Cloud Based Security Center: To Protect Networking Attack by Forensic Scrutiny

Sankara Mahalingam M
PG Student, CSE, Kalasalingam Institute of Technology, Krishnankoil.

Abstract - **Internet security problems are still a big challenge as there are many security events occurred, such as Internet worms, Spam and phishing attacks etc. Botnet, a well-organized distributed network attack, consists of a large volume of bots, which generates huge volumes of spam or launching Distributed Denial-of-Service (DDoS) attacks to victim hosts. To address these problems, a practical Collaborative Network Security Management System is proposed with well- deployed collaborative UTM (Unified Threat Management) and traffic probers. In this paper, we propose a design and implementation of cloud based Security Center for network security forensic scrutiny. We propose to use cloud storage to keep collected traffic data and processing it with cloud computing platform to find the malicious attacks. A workable case, phishing attack forensic analysis is presented and the required computing and storage resources are evaluated based on real trace data.**

Index Terms - **Cloud Computing, Collaborative Network Security System, Computer forensics, Anti-Phishing, Hadoop File System, Eucalyptus.**

1. INTRODUCTION

As Internet, security problems are still a big challenge as there are many security events occurred. The underground economics based on Internet Scam and Fraud is booming. These attackers initiate more and more E-crime attacks and abuse, such as Spams, Phishing attack, Internet worms etc. Firewalls, Intrusion Detection System (IDS) and Anti-Virus Gateway are now widely deployed in edge-network to protect end-systems from the attacks. When the malicious attacks have fixed patterns, they can be easily identified and matching these patterns. For example, the Distributed Denial of service (DDoS) contains very few, if any, signatures strings to identify. Nowadays DDoS attacks are likely launched by a large volume of botswhich forms a Botnet controlled by bot master. The bots are commanded to generate attack new victim machine and enlarge botnet. The bots also commanded to conduct other issues such as disseminating spam or launching Distributed Denial-of-Service (DDoS) attacks to victim hosts. To countermeasure botnet, secure overlay is proposed. To prevent distributed attacks, collaboration is needed, so we used Collaborative Network SecurityManagement System (CNSMS).

The CNSMS aims to develop a new collaboration system to integrated well-deployed UTM such as NetSecu. Such distributed security overlay network coordinated with a centralized Security Center leverage a Peer-to-Peer communication protocol used in UTM's collaborative module and virtually interconnect them to exchange network events and security rules.

In this paper, we evaluate cloud-based solution in Security Center for traffic data forensic analysis. The main contribution of our paper is that we propose a practical solution to collect data trace and analyze these data in parallel in a Cloud Computing platform. We propose to use cloud storage to keep huge traffic data and processing it with cloud computing platform to find the malicious attacks. As we already operate Collaborative Network Security Management System which hasbig data output. A workable case, phishing attack forensic analysis is presented and the required computing and storage resourceare investigated. We have concluded that this phishing filter functions can be effectively scale to analyze a large volume of trace data for phishing attack detection with Cloud computing.The results also show that this solution is economical for large-scale forensic analysis for traffic data.

2. SYSTEM DESIGN AND IMPLEMENTATION

A.CNSMS

Collaborative Network Security Management System (CNSMS) deployed in Multisite deployment. These sites are all managed by Collaborative Network SecurityManagement System in Security Center over Internet.

During the system's operating, the collaborative mechanism runs as we expected to share security events and rulesets, and new rule sets are enforced on demands as instructed by Security Center. Operating reports from each NetSecu node and Prober have been collected and send back to Security Center. Also there are a lot of network security events have been observed and recorded in the deployment, such as DDoS reflect attacks,

Manuscript received February 5, 2015
Manuscript revised February 20, 2015

Spam scatter and ad hoc P2P protocols etc. An information control cycle, which divides several steps, which is Collaborative UTM and Prober acts as sensors and report the security events and traffic data to Security Center. The Security Center aggregates all the events and digs into the collected traffic data. After a detailed analysis and with the assistance of expertise manager, Security Center generates new policy or ruleset to disseminate to each collaborative UTM and Prober for enforcement, and receive the feedback information. The traffic probe is the building block for recording the raw internet traffic at connection level. This can be designed to focus on specific traffic occasioned by centain secutiry events when needed.

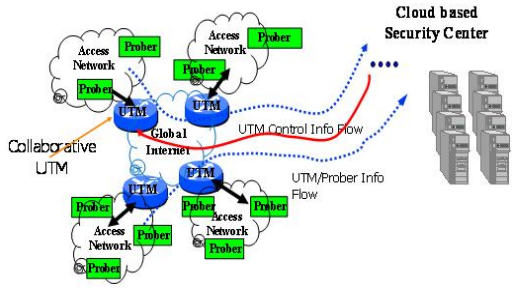

Figure 1. The deployment of Collaborative Network Security Management System

1) Collaborative UTM: Acted as collaborative UTM, NetSecu is introduced in Ref[1]. A NetSecu node consists of the following features:

- Incrementally deployable security elements, dynamically enable/disable/upgrade security functions;
- Policy-instructed collaboration over the Internet.
- NetSecu node contains Traffic Prober, Traffic Controller, Collaborator Element, and Reporting Element to fulfill the above design goals.

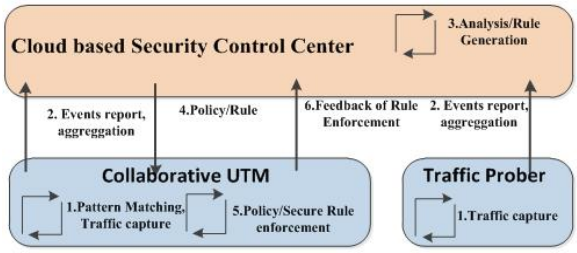

Figure 2. The work principle of CNSMS with Cloud based Security Center.

2)Security Center: Collaborative Network Security Management System (CNSMS) is proposed in [2] and operated in Security Center. As NetSecu nodes could manage security problems in a subdomain and provide P2P communication interfaces, CNSMS orchestrates the communication between these NetSecu nodes. More specifically, CNSMS will achieve the following objectives:

- Security policy collaborative dissemination and enforcement;
- Security event collaborative notification;
- Security ruleset dissemination, enforcement and update;
- Trust infrastructure;
- Scalability.

3) Botnet Control: A botnet is a collection of Internet-connected programs communicating with other similar programs in order to perform tasks. It could be used to send spam email or participate in distributed denial-of-service attacks.

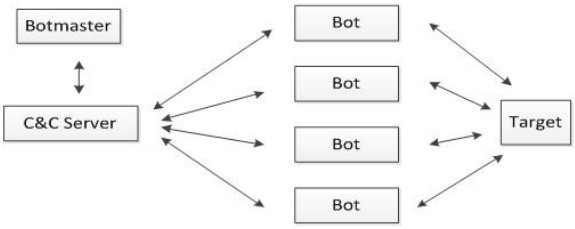

Figure 3: Botnet structure

A typical distributed attack is Botnet, which is extremely versatile and are used in many attacks, for example, sending huge volumes of spam or launching Distributed Denial-of-Service (DDoS) attacks. The work principle of botnet is shown in Figure 3 Suppressing botnets become more and more difficult. There are many reasons, firstly, the Botmaster will keep their own botnets as small as possible not only to hide themselves but also to rent the botnets in an easy way, secondly, bots can automatically change their command and control server (C&C) in order to hide and rescue themselves.

B.Cloud based Forensic Analysis in Security Center

1) Cloud Storage and Computing platform: We focus on the traffic data storage and forensic analysis. The underground cloud storage and computing platform is based on Hadoop and Eucalyptus Cloud Computing. We also give some analysis the use of Cloud Computing platform based on Eucalyptus and Amazon EC2 respectively.

2) Cloud Storage with Hadoop: The Hadoop file system with version 1.0.1 is used for Cloud storage system of collected traffic. The master node is acted as namenode, secondarynamenode, jobtraker, Hmaster, and other node is working as datanode, tasktracker, regionserver. There are totally 4 racks of machines with 5,5,4,4 in each rack. There are 18 slave nodes in total. As the Hadoop system is used for traffic analysis. The traffic collected in individual collaborative UTM is aggregated and uploaded to this cloud platform. Each node has an Intel

four cores CPU with 800MHz, and Memory size is 3GB, and with a 320G HardDisk. We test two scenarios where we write 18 files with each size 300MB and 36 files with each file size 100MB.

C. Cloud Computing IaaS Platform

1)Cloud Computing based on Eucalyptus: In Eucalyptus's term, there is one cloud controller, and the others are compute nodes. Cloud controller acts as the computing portal, task assigner and result aggregation. There is computing instance affiliated with each compute node. In our usage scenario, we run 4 VM instances in each compute node, hence there about 24 running instances simultaneously. Each computing instance runs the pipeline divided into the following phases: data fetcher, data processing, and posting computing results. By this method, we can achieve best working efficiency of hardware and software resource's usage.

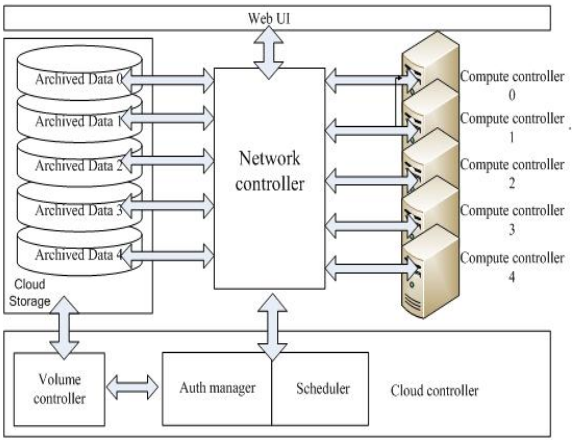

Figure 4. The Cloud Computing Platform based on Eucalyptus

2) Cloud Computing based on Amazon: Amazon EC2 and S3 are used for comparative analysis. The main purpose to use Amazon service is with comparing purpose to our home-brewed Eucalyptus system. As the consideration of user privacy and legal issues, we conduct anonymization processing the data and upload the amazon S3 service.

3) Forensic Analysis of Phishing Attack: Phishing is an intriguing practical problem due to the sensitive information stolen (e.g. monetary user account name and password) and estimated about billion loss in accumulation annually. Not only the users but also the backing financial institutions such as e-banks and e-pay systems have been impaired by phishing attacks.

There is already much research works to countermeasure phishing attacks. To protect web browser user from phishing attacks, plugins to compare visited URL with blacklist URL are already provided by mainstream web browsers. Google also provide safe Browser API for check a URL in Google collected phishing database. It even makes it worse because of the un-awareness of this phishing attack for most of innocent Internet users.

Gregor Maier et al. [3] propose a traffic archiving technology for post-attack analysis in Bro IDS. Using Timemachine, the network trace data is archived and can be feed back to the IDS with current knowledge of modern attacks to find the forensics of attacks was undiscovered in that time. K. Thomas et al.proposed Monarch system [4] for real-time URL spam filtering for tweets and spam mails stream. Compared with Monarch, we put emphasis on phishing forensics analysis of large volume of offline trace with Cloud Computing platform.

With similar idea, we proposed an offline phishing forensic collections and analysis system. This system targeted to solve the following challenging problems:

- *How to collect the original data to search the phishing attack forensics wherein;*
- *How to handle the huge volume data in a reasonably short time.*

Cloud computing platform is used for offline phishing attack forensic analysis. Firstly, our CNSMS collect the network trace data and report to Security Center. Secondly, we have both constructed an IaaS cloud platform and use the existing cloud platform such as Amazon EC2 and S3 for comparabale reason. All phishing filtering operation is based on Cloud Computing platform and running in parallel with "divide and conquer scheme".

1) Data trace collection: Our trace data is an un-interruptible collection about half yearwith multiplevantage points with UTM's deployment. The total size of traffic passed through our vantage points is about 20 TB. The total data is about 20TB and divided into 512MB data blocks. Typically, a typical 512M data block consists of about 40K URLs.

The experimental data is about 1TB when collected in a cut-off mode in a collaborative UTM. The data trace is still growing in the size during our experiments.

2) Data anonymization: To protect user's privacy and avoid legal issues in the research, the trace data is anonymized to replace IP and other user informationbefore the data processing in Amazon EC2.

3) Data processing: The data processing procedure are divided in different phases, which are shown as follows:

File splitting: Each packet capture file created by Time Machine is 512 MB, and is further divided into smaller parts for processing by using tcpdump. This is due to the amount of memory used during the extraction of data from TCP streams will exceed the maximum physical memory.

TCP stream reassembly: This stage is to restore the TCP streams in the captured pcap files using tcptrace.

URL extraction: After extracting data from TCP streams, grep is used to find all URLs contained in the data by searching for lines starting with "Referer: http://".

URL check: URLs found are stored in a file to be checked for phishing by using Google Safe Browsing API [9]. In order to check URLs for phishing sites, we use phishing site data provided by Google. Google provides the first 32 bits of phishing sites' SHA256 values for users to use. If a match is found between a URL's SHA356 value is found, the full 256 bits hash value is sent to Google to check the site. More details on data provided by Google can be found in Google Safe Browsing API's documentation. During the process of comparing URLs' hash values, a prefix tree is used for matching because the data provided by Google is only 32 bits long and a prefix tree can do the matching of a URL's SHA256 value with Google's data in O(1) time.

Result reporter: This stage collects the final results in different machine, and aggregate the final report.

3. EXPERIMENTS RESULTS

We conduct our evaluationexperiment both on Eucalyptus and Amazon AWS for the comparison purpose.

A. Eucalyptus

We also run the phishing data block processing task in home-brewed Eucalyptus platform with Intel Core 2 Quad Processor with 1.333 GHz FSB and 2MB cache, double channel 4GB DDR3 with 1.066GHz, Intel G41 + ICH7R Chipset and Intel 82574L Network Chipset.

Time spending in different process stages in Eucalyptus platform are measured and concluded as shown in Table 1.

stage	TCP stream reassembly	URL extraction	URL check
Time (seconds)	15~20	16~20	~5

Table 1-Time spending in different stage in Eucalyptus.

It seems prefixTreecomparison's speed is quite fast and this time spending can be almost ignored. But before URL check, it need take some time to download the Google Safe Browsing signature libraries, this time spending is quite undetermined due to network status and Google servers' response latencies.

It is also needed to point out that the m1.small instance in EC2 is memory constrained without swap partition support. It will cause problems when consuming a large volume of memory (exceeding the memory usage limit) during trace data analysis.

B. Amazon AWS

Trace file processing is written in Python and executeson an EC2 small instance running UbuntuLinux 10.04.As Linux's command shows, the host CPU is Intel(R) Xeon(R) CPU E5430 @ 2.66GHz with cache size 6MB, and 1.7GB memory (with HighTotal: 982MB, LowTotal:734MB).

Different processing stage incurs different time consumption and is measured in Table 2.

stage	TCP stream reassembly	URL extraction	URL check
Time (seconds)	~287	~47	1~2

Table 2 - Time spending in different micro-stage in processing in Amazon EC2.

Compared with Amazon case, it seems that the CPU used in in Amazon instance has better performance than QX9400 quad core CPU in our physical server.

1) Estimated the number of instances: Assume the time spending in a compute instance to handle a k bytes data block in stage (2), stage (3), and stage (4) are t1,t2, t3 (in seconds) respectively. Assume there are m collaborative UTM or prober to collect traffic data, and the averagetraffic throughput is f bytes/s during the last 24 hours, and the traffic cut-off factor is h. The number of total instances L in parallel needs to handle all last 24 hours traffic iscalculated as follows:

$$T = t1 + t2 + t3 \qquad \text{(Eq. 1)}$$

$$L = (m*f*T*h)/k \qquad \text{(Eq. 2)}$$

L is also affected by several factors such as the percentages of HTTP stream in the traffic, number of URLs in HTTP streams, user's behavior in exploring web sites etc.

In theEucalyptus's case, we only run one instance in each physical server. Assume m=4, f = 100MByte/s (800Mbps) in 1 Gbps link, h =0.2 (means 20% traffic is captured), each block is 200M Bytes, T= 40 s, then the number of physical servers (or instances) in parallel is calculated as follows:

$$L = (m*f*T*h)/k = 4*100*40*0.2/200 = 16$$

In the Amazon EC2 case, T = 330s, and the number of needed EC2 m1.small instances in parallel is calculated as follows:

$$L = (m*f*T*h)/k = 4*100*330*0.2/200 = 132$$

4.CONCLUSION

The Collaborative Network Security Management System is very useful to countermeasure

distributed network attacks. Its operation resulted in big data outputs, such as network traffics, security events, etc. In this paper, we propose to use cloud computingsystems to explore the large volume of collected data from CNSMS to track the attacking events. Traffic archiving is implemented in collaborative UTM to collect all the network trace data and the cloud computing technology is leveraged to analyze the experimental data in parallel. An IaaS cloud platformis constructed with Eucalyptus and the existing cloud platform such as Amazon EC2 and S3 is also used for comparison purpose. Phishing attack forensic analysis as a workable case is presented and the required computing and storage resource are also evaluated by using real trace data. All phishing filtering operation is cloud-based and operated in parallel, and the processing procedure is also evaluated. The results show that the proposed scheme ispractical and can be generalized to forensicanalysis of other network attacks inthe future.

REFERENCES

[1] X. Chen, Beipeng Mu, Zhen Chen, NetSecu: A Collaborative Network Security Platform for in-network Security. Proc. of the 3rd International Conference on Communications and Mobile Computing (CMC), 2011.
[2] Beipeng Mu, Xinming Chen, Zhen Chen, A Collaborative Network Security Management System in Metropolitan Area Network. Proc. of the 3rd International Conference on Communications and Mobile Computing (CMC), 2011.
[3] G. Maier, R. Sommer, H. Dreger, A. Feldmann, V. Paxson, and F. Schneider, Enriching network security analysis with time travel, in Proc. ACM SIGCOMM 2008, Seattle, WA, 2008, pp. 183-194.
[4] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, Monarch: Providing real-time URL spam filtering as a service, in Proc. IEEE Symposium on Security and Privacy, Oakland, California, USA, 2011, pp. 447-462.
[5] J. Dean and S. Ghemawat, MapReduce: Simplified data processing on large clusters, in Proc. 6th Symposium on Operating System Design and Implementation (OSDI 2004), San Francisco, California, USA, 2004, pp. 139-147.
[6] Zhen Chen*, Fuye Han, Junwei Cao, Xin Jiang, and Shuo Chen, Cloud Computing-Based Forensic Analysis for Collaborative Network Security Management System, Tsinghua Science And Technology, pp40-50, Volume 18, Number 1, February 2013.
[7] http://www.antiphishing.org/crimeware.html
[8] http://www.apwg.org/
[9] http://code.google.com/apis/safebrowsing/
[10] http://stopbadware.org/

Sankara Mahalingam M is a PG scholar in department of computer Science and Engineering, Kalasalingam Institute of Technology. He is UG graduated from Syed Ammal Engineering College. His research interests include cloud computing and network security issues.