AN INTRUSION DETECTION SYSTEM TO PREVENT THE K-ZERO DAY ATTACK PROPAGATION

AUTHORS

Bharath A,Praveen S

1.Student,Information Technology,Bannari Amman Institute of Technology,Tamil Nadu,India
2.Student,Computer Science and Engineering,Bannari Amman Institute of Technology,Tamil Nadu,India

ABSTRACT

Network security refers to any activities designed to protect and prevent the network from unauthorized access or denial of the computer network. Denial-of-Service is one of the major attacks of the network. In this attack, the attacker compromises a system in network. The compromised machine in a network involves in the spamming activity. It increases the traffic of the network. It can be prevented by various methods. One of the methods to detect the spam data is by using content-based spam detection technique. The Porter's algorithm is used for the content-based spam detection. This algorithm includes the stop word removal and stemming process.

The main idea of this approach is to monitor the data of the mail. It is carried out using content-based spam detection method in C#.NET. It blocks the mail if it contains any spam data. So, it prevents network degradation and improves the efficiency of the network.

Keyword:*Unauthorized Access, Denial of Service*

1.INTRODUCTION

Network security refers to any activities designed to protect and prevent illegal access to the network or denial of the computer network. Denial-of-Service is one of the major attacks of the network. In this attack, the attacker compromises a system in network. The hacked network device involves in the spamming activity. It increases the traffic of the network. It can be prevented by various methods. One of the methods to detect the spam data is by using content-based spam detection technique. The Porter's algorithm is used for the content-based spam detection. This algorithm includes the stop word removal and stemming process.

The main idea of this strategy is to monitor the data of the mail. It is carried out using content-based spam detection method in C#.NET. It blocks the mail if it contains any spam data. So, it prevents the network degradation and improves the efficiency of the network

2.LITERATURE REVIEW

Customer network activities are a quick and efficient strategy. DBam activates its spam suppressing mechanism to break the internal spam-washing going on in a client network. DBSpam activates its spam suppressing mechanism to break the spam laundering. A prototype of DBSpam based on libpcap, is implemented to validate its efficacy through both theoretical analyses and trace-based experiments.

Yegneswaran, V. et al. (2007) proposed BotHunter is a brand-new form of network perimeter monitoring technology. The recognition of the infection and coordinating conversation that take place during a successful malware infection are its main concerns. In order to detect specific stages of malware infection, such as inbound scanning, exploit exploitation, egg downloading, outbound bot coordination dialogue, and outbound assault propagation, BotHunter uses a correlation engine powered by three malware-focused network packet sensors..

Chen, Z. et al. (2007) proposed a method that describes the behaviours of Localized-Scanning(LS) worms. It is a simple technique used by attackers to search for vulnerable hosts. LS describes the trades off between local and global search of vulnerable hosts. Optimal localized-scanning strategy is designed, which provides an upper bound on the spreading speed of localized-scanning self-propagating codes. The feedback localized scanning and the ping-pong localized scanning adapt the scanning methods based on the feedback from the probed host, and thus spread faster than the original localized scanning and meanwhile have a smaller variance.

Duan, Z. et al. (2007) proposed a method that studies the architectural foundation of the current mailing system that is responsible for the email spamming. The current system is sender driven but the proposed Differentiated Mail Transfer Protocol (DMTP), grants receivers control over Internet. Compared to the current Simple Mail Transfer Protocol (SMTP)-based email system, the proposed email system can force spammers to stay online for longer periods of time that may significantly improve the performance of various real-time Blacklists of spammers.

Xie, Y. et al. (2007) proposed a method UDmap, to identify dynamically assigned IP addresses and analyze their dynamics pattern. It is fully automatic, and relies on application-level server logs that are already available. It highlight the importance of being able to accurately identify dynamic IP addresses for spam filtering, and expected similar benefits of it for phishing site identification and botnet detection.

Gu, G.et al. (2008) proposed a general detection framework that is independent of botnet Command and Control (C&C) protocol and structure, and requires no prior knowledge of botnets. A botnet is defined as a coordinated group of malware instances that are controlled via C&C communication channels. Accordingly, the detection framework clusters communication traffic and similar malicious traffic, and performs cross cluster correlation to identify the hosts that share both similar communication patterns and similar malicious activity patterns.

Junjie Zhang et al. (2008) proposed an approach BotSniffer, that uses network-based anomaly detection to identify botnet C&C channels in a local area network without any prior knowledge or C&C server addresses. BotSniffer can capture this spatial-temporal correlation in network traffic and utilize statistical algorithms to detect botnets with theoretical bounds on the false positive and false negative rates. BotSniffer can detect real-world botnets with high accuracy and has a very low false positive rate.

Zhuang, L. et al. (2008) proposed new techniques to map botnet membership using traces of spam email. The multiple bots participating in same spam email campaign method is used to group bots into botnets. The technique is applied against a trace of spam email from hotmail Web mail services. The method successfully identified hundreds of botnets. It provides new findings about botnet sizes and its behavior.

Xie, Y. et al. (2008) proposed a method AutoRE, that detect botnet-based spam emails and botnet membership. It does not require pre-classified training data or white lists. Also it outputs high quality regular expression signatures that can detect botnet spam with a low false positive rate.

l obfuscation, properties of botnet IP addresses, sending patterns, and their correlation with network scanning traffic.

Sanchez, F. et al. (2010) proposed adetailed analysis of the spam emails' received header fields There hasn't been a thorough study of spammers' forging behaviour, considering how important it is for spam control and identifying the real spam originators. A thorough investigation into the Received header fields of spam emails will look into the extent to which spammers may and do alter the trail data of spam emails, among other things. results and how they may affect various spam management initiatives, such as email sender authentication and spam filtering.

Bacher, P.et al. (2011) proposed a technique called Honeypots for discovering the tools, tactics, and motives of attackers. It focuses on a special kind of threat: the individuals and organizations that run botnets. A botnet is a network of compromised machines that can be remotely controlled by an attacker. With the help of honeynets from the people running botnets, a task that is difficult using other techniques can be observed. Due to the wealth of data logged, it is possible to reconstruct the actions of attackers, the tools they use, and study them in detail.

Duan, Z.et al. (2012) proposed a SPOT spam zombie detecting system by monitoring outgoing messages of a network. It is designed using the statistical tool called Sequential Probability Ratio Test, that detects the ratio of false positives to false negatives. By keeping an eye on the sent messages, it assesses the effectiveness of the developed system and so locates the network's compromised system. Also reducing the energy consumption in wireless adhoc network for larger number of nodes. Hence a source based method called genetic algorithm is used in NS2 that can improve the total

3.PROBLEM STATEMENT

The main problem in the network is the presence of number of compromised machines in the network. Thus these systems commonly referred as spam zombies, that are used for sending spam messages should be detected. Thus the SPOT system is used to find those machines. The major drawback is that it is only used for static IP and the problem occurs when the IP is changed. Thus the content-based spam detection system is used. It monitors the content of the mail along with IP address. So the spam contents are blocked.

There are various anti-spam techniques used by both end users and administrators of email systems to prevent email spam that classifies the web pages as spam or not-spam according to their textual content. Its metrics are intended to extract some information from the textual content of the web pages and are compared with the keywords that are already stored by user. It uses a method called stemming. This process reduces derived words to their stem or root form. So the terms of document are represented by stems. It reduces the dictionary size and makes the search effective.

4.PROPOSED METHODOLOGY

4.1 System Design

A passive attack watches unencrypted communication and scans it for sensitive data and passwords that can be utilised in other attacks. It includes things like traffic analysis, weakly encrypted traffic decryption, monitoring of open communications, and password capture. The attacker attempts to get around or breach secured systems during an active attack. This can be accomplished using malware, Trojan horses, or stealth. Active attacks include attempts to defeat security measures, introduce malicious code, steal or change data, and circumvent other security measures. Active attacks cause Denial-of-Service (DoS), the dispersion of data files, or the change of data.

FIGURE 1 TYPES OF ATTACK

4.2 Methodology

The survey describes different methods to detect the Denial-of-Service attack in a network. Many significant works has been done to prevent the DoS attack on the network through various algorithms and techniques. The approaches discussed here gives the advantages and disadvantages of each method.

Ramachandran, A. et al. (2006) proposed a method to analyze the network-level behavior of the spammers. It includes the IP address ranges of the system that send most spams, common spamming modes on how persistent each spamming host is, botnet spamming characteristics, and techniques for harvesting email addresses. Spammers appear to make use of transient bots that send only a few pieces of email over the course of a few minutes at most. Then spammers briefly announce IP address space through they send spam and withdraw the routes to that IP address space once the spam is sent, in order to remain untraceable.

Xie, M. et al. (2006) proposed a method of laundering email spam through open-proxies or compromised PCs that is a widely used trick to conceal real spam sources and reduce spamming cost in the underground email spam industry. It reveals one salient characteristic of proxy-based spamming activities, namely packet symmetry, by analyzing protocol semantics and timing causality. DBSpam is used to break spam laundering.

4.3 WORKING OF SPAM ZOMBIE DETECTION SYSTEM

The spam zombie detection system formulates the problem in network model and the assumptions made in it. It uses the total number of outgoing messages from a specific system in the network. The messages are assumed to be originated from any system inside the network. The system can either be normal or compromised machine. Here the focus is only on the compromised machine. For managing diverse distributed data centre infrastructures, use OpenNebula, a hyper-converged infrastructure platform. In order to create private, public, and hybrid implementations of Infrastructure as a Service, the OpenNebula platform maintains the virtual infrastructure of a data centre. Data centre virtualization and cloud deployments based on the KVM hypervisor, LXD/LXC system containers, and AWS Firecracker microVMs are the two main applications of the OpenNebula platform. The platform has the ability to provide the cloud infrastructure required to run a cloud on top of already installed VMware infrastructure. Early in June 2020, OpenNebula announced the launch of two editions: the Community Edition and a new Enterprise Edition for business users. Free and open-source software known as OpenNebula CE was made available under Apache License version 2. Access to upkeep is provided without charge with OpenNebula CE. releases, but only users with non-commercial deployments or users who have made significant contributions to the OpenNebula Community are eligible for upgrades to new minor/major versions. OpenNebula EE needs to be subscribed to commercially and is offered under a closed-source licence.

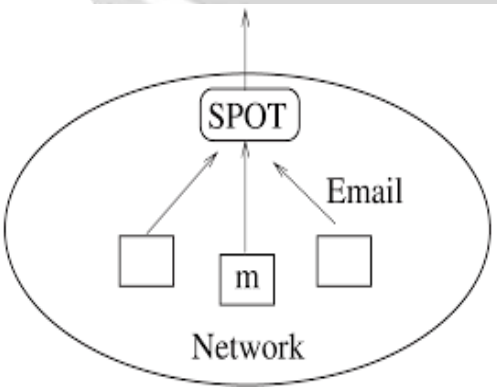

FIGURE 2 NETWORK MODEL

Three algorithms are used by the detection system to find the compromised system in a network. The first algorithm is based on the proportion of spam messages that come from internal machines or are forwarded from them. The second is calculated based on how many spam messages come from or are forwarded by internal machines.The third is based on a statistical method called SPRT. Importantly, SPOT analyses the total number of messages sent by a machine rather than only analyzing the rate at which they are sent to thwart spammers from purposely slowing the rate of message transmission in order to work around the system. The SPOT system enables individual networks to globally monitor computers on their networks and to automatically and accurately detect and efficiently remove compromised computers from their networks in an online manner. This novel detection method is applicable to a wide range of settings in which computer networks play an essential role.

4.4 Software Used

Nebula Sunstone

For managing diverse distributed data centre infrastructures, use OpenNebula, a hyper-converged infrastructure platform. In order to create private, public, and hybrid implementations of Infrastructure as a Service, the OpenNebula platform maintains the virtual infrastructure of a data centre. Data centre virtualization and cloud deployments based on the KVM hypervisor, LXD/LXC system containers, and AWS Firecracker microVMs are the two main applications of the OpenNebula platform. The platform has the ability to provide the cloud infrastructure required to run a cloud on top of already installed VMware infrastructure. Early in June 2020, OpenNebula announced the launch of two editions: the Community Edition and a new Enterprise Edition for business users. Free and open-source software known as OpenNebula CE was made available under Apache License version 2. Access to upkeep is provided without charge with OpenNebula CE. releases, but only users with non-commercial deployments or users who have made significant contributions to the OpenNebula Community are eligible for upgrades to new minor/major versions. OpenNebula EE needs to be subscribed to commercially and is offered under a closed-source licence.

4.5 Hardware RequirementLaptop or PC

- I3 processor system or higher
- 4 GB RAM or higher
- 100 GB ROM or higher

4.6 Software RequirementLaptop or PC

- Windows 7 or higher
- Open Nebula Sunstone

5 .Modules

5.1 MODULE DESCRIPTION

The content-based spam detection system is divided into four major modules. They are:

1. Mail server creation and Mail composing
2. Finding the compromised machines using SPOT
3. Preprocessing and Clustering
4. Spam detection

a) Mail server creation and Mail composing

The project's first portion is dedicated to building mail servers. The mail server was initially developed so that users may connect with one another via email. The mail server environment offers the composer option, which enables users to send emails to any recipient. This allows users to receive emails from a variety of receivers. It includes the inbox's received mail. Also, it offers the option to view sent, spam, and deleted emails.

In the user registration section, the users have to create an account to transfer the mail to various recipients. Once the registration is completed successfully user can login to send mails and receive mails from various users. During the registration process the details about the individual users are gathered and stored into the server. Each and every time user login to the server and check whether the authenticated users login or unauthorized users processing.

In mail composing section, user abstraction is built. The user can send the data to the receiver using the user's mail id, subject and content of the information or attachments of the information and can be able to send the data efficiently. The user can view the list of mails sent by the user in the sent mail folder. The mails that are reported as spam can be viewed in the spam folder.

b) Finding the compromised machines using SPOT

SPOT is the spam zombie detection system, uses SPRT method to find the compromised machine. It analyses the total number of messages sent by a machine rather than only analyzing the rate of message transmission of the system. This system enables individual networks to globally monitor computers on their networks and to automatically and accurately detect and efficiently remove compromised computers from their networks in an online manner.

c) Preprocessing and Clustering

The first stage is to take into account the text-based data present in freely formatted text documents. At the beginning, preprocessing is carried out using spam documents that have already been created. The first step is to eliminate any stop words that provide unnecessary information. Certain verbs, conjunctions, disjunctions, pronouns, etc., as well as stemmed words are among them.

Based on words or concepts defined in the text, a suitable representation of the data is provided. At this stage of information processing, many data representation techniques are applied, such as term frequency and inverse document frequency techniques.

The technique of arranging data or information into groups of comparable types using some physical or quantitative measures is known as clustering. The K-nearest Neighboring algorithm operates under the assumption that the training samples are closest to the new or arriving instance.Here the data are grouped and spam data are classified for identifying spam message.

d) Spam detection

Preprocessing, clustering, and SPOT are all performed by the server when a new email is composed in order to determine whether the user is sending spam. If the content is determined to be spam, the server will reject it, preventing the message from being sent to the recipient and

blocking the content. By doing this, network zombie attacks and bandwidth are reduced.

6.IMPLEMENTATION

6.1 DATABASE STRUCTURE

Database design is crucial to effectively implement a relational database. In this design the parts of a database, data modeling, database construction, and developing a database in a business environment are explained. These skills are fundamental to managing database backed websites, or any relational database application. Database design is a complex subject, no matter how easy some people think it is. A properly designed database is a model of a business, or some 'thing' in the real world.

Column Name	Data type	length	Allow Nulls	Description
Mailed	int	1000	notNull	Mail id
Mailfrom	varchar	50	Null	Mail from
mailtime	datetime	100	Null	Mail time
Mailsub	nvarchar	200	Null	Mail subject
Mailbody	nvarchar	500	Null	Mail body

TABLE 6.2 USER DRAFT TAB

Column Name	Data type	length	Allow Nulls	Description
Mailfrom	varchar	50	notNull	Mail from
mailto	varchar	50	notNull	Mail to
mailtime	datetime	1000	Null	Mail time
mailsub	nvarchar	50	Null	Mail subject
mailbody	nvarchar	500	Null	Mail body

TABLE 6.3 MAIL TABLE

Column Name	Data type	length	Allow Nulls	Description
mailid	int	1000	notNull	Mail id
mailfrom	varchar	50	Null	Mail from
mailsub	nvarchar	200	Null	Mail subject
mailbody	nvarchar	500	Null	Mail body
Upload	varchar	200	Null	User uploads
mailtime	datetime	1000	Null	Mail time
mailattach	varchar	100	Null	Mail attachments

TABLE 6.4 ADMIN INBOX TABLE

Column Name	Data type	length	Allow Nulls	Description
fname	varchar	50	Null	User Firstname
lname	varchar	50	notNull	User Lastname
logname	varchar	50	Null	Login Name
logpass	varchar	50	Null	Login Password
secQues	Varchar	1000	Null	Secret Question
secAns	varchar	1000	Null	Secret Answer

TABLE 6.5 USER LOGIN DETAILS

Column Name	Data type	length	Allow Nulls	Description
mailed	int	1000	notNull	Mail id
mailfrom	varchar	50	Null	Mail from
mailtime	datetime	1000	Null	Mail time
mailsub	nvarchar	200	Null	Mail subject
mailbody	nvarchar	500	Null	Mail body
upload	varchar	200	Null	User uploads
mailattach	varchar	200	Null	Mail attachments

TABLE 6.7 ADMIN SPAM TABLE

Column Name	Data type	length	Allow Nulls	Description
id	Int	1000	notNull	User id
mailid	varchar	50	notNull	Mail id
mailtime	datetime	1000	Null	Mail time
msubj	nvarchar	200	Null	Mail subject
mbody	nvarchar	500	Null	Mail body
upload	varchar	200	Null	User uploads
mailfrom	varchar	50	Null	Mail from

TABLE 6.8 USER SPAM TABLE

CO

7.CONCLUSION AND FUTURE WORK

An effective content-based spam detection technique is designed, that will discover the compromised machines that are involved in the spamming activities. The implementation is also extended to detect the source of spam messages called the spam zombies using SPRT technique. The combination of these methods improves the efficiency of the system and improves the accuracy of detection of spam messages. This method also reduces the number of required observations made to discover the compromised machines. Thus is reduces the flooding of the network and improves the network performance.

The future work includes the implementation of parameter based approach. The content-based approach gives an efficient result, but the processing time is proportionate to the number of messages. Thus the parameter based approach can be used

8.REFERENCES

[1] BURKERT, PETER, ET AL. “AN INTRUSION DETECTION SYSTEM” ARXIV:1509.05371V2, ARXIV.ORG/PDF/1509.05371.PDF.

[2] "CHALLENGES IN REPRESENTATION LEARNING: A REPORT ON THREE NETWORKS CONTESTS." I GOOD FELLOW, D ERHAN, PL CARRIER, A COURVILLE, M MIRZA, B HAMNER, W CUKIERSKI, Y TANG, DH LEE, Y ZHOU, C RAMAIAH.

[3] CHEN, JASON, ET AL. SPAM DETECTION FROM STATIC IMAGES. STANFORD UNIVERSIT DEPARTMENT OF COMPUTER SCIENCE, CS229.STANFORD.EDU/PROJ2016SPR/REPORT/026.PDF.

[4] DACHAPALLY, PRUDHVI RAJ. "INTRUSION DETECTION AND REPRESENTATIONAL AUTOENCODER UNITS." ARXIV:1706.01509, ARXIV.ORG/ABS/1706.01509.

[5] DUMAS, MELANIE. "AN INTRUSION DETECTION SYSTEM TO PREVENT K-ZERO DAY ATTACK." MACHINE PERCEPTION LAB, UNIVERISTY OF CALIFORNIA, 2001. [6] F FENG, R LI, X WANG, D ATHANASAKIS, J SHAWE-TAYLOR, M MILAKOV, J PARK, R IONESCU, M POPESCU, C GROZEA, J BERGSTRA, J XIE, L ROMASZKO, B XU, Z CHUANG, AND Y. BENGIO. ARXIV 2013.