ISSN:0975-9646

Hybrid Solution in Honeypot Mechanism using Hadoop

Girija Srikanth[1,] R.Sathish[2]

[1] *Assistant Professor, Computer Science & Engineering, B.S.Anangpuria Institute of Technology and Management, Faridabad, Haryana, India,*
[2] *Consultant, Capgemini India pvt Ltd, Bangalore,*
Karnataka, India

Abstract -Achieving computer system security is one of the most popular and fastest Information Technology in organization. System security personnel fight a seemingly unending battle to secure their digital assets against an ever increasing onslaught of attacks. Protection of information availability, its access and data integrity are the basic security characteristics of information sources. Any disruption of these properties would result in system intrusion and the related security risk. Advanced decoy based technology called Honeypot has a huge potential for the security community and can achieve several goals of other security technologies, which makes it almost universal. Honeypots- A security resource, whose value lies in being probed, attacked, or compromised, provides a valuable tool to collect information about the behaviors of attackers in order to design and implement better defenses. This review paper is based on an idea of dynamically creating, modifying and managing virtual honeypots. This system combines the concept of honeypots and uses big data analyzer, Hadoop for quick information retrieval and support vector machine. The goal of this proposed system is to create evanescent honeypots at right places and times, on demand

Keywords

Honeypots, support vector machine, hadoop, client-server architecture

1. Introduction

Computer security is among one of the main areas of information technology. Over recent years mentioned area achieves the biggest progress because nobody wants that exactly his system will be attacked and intruder or anybody else will receive the stolen data. Whichever more experienced attacker can exploit weaknesses in the security system and penetrate through its defense mechanism to obtain sensitive data. It's necessary to put high priority to system security, minimize vulnerabilities and secure the computer system against intrusion type of attackers; this technology is called as Honeypot. A physical honeypot is a real machine with its own IP address. Deploying a physical honeypot is often time intensive and expensive as different operating systems require specialized hardware and every honeypot requires its own physical system. In this paper we are focusing more on identifying intrusion detection by using Support Vector Machine (SVM). Hadoop is a "flexible and available architecture for large scale computation and data processing on a network of commodity hardware" It is an open source framework for processing, storing and analyzing massive amounts of distributed unstructured data. It was designed to handle petabytes and Exabyte's of data distributed over multiple nodes in parallel. Hadoop clusters run on inexpensive commodity hardware so projects can scale-out without breaking the bank.

2. Analysis Of Problem

Consider two systems A and B in some network System B was found to be important and had its equivalent honeypot B'. System A did not have its

equivalent honeypot. If an attacker tries to exploit A without falling for honeypot B', the main purpose of having a honeypot in the network is unused. It is expensive to maintain honeypots that yield us no information whatsoever. It is imperative to maintain only those honeypots that could be potential targets for the attacker. Had there been a honeypot for A, it could have provided us a great deal of information.

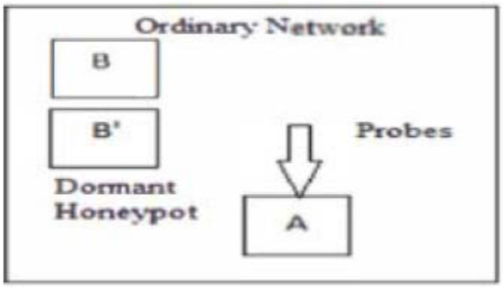

Figure 1: Honeypot Deployed in an ordinary network

Support Machine Vector (SVM) Support Vector Machines (SVM) is the classifiers which were originally designed for binary classification. The classification applications can solve multi-class problems. Decision-tree-based support vector machine which combines support vector machines and decision tree can be an effective way for solving multi-class problems. This method can decrease the training and testing time, increasing the efficiency of the system. The different ways to construct the binary trees divides the data set into two subsets from root to the leaf until every subset consists of only one class. The construction order of binary tree has great influence on the classification performance. In this paper we are using an algorithm, Tree structured multiclass SVM, which has been used for classifying data. This work proposes the decision tree based algorithm to construct multiclass intrusion detection system. If binary SVMs are combined with decision trees, we can have multiclass SVMs, which can classify the four types of attacks, Probing, DoS, U2R, R2L attacks and Normal data, and can prepare five classes for anomaly detection. This paper's aim is to improve the training time, testing time and accuracy of IDS using the hybrid approach.

Hadoop is an open-source software framework for storage and large-scale processing of data-sets on clusters of commodity hardware. Hadoop is an Apache top-level project being built and used by a global community of contributors and users. It is licensed under the Apache License 2.0.

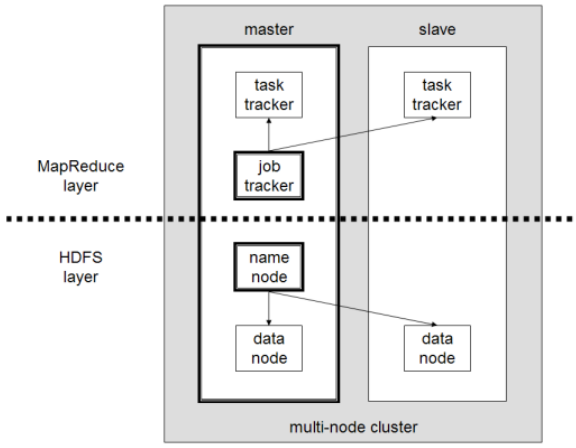

Figure 2: Hadoop

3. Literature Review And Related Work

Honeypots are usually deployed with the intent of capturing interactions with unsuspecting adversaries. In 2011, John P. John, Fang Yuet et al., Heat-seeking Honeypots: Design and Experience proposed that the captured interactions allow researchers to understand the patterns and behavior of attackers. For example, honeypots have been used to automate the generation of new signatures for network intrusion detection systems, collect malicious binaries for analysis, and quantify malicious behavior through measurement studies.

In 2001, L. Spitzner, “The Value of Honeypots, Part One: Definitions and Values of Honeypots,” proposed that it gives actual idea and some definitions of honeypots. For the purposes of this paper, they have defined a honeypot as "a resource whose value is being in attacked or compromised". This means that whatever we designate as a honeypot, it is our expectation and goal to have the system probed, attacked, and potentially exploited. Honeypots are not a solution; they do not 'fix' anything.

In 2003, L. Spitzner, “Honeypots: Tracking Hackers,” Boston, USA: Addison-Wesley, Parson Education, ISBN 0 321-10895-7, here they advocate the use of honeypots as an effective educational tool to study issues in network security. They support this claim by demonstrating a set of projects that they have carried out in a network, which they have deployed specifically for running distributed computer.

In 2003, BAIT-TRAP, proposes the design and implementation of BAIT-TRAP, a

Catering honeypot architecture. By carefully monitoring network activities, BAIT-TRAP dynamically identifies “bait” services and automatically composes “attractive” honeypots in order to capture the expected attacks. Within seconds, a newly composed honeypot will be automatically deployed and exposed to potential attackers.

In 2013, Vidyasagar.S.D proposed a research a paper on a Study on “Role of Hadoop in Information Technology era” and he proposed the Hadoop platform was designed to solve problems where you have a lot of data perhaps a mixture of complex and

structured data and it doesn’t fit nicely into tables. It’s for situations where you want to run analytics that are deep and computationally extensive, like clustering and targeting. That’s exactly what Google was doing when it was indexing the web and examining user behavior to improve performance algorithms. This article has made an attempt to study its need, uses and application, thereby brought to the notice of the readers.

4. Overview

The proposed architecture uses a sophisticated hybrid Honeypot with an autonomous feature as an IDS detection mechanism. Solution for minimizing failures in the detection process and collection of important data based on Honeypot consists of a combination of security tools: Snort IDS, Sebek and Dionaea. Tools were selected based on their properties analysed above. The detection mechanism based on a sophisticated hybrid Honeypot integrated in the client-server architecture consisting of centralized main server and multiple client stations. Client workstations serve to capture suspicious activity or directly record the malicious code which is then send to server for processing. Server analyses received data, decides to issue or not to issue a security warning and displays cumulative information through a web interface. This proposal aims to provide a solution of early warning against any attack on the computer system.

Server Architecture Due to the centralization of collected data is main server at the same time connected to multiple clients and is set to receive all incoming messages which are then stored in the knowledge database. Cohesion of individual reports indicated attackers’ intention to attack aimed computer system areas with widespread attacks or full-range scanning.

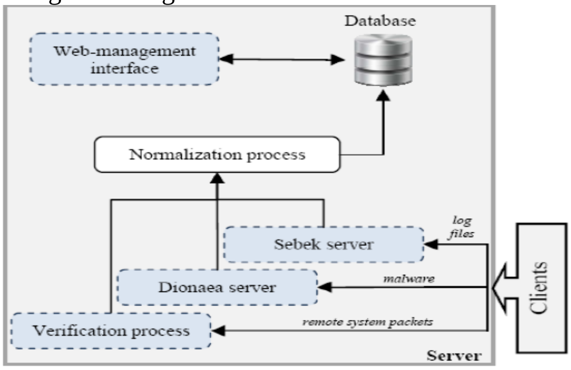

Figure 3: Server Architecture

The proposed server architecture consists of three main parts, which data are normalized before storing to the database:

- Sebek server – at the same time receives and filters several data sources representing instructions or a connection to incoming data storing process.
- Dionaea server – accepts patterns of malicious code that sends the dionaea client part.

• Verification process – a modular scheme of hybrid open-source system for intrusion detection. It's using standard communication format. It can be adapted to the needs of an extensive system from any point of deployment, receives the amount of data from clients and integrates diversified data formats. Web-server interface displays all information about captured attack.

Client Architecture Because of gathering data about attacker activities during an attack are installed clients placed in the same domain. Various parts of the system are independently activated for collecting set of data depending on attack type. Obtained data are subsequently backward delivered to a server to facilitate further analysis and for the subsequent updating system security.

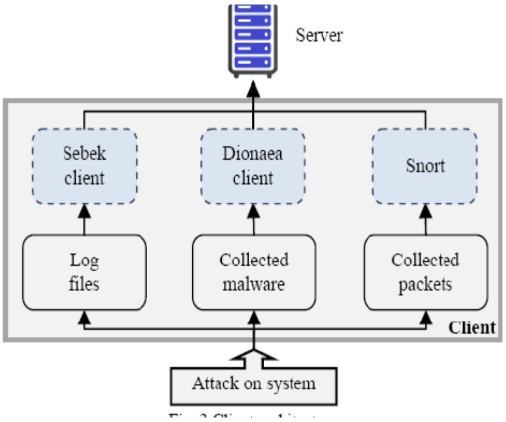

Figure 4: Client Architecture

Client architecture consists of three components/tools:

• Sebek client – records attacker behaviour during interaction with the Honeypots in log files.

• Dionaea client – attracts attackers and captures the patterns of malware by simulating basic system services and vulnerabilities.

• Snort – monitors and filters packets during detecting intrusions, Identifies patterns of separate attacks, information and warning messages.

5. HADOOP ARCHITECTURE

Hadoop is designed to run on a large number of machines that don't share any memory or disks. That means you can buy a whole bunch of commodity servers, slap them in a rack, and run the Hadoop software on each one. When you want to load all of your organization's data into Hadoop, what the software does is bust that data into pieces that it then spreads across your different servers. There's no one place where you go to talk to all of your data; Hadoop keeps track of where the data resides. And because there are multiple copy stores, data stored on a server that goes offline or dies can be automatically replicated from a known good copy. In a centralized database system, you've got one big disk connected to four or eight or 16 big processors. But that is as much horsepower as you can bring to bear. In a Hadoop cluster, every one of those servers has two or four or eight CPUs. You can run your indexing job by sending your code to each of the dozens of servers in your cluster, and each server operates on its own little piece of the data. Results are then delivered back to you in a unified whole. That's Map Reduce you map the operation out to all of those servers and then you reduce the results back into a single result set. Architecturally, the reason you're able to deal with lots of data is because Hadoop spreads it out. And the reason you're able to ask complicated computational questions is because you've got all of these processors, working in parallel, harnessed together. Hadoop implements a computational paradigm named Map/Reduce, where the application is divided into many small fragments of work, each of which may be executed or re-executed on any node in the cluster.

Hadoop Users: The following companies are the users of Hadoop adobe: Alibaba, Amazon, AOL, Facebook, Google, and IBM.

Major Contributors: The following companies are the major contributors of Hadoop. They are Apache, Cloudera and Yahoo.

6. APPLICATIONS

1**. *Intrusion Detection*:** Intrusion Detection is the art of detecting inappropriate, incorrect, or anomalous
activity. The proposed system can be used to determine if a computer network or server has experienced an unauthorized intrusion.

2. Social Networking: Web-based social systems enable new community-based opportunities for participants to engage, share, and interact. This community value and related services like search and advertising disseminators. In an effort to preserve community value and ensure long-term success, we can use proposed for uncovering social spammers in online social systems.

3. Network Forensics: Network forensics deals with the capture, recording and analysis of network events in order to discover evidential information about the
source of security attacks in a court of law. Using this
system we can gather intelligence about the enemy and the tools and tactics of network intruders.

4. Campus Net Security: With the development of digital campus construction, the campus network size has been rapid growth, but there are also many network security problems. If this is applied to the campus network it can make the security of campus network unobstructed.

7. ADVANTAGES AND DISADVANTAGES

1. Minimal resources
   Captures only malicious activity in the system. For its functionality is enough an equipment with low end system parameters.
2. Simplicity
   Honeypots are simple and flexible. For their functionality they do not require complicated algorithms or other complex operations.
   4. Discovering new tools & tactics
   Capture and record everything that interacts with them.
   5. Small data sets
   Produces small amount of data, but it can be a high quality.
5. Reduce false positives and false negatives.

Disadvantages

1. Risk of takeover
   After gaining control over the Honeypot attacker can retrieve all the collected data.
2. Disclosure of identity
   Honeypot has expected characteristics and behaviour. Experienced attacker can detect presence of incorrectly configured decoy in system.

8. Conclusion And Future Work

Honeypots becoming highly-flexible solution, Not only their deployment and management become more cost-effective, but also provide a much better integration into the system, thereby minimizing the risk of human error during manual configuration. Merger with the surrounding system in addition minimizes the risk of identification by attackers. Just as all new technology, the decoys also have some shortcomings that need to be overcome and eliminated. Honeypot is excellent security tool but it is not a panacea for a securing the whole system. The apart of this work is improving the IDS detection mechanism and minimizing the number of generated false positives and also false negatives using advanced technology called Honeypot. The work includes proposal of an autonomous special safety feature using SVM for enhancing security of distributed computer systems. Unique proposal combines a variety of security tools, to order to minimize their disadvantages and maximize the security capabilities in the process of intrusion detection. Honeypots with Hadoop can be found to be more efficient as compared to the conventional honeypot deployment. Standard honeypot deployment yields productive information only if it is explicitly probed or fiddled with by the attacker. This, on the other hand, promises useful data irrespective of the system on the network being targeted. This system would greatly benefit the entire computing community at large. Information security is an unending battle to safeguard our digital assets.

No security mechanism can be classified as 'foolproof' as newer and stronger attacks are being discovered. Honeypots with Hadoop would enable us to get into the attacker's mind to some extent and bolster our defenses.

Acknowledgment

We would like to acknowledge the support of “Techlister Technologies Pvt Ltd.”. We also very grateful to Mr. Pelister, Founder who has given full support.

References

[1] Lance Spitzner,” Honeypots: Definitions and value of Honeypots.”http://www.tracking-hackers.com.

[2] John P. John, Fang Yuet et al.,” Heat-seeking Honeypots: Design and Experience”. In Proceedings of WWW 2011-Session Web Security, 2011.

[3] Christopher Hecker, Kara L. Nance, and Brian Hay, ASSERT Centre, University of Alaska Fairbanks. Dynamic Honeypot Construction. In proceedings of the 10th Colloquium for Information Systems Security Education University of Maryland, University College Adelphi, MD June 5-8, 2006.

[4] L. Spitzner, 2002,” Honeypots tracking Hackers.” lsted. Boston, MA, USA: Addison Wesley.

[5] The Bait and Switch Honeypot, http://www.violating.us/projects/baitnswitch/

[6] The Honeynet Project, http://www.honeynet.org.

[7] L. Spitzner,” Dynamic Honeypots”, http://www.securityfocus.com/infocus/1731, Sept. 2003.

[8] BAIT-TRAP, http://www.cs.purdue.edu/homes/jiangx/BaitTrap, Dec. 2003.

[9] Research paper on A Study on “Role of Hadoop in Information Technology era” by Vidyasagar S.D.

[10]” A Virtual Honeypot Framework” by Neils Provos, Google, Inc.

Authors

Girija Srikanth, working as an Assistant Professor in CSE department in B.S.Anagpuria Institute of Technology and Management, Faridabad, Haryana.

She completed her B.E in Electronics and Communication Engineering and M.Tech in Information Security in CSE department. She submitted papers in 4 International journals with good impact factors and 2 International conferences. She is having membership in International Association of Engineers(IAENG).Her Areas of interest include Network security, RFID, Image Processing, Cyber security.

Sathish R, Currently Working as a Senior Software Engineer in Capgemini India Pvt. Ltd. Bangalore. He completed his Master Degree in Computer Application (MCA) in NITc. He submitted papers in 2 international journals and 2 international conferences His Area of interest include Networking and Application Security mainly in Middleware Stack.