IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

An Intrusion Detection System for Malicious Attacks in Cloud Environment Using Decision Tree Techniques

Gopala B[1*], M. Hanumanthappa[2]

[1,2]Dep. of Computer Science and Applications Bangalore University, Bangalore, India.

[*]*Corresponding Author: gopala.nishanth@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 18/Aug/2018, Published: 31/Aug/2018

Abstract- Secured and reliable services in cloud computing environment is an important issue. A sensitive data in a cloud computing environment is major issues with regard to security in a cloud based system. Many cloud service providers obtain server from other service providers due to it is cost affective and flexible for operation which makes way for data stolen from the external server. To counter a variety of attacks, especially large-scale coordinated attacks, this paper provides a framework for identifying intrusions in cloud environment using Decision Tree Techniques. The proposed system could reduce the impact of these kinds of attacks through providing timely notifications about new intrusions to Cloud users' systems.

Keywords- Cloud computing, IDS, Threats, Decision Tree, cloud service providers**.**

I. INTRODUCTION

Internet has been a driving force towards the various technologies that have been developed. Cloud computing is seen as a trend in the present day scenario with many organizations trying to access data from cloud environment**.** The cloud is emerging as a powerfull delivery models for IT capabilities. It is a way of delivering IT- enabled services in the form of software, infrastructure and more. Cloud computing can be defined as "A computing Cloud is a set of network enabled services, providing scalable, QoS guaranteed, normally personalized, inexpensive computing platforms on demand, which could be accessed in a simple and pervasive way". In simple, Cloud computing is the combination of a technology, platform that provides hosting and storage service on the Internet.

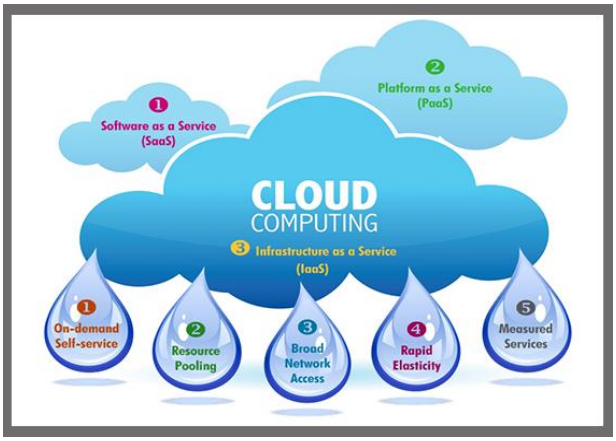

Figure 1

Cloud computing aims to provide scalable and inexpensive on -demand computing infrastructures with good quality of service levels. Cloud Computing is the implementation of engineering principals to obtain high quality applications through Internet. Main goal of the cloud computing is to provide scalable and inexpensive on-demand computing infrastructures with good quality of service levels. Cloud computing provides the internet based, highly scalable distributed computing systems in which computational resources are offered as a service. The advantages of using cloud computing are: i) reduced hardware and maintenance cost, ii) accessibility around the globe, and iii) flexibility and the highly automated process wherein the customer need not worry about software up-gradation which tends to be a daily matter.

II. CLOUD COMPUTING ISSUES

In cloud computing internet connection is necessary for using a cloud service. Another concern of many users have is that their data is no longer stored locally and they may lose control over it as soon as it is stored in the cloud. But some users may not even have a choice to use cloud services or not or may not be aware that they are using a cloud service at all which can also be an issue. The transmission of data over the internet imposes threats to data confidentiality during transmission in unsecured connection and if data stored is unencrypted there may be possible data leaks on a provider's server. As the exact position where the data is stored in the cloud is not known, it is also unclear which laws are applicable. In addition many local laws treat data which is stored locally different than data which is stored somewhere

on the web. This makes it even harder to decide which law is applicable.

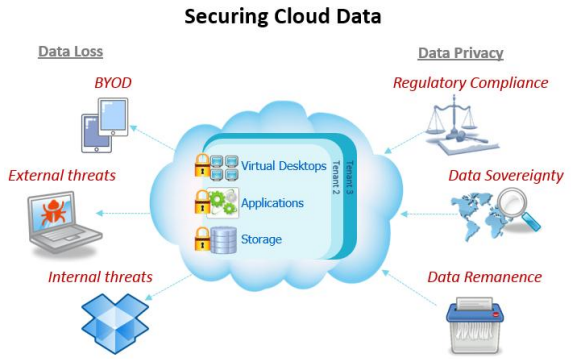

Figure 2

A sensitive data in a cloud computing environment is major issues with regard to security in a cloud based system. Anyone from anywhere anytime can access data from the cloud since data may be common, private and sensitive data in a cloud. So at the same time, many cloud computing service consumer and provider accesses and modify data. Therefore there is a need of data integrity method in cloud computing. Data stealing is a one of serious issue in a cloud computing environment. Many cloud service provider do not provide their own server instead they acquire server from other service providers due to it is cost affective and flexible for operation which makes way for data stolen from the external server. Data loss is a common problem in cloud computing. If the cloud computing service provider shut down his services due to financial or legal problem then there will be a loss of data for the user

III. THREATS AND ATTACKS

The threats and attacks on web applications which are executed in a web browser are grouped as hackers, cloud service providers and governments. For each of them several possible attacks are explained.

a. Hackers

Session Hijacking One of the most well known attacks against cloud services. This attack is only possible if https is not used at all or only during the login process. Unfortunately most providers only offer https for login, not during the whole session for example Google use https by default on all it's services, Facebook has a hidden option to enable https and Yahoo does not provide https. The providers state that there is no customer demand for https. But many customers do not know what risks they are exposed to if not using https therefore the missing customer demand is due to a lack of information. The main reason why the providers do not offer https by default is an economic one. Https needs computing power, therefore more servers are necessary to maintain the service level constant if https is used instead of http. Adding more servers means higher running costs and a lower profit. So many providers tend to offer https as a hidden option, because only a very little percentage of all users will activate this option while the majority does not use https because the do not know about it. This minimizes the additional costs.

The user logs in to the cloud service via https. After the login data is varied, the service provider sends some session information via plain http. At this point an attacker which is able to intercept the network , if he is in the same public Wi-Fi network or uses the same network hub, captures the session information. But he does not have to capture it at this point, as the session information is transmitted with every action the user performs, so the attacker has just to be able to intercept the network track at any time. As soon as he got the session information he can do everything also the user can without anyone, neither the user nor the service provider, noticing. He can impersonate the real user as long as the real user stays logged in and can read all the information which is transmitted between the cloud service provider and the user. Therefore, it is also very important for the user to log out every time he leaves the website or the cloud service and not just to close the browser windows. Else the session cookie or ID remains valid and the attacker can just continue using the website like he is the real user as long as he wants.

b. Man-in-the-Middle Attack

This one only works if the login page is a plain http one and only the actual login information is transferred via https. An attacker tries to modify the DNS information requested by the user, e.g. by forging DNS packets, DNS cache poisoning or ARP spoofing. If he is able to do so, he can redirect the user to his own login pagewhich may look exactly the same as the original login page. The user does not see any difference and since it is plain http also the browser does not warn the user that he is on a different site now. Now the user enters his login information, neither the user nor the service provider notices anything about the attack on the fake login page, the attacker gets this information in plaintext, encrypts it using https and sends it to the cloud service provider. For provider it is exactly the same as if the user logged in directly. Now the attacker can relay all the track between the user and the service provider. Of course he can read this track, as it is unencrypted. In addition the attacker also knows the login information of the user.

The majority of cloud service providers are using an https login page, where the track redirection is no longer possible this way due to the SSL certificates which are used with https. The browser would warn the user if the track is redirected as the certificate does not match the hostname or a plain http site is received instead of an https one. Therefore,

this type of man-in-the-middle attack is not a big concern any more.

c. User Interface Attacks

Most of the cloud applications are accessed using a web browser. Modern web browsers tend to store all kinds of information, including browsing history, login information, form data. All attacks under this category mainly exploit bugs and other security issues present in web browser software. If some specific web browser has known vulnerabilities these can be used to trick the browser to show a fake URL instead of the correct one and therefore showing a fake login page to the user. It may also be possible to gain access to the stored browser cache, including stored login information and passwords. Another possibility is to trick the browser to insert stored auto-complete data on a different website than the one the browser stored it for. The user may also be fooled to think a fake website is the real one by homographic attacks which are using international characters in the URL which look like national ones and there is no difference at the first glimpse. An attacker try to fake the https lock icon by exploiting browser vulnerabilities. Nearly every software has some bugs and security vulnerabilities but most modern web browsers only have minor ones and if there are some more severe, they are usually axed in a few days, so this attacks are only a minor issue if it comes to data confidentiality and privacy in cloud computing.

IV. CLOUD SERVICE PROVIDERS

The cloud computing service provider must make sure that the customer personal information is well secured from other providers, customer and user. As most of the servers are external, the cloud service provider should make sure who is accessing the data and who is maintaining the server so that it enable the provider to protect the customer's personal information. Cloud computing service provider should have the complete access to the server with all rights for the purpose of monitoring and maintenance of server. So this will prevent any malicious user from uploading any infected application onto the cloud which will severely affect the customer and cloud computing service.

a. Exploiting User Data

The business model of many cloud service providers offering free services is built on highly personalized advertisements. Therefore, they scan the user data on tags and keywords and look for other companies which are offering products matching these tags. Then they charge these companies a small fee for showing the user an advertisement on the web interface. But they are not only scanning for simply keywords, also more sophisticated methods are used to obtain all kinds of statistics from a user's emails, documents etc. This is described under the term data mining, which means extraction, analysis and usage of data in a way that it was not originally stored for. The providers can link several different kinds of user data together to get highly accurate user probes which can then be used to do behavior prediction. This can go as far as the cloud provider knowing when the user has to buy a new pack of toilet paper and then showing an advertisement for toilet paper just in the right moment, even if the user might not have thought about buying it just yet. It is obvious that storage encryption is against their business model as long as the provider is not able to decrypt the data. The only thing he could do with encrypted user data is to show random advertisements which is much less efficient and therefore much less portable for the provider. Moreover, service providers claim that users would not even pay for storage encryption or enhanced privacy protection, so there is no need for them to offer it. Some providers state that they do not scan the data if a paid service is used, but the problem is that nobody is able to check if they do so or not. One can only say that they have the technology to scan and analyze the data and it would be easy for them to use it also with paying users.

Data Leaks through Employees is possible if the data is stored unencrypted, a displeased employee of the cloud service provider may access the data and disclose it to a third party. The storage encryption is an effective countermeasure against this type of data leak, especially if only the user is in possession of the key.

V. DECISION TREES

Decision tree is one of the classification algorithms in data mining. The Classification algorithm is inductively learned to construct a model from the reclassified data set. Each data item is defined by values of the attributes. Classification may be viewed as mapping from a set of attributes to a particular class. The Decision tree classifies the given data item using the values of its attributes. The decision tree is initially constructed from a set of pre-classified data. The main approach is to select the attributes, which best divides the data items into their classes. To classify an unknown object, one starts at the root of the decision tree and follows the branch indicated by the outcome of each test until a leaf node is reached. The name of the class at the leaf node is the resulting classification. Decision tree induction has been implemented with several algorithms. Some of them are ID3[7] and later on it was extended into C4.5[8]. C4.5 avoids over fitting the data by determining a decision tree, it handles continuous attributes, is able to choose an appropriate attribute selection measure, handles training data with missing attribute values and improves computation efficiency. C4.5 builds the tree from a set of data items using the best attribute to test in order to divide the data item into subsets and then it uses the same procedure on each sub set recursively.

VI. CLASSIFYING INTRUSIONS USING DECISION TREE

In this paper we proposed a Decision Tree based Intrusion Detection system on cloud environment. The proposed work can be summarized as follows

1. Monitoring and analysis of user and system activities.
2. Checking and comparing vulnerabilities.
3. Abnormal behavior analysis
4. Apply decision tree algorithm for classification of intrusions.

The KDD99 dataset was used for building a network intrusion detector, a predictive model capable of distinguishing between intrusions and normal network connections. In 1998, DARPA intrusion detection evaluation program, a simulated environment was set up to acquire raw TCP/IP dump data for a local-area network by the MIT Lincoln Lab to compare the performance of various intrusion detection methods. It was operated like a real environment, but being blasted with multiple intrusion attacks and received much attention in the research community of adaptive intrusion detection. The KDD99 dataset contest uses a version of DARPA98 dataset. In KDD99 dataset, each example represents attribute values of a class in the network data flow, and each class is labeled either normal or attack. The classes in KDD99 dataset categorized into five main classes (one normal class and four main intrusion classes: probe, DOS, U2R, and R2L).

1) Normal connections are generated by simulated daily user behaviour such as downloading files, visiting web pages.

2) Denial of Service (DoS) attack causes the computing power or memory of a victim machine too busy or too full to handle legitimate requests. DoS attacks are classified based on the services that an attacker renders unavailable to legitimate users like apache2, land, mail bomb, back, etc.

3) Remote to User (R2L) is an attack that a remote user gains access of a local user/account by sending packets to a machine over a network communication, which include sent mails.

4) User to Root (U2R) is an attack that an intruder begins with the access of a normal user account and then becomes a root-user by exploiting various vulnerabilities of the system. Most common exploits of U2R attacks are regular buffer overflows, load module.

5) Probing (Probe) is an attack that scans a network to gather information or find known vulnerabilities. An intruder with a map of machines and services that are available on a network can use the information to look for exploits.

In order to evaluate the performance of algorithm for intrusion detection on cloud environment, we performed classification using KDD99 dataset. The results of the comparison of ID3 and C4.5 algorithms are tabulated in table 1.

Table 1 Comparision Results

Method	Normal	Probe	DOS	R2L	R2L
ID3 (DR%)	96.63	96.33	96.25	55.21	93.5
ID3 (FP%)	0.66	0.46	0.55	0.63	0.51
C4.5 (DR%)	97.63	96.61	97.25	56.21	94.5
C4.5 (DR%)	0.12	0.16	0.15	0.42	0.21

Table 2 Test Data for detecting intrusions

Tracks	No.of files sent	No.of Intrusions detected	Time taken	Througput in bytes
1	25	5	370	200
2	30	8	380	200
3	35	12	410	292
4	40	18	515	490
5	50	22	625	595

It is seen in Table 2 that the number of intrusions detected gradually increases with increase in time and hence the throughput increases

VII. CONCLUSION

The chances of intrusions are more with the erudition of intruders attacks in cloud computing over internet.. Different IDS techniques are used to counter malicious attacks in traditional networks. .In this work we used Decision Tree method to classify the intrusion data set, the proposed approach is to identify the unseen or unknown attacks in Cloud Environment.

REFERENCES

[1] Xinfeng Ye Access Control for Cloud Applications IEEE 2015.

[2] Sonia Bassi et al. Cloud Computing Data Security-Background & Benefits IJCSC 2015.

[3] Navadeep Agganwal et al. Cloud Computing: Data Storage Security Analysis and its Challenges International Journal of Computer Applications Volume 70– No.24, May 2013.

[4] Keiko Hashizume et al. An Analysis of security issues for cloud computing, Journal of Internet services and applications, Springer 2013.

[5] Manoj K et al. Unsupervised Outlier Detection Technique for Intrusion Detection in Cloud Computing, ICCT IEEE 2014.

[6] Roshanak Roshandel et al. User-Centric Monitoring of Sensitive Information Access in Android Applications, 2nd ACM International Conference on Mobile Software Engineering and Systems IEEE 2015.

[7] Richa et. al, To Improve Security in Cloud Computing with Intrusion detection system using Neural Network. IJSCE 2013.

[8] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.

[9] J. R. Quinlan. Induction of Decision Trees. Machine Learning, 1:81-106, 1986.

[10] Manish, Dr. Hanumanthappa M, Intrusion Detection System Using Decision Tree Algorithm, IEEE 2012.

