User Profiling System for Detection of Masquerading Attack on Private Cloud

1 Chalonsh D'silva, 2Deepika Mulchandani, 3Rujuta Pimprikar, 4Sneha Nair, 5Prof.Priya R.
Department Of Computer Engineering
Vivekanand Education Society’s Institute of Technology
Mumbai, India.
2deepika.mulchandani@ves.ac.in, 3priya.rl@ves.ac.in

Abstract— Cloud computing has advanced over the years as utility computing, application service provision and software as a service model. Cloud computing has competency of remodeling the entire architecture of Information Technology; however, not much has been done to overcome the threats to this type of computing model related to intruder utilizing resources of registered cloud user. This paper proposes a dynamic user profiling system which tries to overcome the security setbacks of cloud computing concerning masquerading. The User Profiling System monitors the user's behavior looking for divergence of behavior from normal thus improving cloud security by taking proactive and reactive measures on encountering atypical behavior. It analyzes the behavior of the users using soft computing technique of neural networks and fuzzy logic and identifies the malicious users of the cloud computing model.

Index terms— user profiling, neuro-fuzzy, masquerading

I. INTRODUCTION
Cloud computing is a model in the field of computing technology which is designed to provide scalable and measured resources. It is a pay-as-you-consume model wherein the infrastructure is shared by multiple users and resources can be accessed from anywhere across the world. Though cloud computing is a recent model, the users of this type of computing model are increasing phenomenally. Cloud computing model provides users and organizations with the facility to store their data at third party data centers. Though it is being used largely for storing data, this model has not been able to ensure the security of the data stored from various threats. Data stored on the cloud are of great value and due to this critical nature of the stored data cloud security is paramount. Traditional methods of authentication, authorization and encryption are incompetent to handle the nature of attacks possible on cloud as the cloud resources are distributed, virtual and heterogeneous.

Insider data theft or masquerading attack is an attack that cannot be avoided using the existing measures present for data security. Solutions like using fully homomorphic encryption are insufficient for data protection. Masquerade detection is very difficult if the attacker is an insider. In this type of attack the intruder poses as a legitimate user of the system. A malicious attacker can get the credentials of the legitimate user by means of password sniffing, key logger or through a break in. The attacker may steal passwords or exploit the trust of a legitimate user to assume his identity. Due to lack of resources and evidence it becomes very difficult to identify this type of attack. [5, 6, 7]

In this paper, we propose a user profiling system for detecting masquerade attacks in a cloud computing environment. This user profiling system will monitor the user’s behavior pattern to provide security against masquerading. The user profiling system will study the pattern of user data access and takes corrective action, based on a hybrid approach i.e. Neuro-fuzzy algorithm.

The flow of the paper is as follows: The first part comprises basic concepts related to cloud computing and also elaborates on the masquerading attack addressed in the paper. The second part describes all the methodologies that have been used for detection of the attack. It is followed by the work that has been previously implemented in this domain. The fourth and fifth parts throw light on the proposed design and system model. The sixth part explains the implementation details of the system, followed by the results of neural networks and fuzzy inference system in the seventh and eighth part respectively. The ninth part compares the accuracy of the obtained results with that of a predefined neuro-fuzzy system. The final part concludes along with the limitations as well as future scope for the user profiling system.

II. LITERATURE SURVEY
Cloud computing is the new buzzword in the IT industry. This model is nothing but the easy provision of computing resources over a network. The resources are provided by an entity known as the cloud service provider who handles the physical maintenance of hardware and software required for these resources .This new computing model has many perks. Many different types of services and applications are incorporated in this model. Scalability or elasticity is one of the features of cloud computing where the resources can be scaled up when the need arises and scaled down when not needed anymore. Another important benefit of cloud computing is that cloud services can be gained without association in person with the cloud service provider. Services
can be accessed by multiple platforms at anytime. Resource utilization of a user is managed by monitoring storage usage, CPU hours, bandwidth etc. Several technologies such as virtualization, grid computing and SOA have a part in making this model work. There are different architectures of cloud computing based on different services that are provided which are dependent on the service provider.

There are three different service models of cloud computing:

1. SaaS- Software as a service: Here, a service is provided usually through a web browser. It is basically an application running on cloud infrastructure. E.g. Google Docs
2. PaaS- Platform as a service: Here, an application can be created by the user using programming languages and tools provided by the PaaS provider and can be deployed on the cloud. User has control over applications that he has created. E.g. Azure
3. IaaS- Infrastructure as a service: Here, a user can run software using the processing, storage and network resources provided by the IaaS provider. The user has control over the operating system, the server and the applications. E.g. Amazon EC2 services

Also, there are four deployment models of cloud computing:

1. Private Cloud: The cloud is used by one organization or company. The cloud itself can be provided to the organization by a third party service provider.
2. Public Cloud: The cloud is public and is used by everyone. It is very cost-effective for its users. However, large investment is needed and hence such types of clouds are deployed by Microsoft, Google, Amazon, etc.
3. Community Cloud: This cloud can be owned by two or more organizations. It is usually deployed by schools.
4. Hybrid Cloud: This type of deployment model can be mixture of any two or more of the above models.[1]

Even though cloud computing has arrived recently in the IT industry; it has advanced to the point where it has large number of users. Many organizations are shifting critical data, key applications from internal (local) storage to the cloud. The data of an organization is stored away from their local machines, on virtual machines provided by the cloud service provider. It relieves the organizations from several management issues such as software updates, server management, configuration, etc. which are now handled by the cloud service providers. Even though it has several benefits with respect to management of physical network and infrastructure, ease of access and round the clock availability, there are risk factors associated with this model.[2]

Public clouds provide SaaS, PaaS and IaaS and are accessed by general public. They are owned by Cloud Computing providers. They have unstable network and the service is not individual enough. There is enormous implication on existing IT management process. Private clouds also provide SaaS, PaaS and IaaS but are accessed by internal organization of enterprise or community. They are owned by enterprises. They have a more stable network and individual service. They have hardly any effect on existing IT management process. Private clouds are constructed by enterprise or institution for provision of better service, safety and control. Enterprises control the infrastructures, and can also control the way how applications are deployed on it. Generally, private cloud is deployed in enterprise’s data center which is located behind firewalls, and it can also be deployed in a safe hosting place. [3]

Over a public cloud, the data of users is at multiple locations which are unknown to them and hence susceptible to various attacks. Over a private cloud the virtualized infrastructure and data store is on-site providing assurance to the users about where their data resides. However, in this case the threat to data is not completely eliminated as outsiders can still attack and the additional high risk of malicious insiders persists. These malicious insiders pose a high threat in cloud computing environment as they can be cloud service provider employees, i.e., the technical staff or cloud administrator.[12] The threat to private cloud is mainly from competitors of the organization, illegitimate users, etc. The cloud administrators have the highest possibility of violating the user’s privacy. The cloud administrators have privileged physical access to the machines as well as the technical expertise to deliberately violate the cloud users for their advantage. The administrator can steal data and provide it to a competitor or can change data leading to issues providing him monetary gain. Apart from the administrator who has privileged access to the database files, other insiders can pose a threat if they get access to these files. Hence, basic countermeasures such as encryption of user’s personal information should be taken. [12]

Masquerading attack is a consequence of identity theft. These attacks result due to stealing user’s credentials or may also be due to laziness or misplaced trust of the user. The countermeasures for data leakage and account hijacking alone are not sufficient to detect masquerading attack. In detecting masquerading attacks, the most important factor to be considered is that the attacker has already gained access to the legitimate user’s account and he does not try to exploit the access privileges of the user. Hence, normal access control mechanisms are not sufficient for detecting a masquerade attack. [6]

III. METHODOLOGIES

A. User Profiling

In any user profiling, the behavior of a user is captured and analyzed. This ‘behavior’ has to be unique for each user to detect malicious activities. ‘Search behavior’ is one such behavior which is unique and difficult to impersonate. Profiling this search behavior includes gathering data such as what the user searches or explores, at what time, how many times, and for what purpose. Though search behavior profiling has been used for personalized content display and web usage mining, it has not been used in security systems.[5,6,7]

B. Fuzzy System

For generating the user profile, user activities are monitored. User activities can be highly fuzzy and drawing a hard line between ‘malicious’ and ‘non-malicious’ activities is not reasonable. Quite often there may be unique and unusual access by user for performing some legitimate activity. Also Fuzzy algorithm presents a detailed view of the system instead
of representing it in an ambiguous manner. One cannot consider a user to be malicious depending on a few activities, and therefore in such a case a fuzzy system is most suitable as it generalizes the discrete values of normal and malicious into continuous probabilities.

The first step is Fuzzification where the degree of truth is determined using functions that are defined on input variables.

The second step is Inference wherein the computed truth values are applied to the conclusion part of each rule to generate fuzzy subsets which are assigned to each output variable. The third step is Composition, in which the all the fuzzy sets are combined to form a single fuzzy set for the final output variable. The last step is Defuzzification that converts the fuzzy set to crisp values.

C. Neural Networks

Smallest processing unit of a neural network is a neuron. Neural Networks receive input from processing units which are called input parameters. Neural networks have several hidden layers which receive input from other processing units and the processing is done in parallel. The set of processing units that are obtained as the result of processing are called output units.

Neural network algorithm consists of two stages namely training phase and testing phase. In the training stage, training data set input parameters are used which derive weights and bias for producing an output that matches required categorization of user type, known as target set. In the testing stage, new values of input parameters are used to produce an output based on training set outputs. There are two methods of implementing the learning algorithm for neural networks: supervised and unsupervised.

D. Neuro-fuzzy System

The neurons in a neural network work in parallel and each of these neurons communicate with each other using the weights between them. But, it is heuristically difficult to initialize weights in a neural network. Therefore, one cannot extract If-then rules as it can only take trained crisp values as input.[9] On the other hand, knowledge acquisition is difficult for fuzzy systems, and though it can encode this knowledge using rules, it is very time consuming. Using neural networks we can automate this process which greatly reduces the development time. Neuro-fuzzy networks thereby fill up each other’s disadvantages and provide a system that will be able to perform analysis in a precise manner.

As shown in Figure 3.1, the neural network is used to determine patterns for user profiling and this is given as input to the fuzzy system. The fuzzy system is trained by a learning algorithm derived from neural networks. It then takes the decision based on input from neural system and previous knowledge. This system can be represented to be made up of three layers: the input variables, the fuzzy rules and the output variables.

IV. RELATED WORKS

Sahil, Sandeep Sood, Sandeep Mehmni, Shikha Dogra,[10] analyzed the dynamic nature of security threats over the cloud and realized that security cannot be provided using only one mechanism, e.g., encryption or authentication and hence proposed a user profiling system with hybrid approach of artificial intelligence.

This system would records user’s activities and gives in-depth information about user’s activities using artificial intelligence techniques of genetic algorithms and fuzzy systems. The authors categorize user’s character (malicious or safe) based on user’s usage patterns. This usage pattern included resource utilization and traffic patterns. They checked for anomalies, e.g., suspicious traffic means the user can be an attacker or victim or both based on the way the traffic is outbound or inbound.

The User profiling system using fuzzy systems was hierarchical in their experiment. The historical data, i.e., previous usage pattern of the user was taken into account. In the first round, the user would be assigned a character (safe, malicious or highly malicious) which gets changed in the second round based on the analysis of historical pattern with current pattern of usage.

The genetic algorithm based UPS took enumerated characters of users to calculate fitness values and divided the users into ELITE parents (safe) having better fitness values and other parents (malicious and highly malicious) with worst fitness values. Then it searched for crossover and mutated other parents and limits the resources. Since both the above methods have problems,[10] suggested a hybrid system which would better the detection of malicious user.

M. Ben-Salem and S. J. Stolfo [5] proposed a system which tracked user’s activities and measured any significant change in the behaviour. Their system was based on the assumption that a masquerade’s intent would be visible through behaviour based on the volume of operations performed.

H. A. Kholidy, A. Erradi, S. Abdelwahed, and F. Baiardi,[13] presented a hierarchical analysing system which monitored and detected security threats and attacks in the cloud and autonomously suggested preventive measures for the same.

C. Chen, D. J. Guan, Y. Huang, and Y. Ou , [14] proposed a system which analyzed multiple logs files to determine intentions for an action which would help identify attacks from inside the computing environment and stealth attacks.

V. SYSTEM MODEL

The User Profiling System creates a profile for each user of the cloud using log files generated by using the interface of the cloud. The system then uses this profile for determination
of intrusion in the environment by using novelty detection of user behavior. The architecture can in general be divided into two broad areas: The Profile Generation phase and Classification phase.

- The profile generation phase is responsible for acquiring the user’s interests and inducing the user profile, and consist of three stages: observation stages, the feature extraction stage and the profile generation stage.

Actions performed by the user over the interface are captured. Features such as session time, number of clicks, number of instances and security groups created, etc. are extracted from these observations, and used to create training instances. The training instances are used to induce user profile.

- In the classification phase, based on analysis performed on the user profiles by neuro fuzzy system, i.e., by using fuzzy rules, the users are categorized into malicious, highly malicious and normal users. Figure 2, depicts the entire flow of the user profiling system from creation of log files of each of cloud to notification being send on detection of malicious activity.

![Flow of the system](image)

Figure 2. Flow of the system

The system is comprised of the following:

- The user makes an account on Eucalyptus Cloud, in order to avail the cloud services.
- While the user creates instances, volumes, etc. and navigates through the system, his activities get recorded in the log files.
- Structured data in the form of number of clicks per session, session time, time taken to create instances, volumes and security groups, number of instances, volumes and security groups, time spent on each page and navigation through the interface are stored in the database.

- This data is given as input to the neural network, which first trains the system and then compares the navigational patterns of the normal session of the user with that of the current session of the user.

- This output is sent to the fuzzy system, which then uses fuzzy IF-THEN rules, to decide the degree to which a user’s activities can be considered malicious.

- The output of fuzzy system is sent to the administrator, who then sends an email to the legitimate user to notify him of the possible attack.

VI. PROPOSED METHODOLOGY

The goal of the system is to prevent intrusion into the cloud resources of the users by masqueraders. For each user a profile is created which captures user behavior with respect to the following parameters.

- Session time.
- Clicks per session.
- Time taken to create an instance over the cloud.
- Time taken to create a volume over the cloud.
- Time taken to create security groups.
- Number of instances, volumes and security groups created in a given session.
- Time spent on different pages of the application interface.
- Navigation of interface.

These parameters are stored for each session of the user. For each user a separate table is created in the database where data is entered session wise. If the user is new and no previous session is present then a table gets created for him. This way each user’s behavior is gathered separately to generate profile for each user of the private cloud. This captured data is used as training data for generating user model of each user. This data acts as input to the neural network which trains the system with a particular user’s normal access behavior.

When a masquerader attacks, his behavior will differ from the legitimate user with respect to the parameters mentioned earlier. When the malicious attacker gains control of the user’s account, his activities are recorded. It is assumed that the normal user is familiar with his account on the cloud. Hence, this malicious attacker will be identified by the security system by testing his behavior against the trained user model. Anomalous activities can thus be efficiently detected by using user profiling.

Neural network output is then further used to tune membership functions in the fuzzy system. Neural network combined with fuzzy system provide a better detection efficiency as fuzzy rules are used to determine the degree to which the abnormality has been detected. The fuzzy system works in a hierarchical manner by using present neural output as well as previous neural outputs of the user for deciding the level of maliciousness which further increase the reliability and efficiency of the system.
Software agents are learning agents that perform activities with little or no human intervention. Software agents can evaluate and perceive human behavior to produce intelligent output. The user profiling system is created by software agents who are trained by neural network algorithms and feedback which contains the categorization (normal or malicious) based on the past behavior and is further segregated on the basis of maliciousness using fuzzy inference system. This ensures that the user profiles though initially created with minimum user behavior analysis are enhanced with time. [8]

The search features which are captured will act as the input parameters for the neural network in the neuro-fuzzy system. The input parameters will generate the neural outputs using the neural network hidden layers. The neural outputs will act as input to the fuzzy inference system. The fuzzy system will consult the knowledge base having the fuzzy if-else rules to generate the decisions which will be used to detect malicious users.

The proposed neuro-fuzzy system:
- Observes each user’s behavior.
- Fetches user access history if any, and learns and identifies the access pattern and uses the fuzzy system to identify unusual activities.
- It updates user’s category according to current inferences.
- Analyses the degree of maliciousness.
- Notifies the admin who takes any of the following actions according to the degree of threat:
  - Send email notification about malicious activity in the account with recovery measures.
  - Put the user in highly malicious category for constant and advanced monitoring

VII. IMPLEMENTATION

For the first phase, i.e., the profile generation phase a private cloud was setup using the Eucalyptus open software framework. It implements Infrastructure as a Service (IaaS) and provides the ability to run and control virtual machines. It is portable, modular and simple to use. Eucalyptus is a framework mostly used for academic research as it is open source. It is easy to install and as non-intrusive as possible.

It has a web interface and users interact with the system using the exact same tools and interfaces that they use to interact with Amazon EC2. It supports VMs that run on top of Xen hypervisor and VMWare.[4] It is very beneficial as it is AWS compatible. AWS is amazon web services which provide public cloud services using S3, EC2 and EBS. Eucalyptus uses the same interfaces and protocols as AWS and hence is able to be compatible with public cloud services of AWS like S3, EBS and EC2. The components of Eucalyptus are node controller (NC), cloud controller (CLC) and cluster controller (CC).[11]

Eucalyptus private cloud was setup using VMware workstation and Cloud-in-a-box faststart ISO of Eucalyptus. The cloud-in-a-box type of setup installs all the components of Eucalyptus together. Node controllers can be installed separately to provide scalability. The management console of eucalyptus was used to create private cloud users. User console allows users to create and launch instances. The user can create volumes, security groups, etc. He can access all the IaaS features provided by Eucalyptus easily. The Eucalyptus console is written in Python and using the ‘RotatingLogFileHandler’ logging handler of python, the logs of the user console were generated in the file ‘console.log’. A data extraction module was written in Java to extract the features mentioned above. These parameters are stored for each session of the user. For each user a separate table is created in the database where data is entered session-wise. If the user is new and no previous session is present then a table gets created for him. This way each user’s behavior is gathered separately to generate profile for each user of the private cloud.

This completes the profile generation phase. Neural network is used as the first step for classification. The neural network used by this system is multilayer perceptron. The neural network uses the first 5 sessions of the user for the initial training process to learn the behavior of each user of the system. These sessions are given non-malicious output, i.e., 0. Now, testing data is provided to the neural network to recognize the pattern and detect anomalies. The parameters from the feature extraction module are input to the neural network. The output is a value in the range of [0,1]; 0 being non-malicious and 1 being malicious.

VIII. FUZLY INFERENCE SYSTEM

The fuzzy inference system has two input variables and a single output variable. One of the input variables is the current neural network output, i.e., the current session which is being tested and the other input is the previous session output of neural network. The output of FIS is the degree of maliciousness in the session.

Here, we have used Mamdani fuzzy inference system.

Figure 3. Neural Network

The neural network classifies each session as malicious or non-malicious. But drawing a hard line between malicious and non-malicious cases is unreasonable and hence we proceed with a FIS (fuzzy inference system) which outputs the degree of maliciousness in the session.

Figure 4. Fuzzy System0
values between 5 and 7 depict a malicious scenario and between 0 and 5 indicate non-malicious or normal scenario.

Once the degree of maliciousness of a session of a user is identified then desired action has to be taken to inform the user. If the degree of maliciousness is greater than 7 then an email is sent to the user advising him to change his password and logout from the session before leaving a machine. Other safety guidelines are also mentioned. If the degree of maliciousness is less then no email is sent.

VIII. NEURAL NETWORK RESULTS

A. Non-Malicious User

Since 0 is specified as output for a non-malicious user, values closer to 0 signify that it is a normal user. Figure 5, shows neural output of a normal user session.

![Figure 5. Neural output for normal user](image1)

B. Malicious User

Since 1 is specified as output for a malicious user, values closer to 1 signify that it is a malicious user. Figure 6, shows neural output of a malicious session.

![Figure 6. Neural output for malicious user](image2)

IX. FUZZY RULES

If a user is highly malicious in the current session, then irrespective of his previous behavior, he will be given the highest degree of maliciousness. If the user is normal in the current session, then considering that he has been notified if previously he was malicious, he will be given the least degree of maliciousness indication he is normal. If the user is currently identified as malicious then if he was highly malicious or malicious in the previous sessions, he will be given the highest degree of maliciousness. If the user is malicious in the current session, and he was normal previously then he will be given malicious degree.

Figure 7, depicts the scenario where user was malicious in the previous session and is currently highly malicious then he is given the highest degree of maliciousness, in this case 8.5.

![Figure 7. Fuzzy rules](image3)

X. ANALYSIS OF RESULTS

The above inference regarding the degree of maliciousness of the user has been drawn based on neural network results, followed by fuzzy inference system. However, there exists a comprehensive neuro-fuzzy system in Matlab that trains data and applies its own set of fuzzy rules, to give a direct output.

![Figure 8. Neuro-fuzzy output for non malicious user](image4)

As in previous case, values closer to 0 signify non-malicious user and values closer to 1 signify malicious user. The output obtained from comprehensive neuro-fuzzy system for non malicious user is 0.445, whereas the output for malicious user is obtained as 0.489.
It was observed that the system is unable to clearly distinguish between malicious and normal activities and gives a near same value for both, resulting in a high rate of error. This occurs due to the fact that the combined neuro-fuzzy system in Matlab considers the maximum and minimum value for every parameter, thereby giving erroneous results.

Therefore it can be concluded that the procedure of using neural network results as input to FIS provides a much precise and accurate outcome as compared to combined neuro-fuzzy system.

CONCLUSION

The existing traditional security mechanisms are insufficient to handle threats as they rely on the secrecy of one’s login credentials for providing security. In such cases, a masquerading attack goes undetected. The User Profiling System, however, can identify malicious activities taking place over a private cloud even in the case of a masquerading attack. It extracts the traits and characteristics of the user’s behavior from the log files generated in the cloud. The neural network is trained using supervised learning to identify behavioral patterns in the given data set, and consequently perform anomaly detection. Further, fuzzy inference system is implemented to decide the degree of maliciousness of the user – normal, malicious or highly malicious. Depending on the degree of maliciousness inferred, an email is sent to the user notifying about the observation of abnormal activity & requesting to change passwords.

There are a few limitations of the user profiling system –
- It is not a real time working model. The system can successfully detect a masquerading attack, but cannot prevent it.
- The fuzzy inference system used does not take into consideration the cumulative maliciousness of the user from all previous sessions while deciding the current degree of maliciousness.
- A common problem in online classification tasks is concept drift, which is when the target concept changes over time. Identifying concept drift is often difficult. The system does not take fully into account the problems with concept drift.

However, the User Profiling System produces accurate results with all types of datasets when not considering the concept drift factor.

XI. FUTURE SCOPE

The future scope for the project is as follows:

- Using unsupervised learning algorithm for training the neural network as that can help capture a wider range of malicious activities.
- Processing data in a manner such that the fuzzy inference system can access the cumulative degree of maliciousness of the user before deciding the current degree.
- Expanding the domain of the project to Big Data Analytics so that real time processing of data can be implemented. Consequently the action taken on detecting a malicious user can be preventive in nature, such as blocking access, asking security questions, etc.

References

Advances in Computer Engineering and Applications (ICACEA).


