$. If $Pr(O| \lambda)$ is greater than $Pr(O| \lambda)$, it implies that a critical point in likelihood has been reached, thereby finish the reestimation. Otherwise, $Pr(O| \lambda)$ is substituted by $Pr(O| \lambda)$ and reestimation continues.

5. EXPERIMENTAL ANALYSIS AND RESULT

Table 1. Actual Values and Discrete Observation Symbol values of the Features of one of the TCP sessions of KDD Cup 1999 data set

Feature No.	1	2	3	4	5
Values from a TCP session	22	185	554	8	15
Observation symbol values	1	1	1	1	1

Table 2 . Initial state distribution (parameter 'π' of HMM) for one of the TCP sessions of KDD Cup 1999 data set

States	Initial State Distribution values(π_i)
1	0.000581
2	0.261902
3	0.089830
4	0.375828
5	0.271858

Table 3 . State transition probability distribution (parameter 'A' of HMM) for one of the TCP sessions of KDD Cup 1999 data set

State	1	2	3	4	5
1	0.1456	0.1062	0.2718	0.2496	0.2265
2	0.0679	0.3353	0.2774	0.2005	0.1186
3	0.0191	0.1165	0.4648	0.1878	0.2115
4	0.6308	0.2844	0.0760	0.0029	0.0056
5	0.1404	0.1976	0.2123	0.2237	0.2257

Table 4. Re-estimated values of State transition probability distribution (Parameter 'A' ofHMM) for one of the TCP sessions of KDD Cup 1999 data set

States	1	2	3	4	5
1	0.0124459	0.0726424	0.670147	0.00490452	0.0109025
2	0.00298626	0.143447	0.45335	0.00209822	0.00365321
3	0.00041758	0.024481	0.362824	0.00104454	0.00515212
4	0.079060	0.337855	0.325613	9.8928E-05	1.4595E-09
5	3.2758E-215	1.804E+55	3.3E- 307	1.283E-307	3.275E-243

Table 5. Re-estimate values of State observation probability distribution (Parameter 'B' of HMM) for one of the TCP sessions of KDD Cup 1999 data set

States	1	2	3	4	5
1	0.752849	0.19004	0.0471515	0.00864518	0.00124678
2	0.559506	0.263164	0.111445	0.0409537	0.019992
3	0.309365	0.369747	0.22593	0.0751317	0.0175932
4	0.7508	0.179011	0.0562065	0.11656	0.00254451
5	0.237981	0.732203	0.0248835	0.391111	0.000877156

Table 6 . Re-estimeted Value Initial state distribution (parameter 'π' of HMM) for one of the TCP sessions of KDD Cup 1999 data set

States	Initial State Distribution values(π_i)
1	5.02736E-06
2	0.19047
3	0.801955
4	0.000226119
5	0.00734417

6. GRAPHICAL REPRESENTATION

Actual and Trained TCP Session Transition Probability Distribution

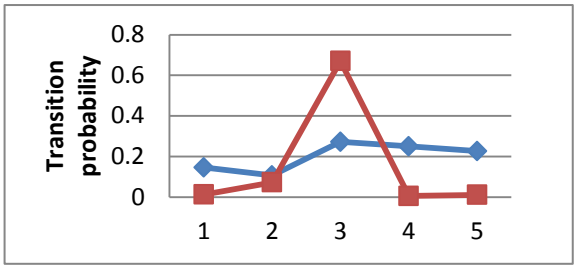

Figure 2. (State 1 transition probability distribution)

Actual and Trained initial state probability Distribution

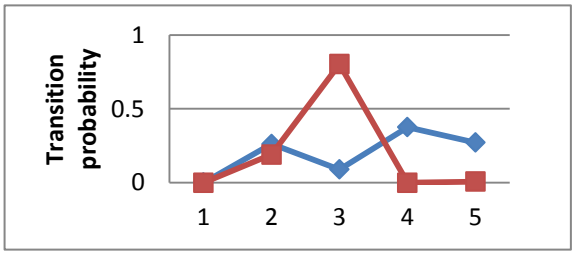

Figure 3.(Initial state distribution of 5 states of a TCP Session)

7. CONCLUSION

This paper investigated the capabilities of Hidden Markov Model in Anomaly Detection technique of Intrusion detection System. As described above, one HMM has been trained for each TCP session of the KDD Cup1999 data set. While training the model, special attention is given to the initialization of A, B, and π parameters and model selection issue. Training is performed using standard Baum-Welch procedure.

Network traffic for a TCP session is observed and modeled using different procedures like forward-backward procedure, Baum-Welch algorithm and viterbi algorithm and using the maximum likelihood principle the traffic is identified as either normal that is legitimate or abnormal that is intrusive using the recognition phase.

8. FUTURE WORK

For further work we have to incorporate data mining approach with the other network layer protocols to design a more sophisticated detection system to detect all variants of attack on the data packets transferred over network to maintain the network and information security.

9. ACKNOWLEDGEMENT

Our thanks to the experts who have contributed towards development of the template. We are extremely thankful to our parent and family members for their moral support and help. We are also thankful to our colleague for their idea to develop this paper.

10. REFERENCES

[1] Barford P., Kline J., Plonka D., Ron A.: A Signal Analysis of Network Traffic Anomalies, Proc. of the 2nd ACMSIGCOMM Workshop on Internet Measurements, 71 - 82 (2002)

[2] Biswanath Mukherjee, Todd L. Heberlein, and Karl N. Levitt. Network intrusion detection, IEEE Network, 8(3): 26-41,May/June 1994

[3] Cisco Systems, Inc., "Defining strategies to protect against TCP SYN denial of service attacks," July 1999, http://www.cisco.com/warp/ public /707 /4.html.

[4] Hajji H. Statistical Analysis of Network Traffic for Adaptive Faults detection. IEEE Trans. Vol. 16, no.5, 1053-1063 (2005)

[5] J. Mirkovic, G. Prier, and P. Reiher, "Attacking ddos at the source," 2002.

[6] KDD Cup1999 Data, Information and Compter Science, University of California, Irvine. http:// kdd. ics.uci.edddatabases/kddcup99/kddcup99 .html

[7] Kevin J. Houle and George M. Weaver, "Trends in denial of service attack technology," http://www.cert.org /archive /pdf/DoStrends.pdf, October 2001.

[8] Kim S. S. Reddy A. : Statistical Techniques for Detecting Traffic Anomalies Through Packet Header Data. Accepted by IEEE/ACM Tran. Networking (2008)

[9] L. Lee W., Xiang D. : Information-Theoretic Measures for Anomaly Detection. Proc. Of IEEE Symposium on security and privacy.(2001)

[10] Nong Ye,Xiangyang Li,Qiang chen, Sayed Masum Emran and Ming ming Xu *"Probabilistic Techniques for Intrusion Detection Based on Computer Audit Data. "* IEEE Transaction Systems, mans and cybernetics

[11] S.Bellovin, "Distributed denial of service attacks," Feb. 2000, http://www.research.att.com /~smb/talks.

[12] Sung-Bae Cho, *"Incorporating soft computing Techni*

ques Into a Probabilistic Intrusion Detection System" IEEE Transaction Systems, mans and cybernetics, Vol.32, No.2.

[13] Thotta M., Ji C.: Anomaly Detection in IP Networks. IEEE Trans. Vol. 51, No. 8, 2191-2004 (2003)

[14] T. H. Ptacek and T. N. Newsham, "Insertion, evasion, and denial of service: eluding Network intrusion detection," Secure Networks, Inc., Jan. 1998

[15] Yang Y., eng F., Yang H. : An unsupervised Anomaly Detection Approach using Subtractive Clustering and Hidden Markov Model. Communications and Networking in China. (2007)