Intrusion Detection using Hidden Markov Model

Sanjay Kumar Sharma
Ph.D. Scholar,
Mewar University,
Rajasthan

Manish Manoria
Director and Professor,
TRUBA Institute of Engineering and I.T. Bhopal,
MP

ABSTRACT

The success of modern day technologies highly depends on its effectiveness of the world's norms, its ease of use by end users and most importantly its degree of information security and control. Cloud computing is a new and emerging information technology that changes the way of IT architectural solutions and put forward by means of moving towards the theme of virtualization of data storage, local networks (infrastructure) as well as software. Cloud computing has been envisioned as a next generation information technology (IT) paradigm for provisioning of computing services with a reduced cost and fast accessibility. It provides greater flexibility with lesser cost like on demands services, scalable network, and virtualized services to the end users[13][14]. Security threats in existing technologies and legacy will remain for intruders [14]. In this paper, different intrusion detection and prevention techniques are studies which affect availability, confidentiality and integrity of Cloud resources and services. Also examines proposals incorporating Intrusion Detection Systems (IDS) in Cloud and types of attacks. Proposal of new ideas for detection and preventions of intruders to achieve desired security in the cloud [24].

General Terms

Cloud computing, Hidden Markov Model

Keywords

Cloud computing, IDS, IPS

1. INTRODUCTION

The "cloud" in cloud computing can be defined as the set of hardware, networks, storage, services, and interfaces that combine to deliver aspects of computing as a service. Cloud services include the delivery of software, infrastructure, and storage over the Internet (either as separate components or a complete platform) based on user demand. Technologies are changes day to day. Development of the communication technology, computer network is very fast. Cloud computing is increasingly studied and applied, which will promote profound changes in IT industry, but it will also bring enormous impact and challenges to users information and asset security realization and privacy protection[14].

1.1 Attacks on Cloud System

In this section we illustrate different type of attacks, which causes availability, confidentiality and integrity issues in cloud.

1) Wrapping attack

2) Malware Injection attack

3) Flooding attack

4) Data stealing problem

5) Accountability checking

Some details of above attacks are as follows:

Wrapping attack:- Injecting a fake element into the structure of the message is called wrapping attack.

Malware Injection attack:- An adversary attempts to inject malicious service or code, which appears as one of the valid instance services running in the cloud.

Flooding attack:- When an adversary has achieved the authorization to make a request to the cloud, then they can easily create bogus data and pose these requests to the cloud server. When processing these requests, the server first checks the authenticity of the requested jobs. Because non-legitimate requests must be checked to determine their authenticity, checking consumes CPU utilization, memory and engages the IaaS to a great extent. While processing these requests, legitimate services can starve, and as a result the server will offload its services to another server. Again, the same thing will occur and the adversary is successful in engaging the whole cloud system just by interrupting the usual processing of one server, in essence flooding the system.

Data stealing problem:- This type of attack is the common approach to breach a user account. The user account and password are stolen and the subsequent stealing of confidential data or even the destroying of data can hamper the storage integrity and security of the cloud.

Accountability checking:- As we all know that the payment method in a cloud system is based on the usage of the resources. When a customer launches an instance, the duration of the instance, the amount of data transfer in the network and the number of CPU cycles per user are all recorded. Based on this recorded information, the customer is charged. So, when an attacker has engaged the cloud with a malicious service or runs malicious code, which consumes a lot of computational power and storage from the cloud server, then the legitimate account holder is charged for this kind of computation. Though the customer is not aware of the attack and until the main cause of the CPU usage is detected, the providers will charge the customers first. As a result, a dispute arises and business reputations are hampered. All the focus for charging is based on the recorded parameters.

1.2 IDS AND IPS Techniques

There are many Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) techniques exist:

1. Signature Based Detection:- This technique identifies intrusion by matching captured patterns with preconfigured knowledge base and has high detection accuracy for previously known attacks in low computational cost.

2. Anomaly Detection:- It uses statistical test on collected behavior to identify intrusion. It reduces false alarm rate for unknown attacks also.

3. Artificial Neural Network Based IDS:- This method Classifies unstructured network packet efficiently using multiple hidden layers in artificial neural network and also increase the efficiency of classification.

4. Fuzzy Logic based IDS:- This method is used for quantitative features and provides better flexibility to some uncertain problems.

5. Association Rule based IDS:- It detect known attack signature or relevant attacks in misuse detection.

6. Support Vector Machine (SVM) Based IDS:- It classify intrusions correctly, when we have limited sample data. It can also handle massive number of features.

7. Genetic Algorithm (GA) based IDS:- Used to select best features for intrusion detection. Also has better efficiency as compared to others.

8. Hybrid Techniques:- It is used to classify rules efficiently by combining different techniques.

2. RELATED WORK

Sang-Jun Han and Sung-Bae Cho proposed [1] novel intrusion-detection technique based on evolutionary neural networks (ENNs). Many learning methods of neural network by training have explained like MLP (multi layer perception, Elmann neural network). In ENN method hidden nodes are generated automatically depending on output. This method is based on topology of neural network and works accordingly. They worked basically on User to Root (U2R) attack and compare their results with previously intrusion detection methods. They found good results using ENN method. We can improve the detection accuracy rate in less time by combining expert neural network algorithms then we can get good results.

Debin Gao, Michael K. Reiter and Dawn Song proposed [2] Novel HMM based behavioral distance to detect carefully crafted mimicry attacks that would evade detection by a system that utilizes traditional host-based anomaly detection or output voting. This approach compares the low level behaviors (like, system calls) of two diverse replicas when processing the same, potentially malicious inputs. If the two replicas are diverse and vulnerable only to different exploits, a successful attack on one of them might induce a detectable and increase in the behavioral distance. This makes mimicry attacks potentially more difficult, because to avoid detection, the behavior of the compromised process must be close to the behavior of the uncompromised one. Behavioral distance is based on measuring the evolutionary distance (ED) between replicas observable behaviors. We can work on behavioral distance method to find more accurate results in less time using other models.

Adel Nadjaran Toosi, Mohsen Kahani [3] discussed various soft computing techniques like neuro-fuzzy, fuzzy decision, genetic algorithm are combined and process the data using these techniques. The outputs of these techniques are provided to fuzzy inference which detects intrusions on the bases of training data set. KDD cup 99 is used as input dataset. Genetic algorithm is used to optimize the fuzzy decision making engine. It works on DoS, U2R, R2L attacks. By using feature selection method we can reduce the features for the classifiers.

Afroza Sultana, Abdelwahab Hamou-Lhadj, and Mario Couture proposed[4] Improved Hidden Markov Model and focuses on Hidden markov model(HMM) training time. Training time is reduced from 31.96% to 48.44% without affecting the detection accuracy. By frequent common patterns in the trace sequence not all patterns. n-grams extraction algorithm is used to extract common patterns.

Chenfeng Vincent Zhou, Christopher Leckie and Shanika Karunasekera[5] focuses on coordinated attacks such as large-scale stealthy scans, worm outbreaks and distributed denial-of-service (DDoS) attacks, which occurs in multiple networks simultaneously and proposes collaborative intrusion detection system(CIDS) which works on two phases one detection phase and another is correlation phase. The objective of Collaborative intrusion detection systems(CIDS) is to reduce the number of false alarms and irrelevant alerts that would be generated by individual IDS acting in isolation and produce a high level overview of the security and correlating the alerts from individual IDS. In detection phase they spread number of sensors in their sub network or individual host then generates low level alerts and in correlation phase, which transforms the low-level intrusion alerts into a high level intrusion report of confirmed attacks.

Suseela T. Sarasamma, Qiuming A. Zhu, and Julie Huff[6] proposes Multilayer hierarchical Kohonen Net or Kohonen self-organizing map (K-Map) to implement an anomaly based intrusion detection system (IDS sensor). By observing specific features in the packet headers that are more significant indications of abnormal activities.

3. PROPOSED SYSTEM

1.Construct cloud simulation environment which consist of active number of users and the datacenters.

2.The sender can send packets to the datacenter while the cloud environment is establish.

3.Select the dataset on which the training and testing of detecting intrusions is done.

4.Train the input dataset, as soon as the training finished it will generate a number of rules.

5.After that HMM is initialized with definite parameters at the server or at the broker level of the cloud.

6.The number of states present in the model will depends on the users in the cloud.

7.As soon as the user starts distribution of the packets to datacenter, then the HMM starts to compute the probability of each of the packet in the transition state. If the probability of any packet exceeds previously defined threshold value; then a rule system is developed based on the feature values of the dataset IDS.

8.Based on threshold value, discard the packets.

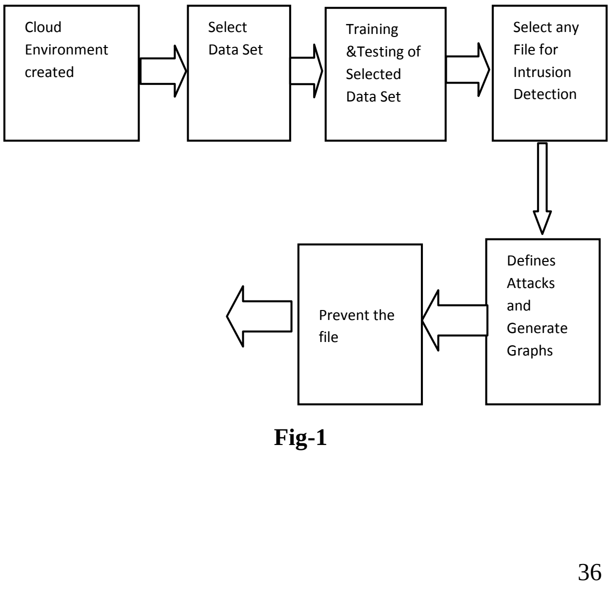
Cloud Environment created
Select Data Set
Training &Testing of Selected Data Set
Select any File for Intrusion Detection
Defines Attacks and Generate Graphs
Prevent the file

Fig-1

3.1 Probability Distribution using HMM

A Hidden Markov Model consists of five tuple:

N – Is the number of states available in the model Q {Q1, Q2, Q3…}

M – Is the number of observation symbols V {V1, V2, V3…..}

A – Transition Probabilities of State

B – Is the distribution of each of the state

Π – Is the initial state distribution

1.The preliminary transition probability from one state Q1 to a different state Q2 at a particular time occurrence t+1 depends on the state at time t according to the markov hypothesis i.e.

$$a_{ij} = p\ (q_{t+1} = s_j | q_t = s_i\quad)$$

2.The probabilities of the transition of the states is independent of the real time where the transition takes place according to the fixed hypothesis i.e.

$$p(q_{t1+1} = s_j | q_{t1} = s_i) = p(q_{t2+1} = s_j | q_{t2} = s_i)$$

3.Lets 'n' is the amount of packets 'pkt' send at a particular transition at a particular occurrence of time.

4.Calculate each step of the transition the state which is most possible $\hat{q}_i\ , 1 \le i \le T$ for the observation $z_i\ , 1 \le i \le T$, probability of state transition δt can be computed using Viterbi algorithm of HMM.

5.After each step of the transition; compute the general probability of each packet to be transmitted at each step Q.

6.The average probability of the state can be computed using

$$\delta_{avg} = {\textstyle\sum_{k=1}^{T}} \delta_k^{(i)} \Big/ T$$

7.The condition is checked i.e. if the average probability is less than the threshold value then the intrusion is detected in the packet.

$$\delta_{avg} < L\ (initial\, threshold\ value)$$

4. CONCLUSION AND FUTURE SCOPE

We can make efficient intrusion detection system by combining expert neural network algorithm and hidden markov model for efficient intrusion detection based on structure of neural network. To detect intrusion in less time and more accurate way. To classify large data we can use feature reduction by feature selection methods.

5. REFERENCES

[1] Sang-Jun Han and Sung-Bae Cho" Evolutionary Neural Networks for Anomaly Detection Based on the Behavior of a Program ",IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 36, NO. 3, JUNE 2006.

[2] Debin Gao, Michael K. Reiter and Dawn Song, "Beyond Output Voting: Detecting Compromised Replicas Using HMM-Based Behavioral Distance", IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 6, NO 2, APRIL-JUNE 2009.

[3] Adel Nadjaran Toosi, Mohsen Kahani," A Novel Soft Computing Model Using Adaptive Neuro-Fuzzy Inference System for Intrusion Detection", Proceedings of the IEEE International Conference on Networking, Sensing and Control, London, UK, 15-17 April 2007.

[4] Afroza Sultana, Abdelwahab Hamou-Lhadj, and Mario Couture, "An Improved Hidden Markov Model for Anomaly Detection Using Frequent Common Patterns ", IEEE ICC Communication and Information Systems Security Symposium, 2012.

[5] Chenfeng Vincent Zhou, Christopher Leckie and Shanika Karunasekera," A survey of coordinated attacks and collaborative intrusion detection", computers & security , pp 124-140 2 9 (2010).

[6] Suseela T. Sarasamma, Qiuming A. Zhu, and Julie Huff, "Hierarchical Kohonenen Net for Anomaly Detection in Network Security", IEEE TRANSACTIONS ON SYSTEMS,MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 35, NO. 2, APRIL 2005.

[7] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou and Jin Li, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 22, NO. 5, MAY 201.

[8] Wang Jun-jie, MuSen, "Security Issues and Countermeasures in Cloud Computing", IEEE, 2011.

[9] Farzad Sabahi, "Virtualization-Level Security in Cloud Computing", IEEE,2011.

[10] Chirag Modi, Dhiren Patel, Hiren Patel, Bhavesh Borisaniya, Avi Patel, Muttukrishnan Rajarajan, "A survey of intrusion detection techniques in Cloud", Journal of Network and Computer Applications, 2012.

[11] S.V. Narwane, S. L. Vaikol, " Intrusion Detection System in Cloud Computing Environment", International Journal of Computer Applications, 2012.

[12] Sung-Bae Cho, "Incorporating Soft Computing Techniques Into a Probabilistic Intrusion Detection System", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 32, NO. 2, MAY 2002.

[13] Qian Wang, Cong Wang and Kui Ren, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 22, NO. 5, MAY 2011.

[14] Wang Jun-jie and MuSen, "Security Issues and Countermeasures in Cloud Computing", 978-1-61284-491-6/111 2011 IEEE.

[15] Ramgovind S, Eloff MM and Smith E, "The Management of Security in Cloud Computing", 978-1-4244-5495-2/10 2010 IEEE.

[16] Yun Yang, Lie Wu and Wenping Hu, "Security Architecture and Key Technologies for Power Cloud Computing", 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE) December 16-18, 978-1-4577-1701-7/11 2011 IEEE Changchun, China.

[17] Ping Wang, Wei Huang and Carlos A. Varela, "Impact of Virtual Machine Granularity on Cloud Computing

Workloads Performance", 11th IEEE/ACM International Conference on Grid Computing, 978-1-4244-9349-4/10 2010 IEEE.

[18] Farzad Sabahi, "Virtualization-Level Security in Cloud Computing", 978-1-61284-486- 2/1112011 IEEE.

[19] Jinhui Yao, Shiping Chen and Chen Wang, "Accountability as a Service for the Cloud", 2010 IEEE 6th World Congress on Services.

[20] Joel-Ahmed M. Mondol, "Cloud Security Solutions using FPGA", 978-1-4577-0253-2/11 2011 IEEE.

[21] Modi, C., Patel, D., Patel, H., Borisaniya, B., Patel, A. & Rajarajan, "A survey of intrusion detection techniques in Cloud" Journal of Network and Computer Applications, doi: 10.1016/j.jnca.2012.05.003 <http://dx.doi.org/10.1016/j.jnca.2012.05.003>

[22] S.V. Narwane and S. L. Vaikol, "Intrusion Detection System in Cloud Computing Environment" , International Conference on Advances in Communication and Computing Technologies (ICACACT) 2012 Proceedings published by International Journal of Computer Applications® (IJCA).

[23] Irfan Gul, M. Hussain, "Distributed Cloud Intrusion Detection Model", International Journal of Advanced Science and Technology Vol. 34, September, 2011

[24] https://cloudsecurityalliance.org/