Log Visualization to track Compromised Host on the Network

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Software Engineering

by:
Sheenam Goyal
(80731022)

Under the supervision of:
Dr. Maninder Singh
Associate Professor (CSED)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

JULY 2009

Certificate

I hereby certify that the work which is being presented in the thesis entitled, **"Log Visualization to track Compromised Host on the Network"**, in partial fulfilment of the requirements for the award of degree of Master of Engineering in Software Engineering submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Maninder Singh* and refers other researcher's works which are duly listed in the reference section.

The matter presented in this thesis has not been submitted for the award of any other degree of this or any other university.

Sheenam Goyal

(Sheenam Goyal)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Maninder Singh)
Computer Science and Engineering Department,
Thapar University,
Patiala.

Countersigned by

14/07/09.

(RAJESH KUMAR BHATIA)
Assistant Professor & Head
Computer Science & Engineering Department
Thapar University
Patiala.

15/7/09

(R.K.SHARMA)
Dean (Academic Affairs)
Thapar University,
Patiala.

Acknowledgement

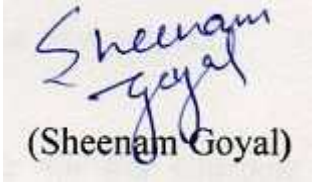
First and foremost, I would like to express my sincere gratitude to my guide **Dr. Maninder Singh,** Associate Professor, Computer Science and Engineering Department for immense help, guidance, stimulating suggestions, and encouragement all the time with this thesis work. This work would have not been possible without his encouragement. He always provided a motivating and enthusiastic atmosphere to work with, it was a great pleasure to do this thesis under his supervision.

I am equally grateful to **Dr. Rajesh Kumar Bhatia,** Assistant Professor and Head, Computer Science and Engineering Department for their appreciation and satisfactorily healing me off my inexperienced inquisitions about the new subject.

I am greatful to **Dr. R.K. Sharma**, Dean of Academic Affair for his constant encouragement that was of great importance in the completion of the thesis.

I would also like to thank all the staff members and PhD scholar Ms. Shashi Bhanwar who were always there at the need of the hours and provided with all the help and facilities, which I required for the completion of my thesis. I am deeply indebted to my parents and friends for their inspiration and ever encouraging moral support, which enabled me to pursue my studies.

I am also very thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love and affection which made my stay at Thapar University memorable.



Abstract

The expansion of network processing power and higher capacity network links have increased the amount of data and information that can be exchanged over the Internet in networks. As the trend of successful network attacks continue to rise, better forms of intrusion detection and prevention are needed. Organizations are facing the problem to look at a complex data set and understand the various entries in it. Network traffic visualization techniques aid an administrator in recognizing attacks, Instead of reading through the file, line by line, visualize the data with the graph using various tools.

The goal of visualizations is to provide a new way of analyze the data. Visualization is the practice of mapping of data into visual form for exploration analysis and presentation. Visualization is about filtering data, mapping the remaining data to graphics primitives, and rendering the result for repeated playback and study. Log visualization is a very interesting way of representing large amount of data, with this method the interesting data of the log files can be taken in a quick way.

The focus of this thesis is to identify the compromised host on the network through visualizing the log files generated by network security solutions and make ease to administrator to take decision accordingly.

Table of Contents

Certificate *i*

Acknowledgement *ii*

Abstract *iii*

Table of Contents *iv*

List of Figures *vii*

Chapter 1: Introduction **1**

1.1 Need of Network Security 1

1.2 Security Management 2

1.3 Importance of log files in network Security 2

1.4 Log file visualization 3

Chapter 2: Literature Review **5**

2.1 Network Security Goals 5

2.2 Security Life Cycle 6

2.3 Various Threats to Network Security 7

2.4 Types of Network Attacks 10

2.4.1 Reconnaissance Attack 10

2.4.2 Access Attack 11

2.4.3 Denial of Service Attack 11

2.4.4 Smurf Attack 13

2.4.5 Data Interception 13

2.4.6 Social Engineering 14

2.4.7 Vandals 15

2.5 Approaches to Network Security 15
2.5.1 Reactive Approach 15
2.5.2 Proactive Approach 16
2.6 Network Security Solutions 16
2.6.1 Firewall 16
2.6.2 Intrusion Detection System 27
2.6.3 Intrusion Prevention System 18
2.6.4 Honeypot 19
2.7 Network Monitoring and Analysis 20
2.7.1 Active Monitoring 21
2.7.2 Passive Monitoring 23
2.8 Role of Log files in Security 23
2.9 Log Files 24
2.10 Need for Log Management 24
2.11 Log Management 25
2.12 Architecture of Log Management 27
2.13 Limitations of Log Files 28
2.14 Log Visualization 28
2.14.1 Visualization 28
2.14.2 Need of visualization 29
2.14.3 Advantages of using Visualizations 29
2.14.4 Types of Visualization 30
2.14.4.1 Online visualization 30
2.14.4.2 Offline visualization 30
2.15 Log File Visualization 31
2.16 Goals of Log File Visualization 31

2.17 Stages of the visualization process: 32

2.18 Current visualization applications 32

2.18.1 MRTG Tool 33

2.18.2 Afterglow 33

2.18.3 RRD tool 34

Chapter 3: Problem Statement 35

Chapter 4: Proposed Solution and Implementation 36

4.1 Proposed Framework for Log Visualization 36

4.2 Implementation and Results 38

4.2.1 The Virtualization software 38

4.2.2 Fedora 8 under VMware 39

4.2.3 Capturing Network traffic 40

4.2.4 Visualizing logs 42

4.2.5 Identifying Compromised Host 45

4.2.6 Visualizing Log for a particular source IP Address 47

Chapter 5: Conclusions and Future Scope 49

References 51

Papers Communicated 54

List of Figures

Figure No.	Title	Page No.
Figure 2.1	Security Life Cycle.	6
Figure 2.2	Distributed DoS attack.	12
Figure 2.3	TCP SYN attack.	12
Figure 2.4	Smurf attack.	13
Figure 2.5	Firewall.	16
Figure 2.6	Intrusion Detection Systems.	18
Figure 2.7	Honeyd is a daemon application.	20
Figure 2.8	ICMP ping command (Active Measurement).	22
Figure 2.9	Setup of a passive monitoring.	23
Figure 2.10	Online visualization and Offline visualization.	30
Figure 4.1	Framework for Log Visualization.	37
Figure 4.2	Virtualization Software.	39
Figure 4.3	Fedora 8 under VMware.	39
Figure 4.4	Capturing data using tcpdump.	41
Figure 4.5	Parsing of data fields.	42
Figure 4.6	Connections on the server.	43
Figure 4.7	Graph showing the network traffic of Thapar University.	44
Figure 4.8	Compromised machines 172.31.5.10.	45
Figure 4.9	Showing compromised host.	46
Figure 4.10	Closer view of machine 172.31.5.99(compromised)	47
Figure 4.11	Packets captured for extracting a particular source IP address.	48
Figure 4.12	Machine 172.31.5.159 is communicating with 12 IP addresses.	48

Chapter 1

Introduction

A network is one of the most important basic resource in large institutes. Today, networks play a very important role in every organization. With widespread distributed deployment, managing the security of a network becomes very complex. Every system connected to the Internet is a potential candidate for a malicious attack. Even the most secure Internet sites, corporations, and government agencies have experienced security violations and network attacks. To deal with those risks, organizations are deploying network infrastructure devices such as firewalls and intrusion detection systems to increase the security of their networks. Firewalls, intrusion detection systems computer networks against unauthorized access, intrusion attempts, and network attacks, best practices dictate deploying multiple layers of devices and technologies to ensure that networks are secure and can recover rapidly from security breaches. Security policies are used that reduce risk to support the network and computer business requirements [1].

Securing information is vital for the survival of many organizations. Network security involves all activities that organizations, enterprises, and institutions undertake to protect the value and ongoing usability of assets and the integrity and continuity of operations [2]. An effective network security strategy requires identifying threats and then choosing the most effective set of tools to combat them.

1.1 Need of Network Security

Network security is an ongoing process that helps keep unauthorized parties from gaining access to the network. It can also be extremely helpful in detecting whether or not a hacker tried to access a system, what areas were accessed and what damage was done. Network security is vital to keeping hackers from viewing sensitive information whether that sensitive information is about the organization which is not public. Hackers can also use different information found on the network to gain full access to ones system. A simple visible code found on a network can be one of many unlikely keys that can open the gates to control of network. Once the gates are open hackers

not only have access to sensitive stored information, but also can gain control of computer and use it to send out spam, surf the internet and attack other computers and networks. The best way to secure ones network is to perform ongoing security checks on a network and have a security management for it.

1.2 Security Management

Security management plays a larger role as all communication is over the insecure internet security management. The increasing number of people, organizations, and enterprises which subscribe to the internet makes security management an important issue. Security management aims to maintain the integrity, confidentiality and availability of systems and services. It involves protecting a network from all kinds of unauthorized modification, destruction, or disclosure. This includes many sub functions like collecting and reporting security related information, proactively detecting and preventing intrusions. It provides assurance that the networks perform its critical functions correctly and there are no harmful side effects Security management tells the challenges of securing information in a global, dynamic networked systems environment [3]. It gives range of vulnerabilities and threats to which an organization's network assets may be exposed and provides strategies to deal with information security problems and ways to improve the network security of an organization.

1.3 Importance of log files in network Security

Examining log files is one of the most useful ways in detecting and investigating problems with computer systems. Logs can provide information about systems faults and misuse as well as early warning of problems All network systems and devices like Windows/Linux desktops & servers, routers, switches, firewalls, proxy server, VPN, IDS and other network resources generate logs by every second. The logs contain all information of all the system, device, and user activities that take place within these network infrastructures. Log files are important forensic tools for investigating an organizations security posture. Analysis of these log files provide plethora of information on user level activities like logon success or failure, objects access , website visits, system & device level activities like file read, write or delete, host

session status, account management, network bandwidth consumed, protocol & traffic distribution and network security activities like identifying virus or attack signatures and network anomalies [4].Log files are created and continuously updated by the system to provide a tracking mechanism of all the activities that transpired during the system uptime.

Today's system administrators, under the burden of rapidly increasing network activity, need the ability to rapidly understand what is happening on their networks. While text-based systems are able to assist with awareness by alerting the user to potential problems and not good at helping the user forms an accurate mental model of the situation. Since text-based systems cannot provide a dynamic overview of the whole network, network administrators need visualization tools [5].

The examination of logs plays a big role in modern computer security, but it has become a time consuming and daunting task due to the sheer amount of data involved. It is therefore necessary to make specialized tools to aid the investigation, so that the digital evidence can be extracted in a fast and efficient manner.

1.4 Log file visualization

Network log analysis can be of great assistance in such diverse areas as intrusion detection, incident handling and problem troubleshooting. As log files contains so much information which is very redundant in nature. Administrator has a great problem while reading the logs which can be resolved by visualizing. The log files visualization provides the pictorial view of the events which are easy to understand and identify the suspected event.

In many organizations, system administrators ignore network log files until a major incident occurs. The size of log file is too large that administrator does not feel comfortable to store and read them. It contains every event which is not desired and it is very difficult to find out the required fields from the log files. Visualizing the log files make it easy to find out the required event happened on the network. Pictorial representation takes a short time to take a view at a glance of records and any malicious activity can be identified easily.

Visual Information Utility for Administration helps network administrators by showing the overview of their network traffic or logs in graphical representation. Aside from seeing abnormal traffic, Visualization techniques offer a first step in qualitative understanding of the traffic trends and the relationships that can exist between various traffic parameters. A basic problem in relation to the data network is the degree of traffic predictability [6]. Visualization tools used in conjunction with the ability of the human visual system.

In the process of log visualization which takes raw data as input and makes different images of log files represents every information in pictorial forms. The framework for log visualization consists various steps in which network traffic is captured and then analyzing the data includes only specific fields which can be accomplished by filtering the log files and then this is parsed as input of a tool and generates beautiful graphs which help the administrator to examine malicious activities and compromised hosts on the network. So in this way an administrator can easily find compromised hosts and can take decision effectively.

Chapter 2

Literature Review

Network security refers to all hardware and software functions, characteristics, features, operational procedures, accountability measures, access controls, and administrative and management policy required to provide an acceptable level of protection for hardware , software, and information in a network. Network security operators are overloaded with textual logs and interfaces using primitive graphs generated from these security appliances that prevent them from accurately determining significant problems. An effective network security strategy requires identifying threats and then choosing the most effective set of tools to combat them [7]. Network security consists of the provisions made in an underlying computer network infrastructure, policies adopted by the network administrator to protect the network and the network-accessible resources from unauthorized access, and consistent and continuous monitoring and measurement of its effectiveness.

2.1 Network Security Goals

A secure network must have integrity such that all of the information stored there in is always correct and protected against fortuitous data corruption as well as wilful alterations. To secure a network there must be confidentiality, or the ability to share information on the network with only those people for whom the viewing is intended. Finally, network security requires availability of information to its necessary, recipients at the predetermined times without exception [8].

A successful security management solution begins as an integral part of an organization's overall business strategy. Once security management is accepted as a core business operation, it necessitates the development of guidelines that create the security practices necessary to support the business strategy. Security is a not a technology problem, it's a people problem. There is no computer security product that acts as magical security dust, imbuing a network with the property of "secure." It's not the way business works [9].

2.2 Security Life Cycle

A process for maintaining an acceptable level of perceived risk is security. It is not an end state, but rather a process. There are four steps in a security lifecycle: assessment, Protection, detection, and response.

A. Assessment

Assessment pertains to policies, procedures, laws, regulations, budgeting, and other functions that are preparation for the other three dimensions of the process.

B. Protection

Although protection is the desired end result, in this context protection refers to the process of implementing countermeasures to reduce the likelihood of a compromised network.

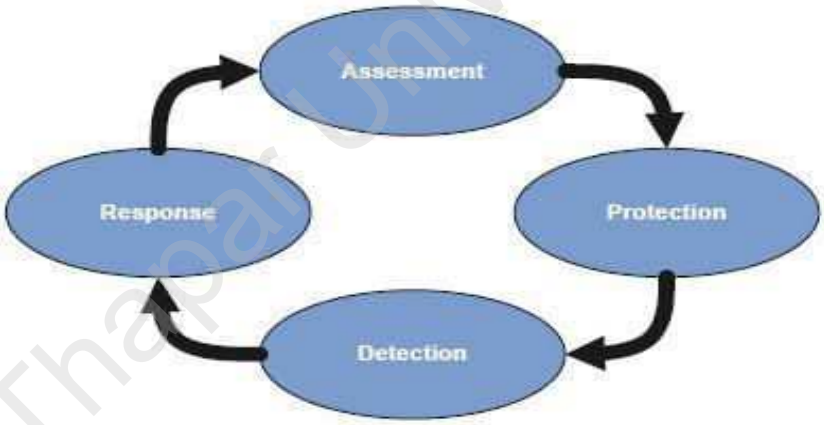


Figure 2.1: Security Life Cycle.

C. Detection

Detection is the process of identifying policy violators who are unauthorized to use or perform unacceptable actions on a computer network.

D. Response

Response is acting on the information gathered during the detection phase. The activities that occur during the response phase include maintenance on the

countermeasures implemented during the protection phases and prosecution of intruders [10].

2.3 Various Threats to Network Security

Viruses

A virus is a piece of programming code inserted into other programming to cause some unexpected and for the victim, usually undesirable event. Computer programs written by devious programmers and designed to replicate themselves and infect computers when triggered by a specific event Viruses can be transmitted by down loading code from other sites, from email or internet, or by using an infected diskette. The source of the file downloaded, or of a diskette received, is often unaware of the virus [11].

Some common symptoms that could indicate that system has been infected are:

- Computer slows down without reason.
- Unusual messages or displays on the monitor.
- Unusual sounds or music played at random times.
- System has less available memory than it should.
- A disk or volume name has been changed.
- Programs or files are suddenly missing.
- Unknown programs or files have been created.
- Some of files become corrupted or suddenly don't work properly.

Main categories of viruses:

- Micro Viruses
- Macro Viruses

Micro Viruses

Boot Sector Viruses

Boot viruses place their code in the disk sector, whose code the machine will automatically execute when booting. Thus, when an infected machine boots, the virus loads and runs. After boot viruses are finished loading, the original boot code is loaded, which moved to another location or take other measures to ensure the machine appears to boot normally.

File Infector Viruses

File viruses attach to program files containing executable or interpretable code in such a way that when someone runs the infected program, the virus code executes. The virus code is added in such a way that it executes first. After the virus code has finished loading and executing, it will normally load and execute the original program it has infected, or call the function it intercepted, so as to not arouse the user's suspicion.

Macro Viruses

Macro viruses are the most prevalent and successful type of virus. Macro viruses attach themselves to documents such as word processing files and spreadsheets. A macro virus uses an application's macro programming language to execute and propagate. Many popular software packages, such as Microsoft Office, use macro programming languages in their products to automate complex or repetitive tasks. Attackers have taken advantage of macro programming capabilities to distribute malicious code. Macro viruses tend to spread quickly because users frequently share documents from applications with macro capabilities. The Concept, Marker, and Melissa viruses are well-known examples of macro viruses. The example of macro virus is Trojan horse which is a delivery vehicles for destructive code, which appear to be harmless or useful software program.

Script viruses

Script viruses also became quite successful around the beginning of this century. This was mainly due to the increase in machines running Windows Scripting Host, which was first installed by default in Windows 98 and 2000 and with Internet Explorer 5.0 and later versions. Representing new types of 'program file', but with icons more like that of 'safe' text files, standalone Visual Basic Script and JavaScript programs became a popular target of the writers of mass mailing viruses.

Mobile Code

Mobile code is software that is transmitted from a remote system to a local system and then executed on the local system without the user's explicit instruction. Popular vehicles for mobile code include Java applets, ActiveX, JavaScript, and VBScript.

Worms

Worms are similar to viruses in that worms make copies of themselves, but different in that worms need not attach to particular files or sectors at all. Once such a worm is executed, it seeks other systems rather than parts of systems to infect, then copies its code to them in such a way as to have the code execute directly from memory. The primary difference is that worms do not require access to programs to replicate themselves. Worms take advantage of shared directories, interconnected machines and vulnerable network services [12].

Trojan horse programs

Just as the mythological Trojan horse appeared to be a gift, but turned out to contain Greek soldiers who overtook the city of Troy, today's Trojan horses are computer programs that appear to be useful software, but instead security is compromised and causes a lot of damage. A recent Trojan horse came in the form of an e-mail that included attachments claiming to be Microsoft security updates, but turned out to be viruses that attempted to disable antivirus and firewall software [13].Trojan horses spread when people are lured into opening a program because it comes from a legitimate source.

2.4 Types of Network Attacks

Including reconnaissance attacks in which information gathering activities is performed to collect data that is later used to compromise networks, access attacks which exploit network vulnerabilities in order to gain entry to e-mail, databases, or the corporate network, and denial-of-service attacks which prevent access to part or all of a computer system.

2.4.1 Reconnaissance Attack

Reconnaissance is the unauthorized discovery and mapping of systems, services, or vulnerabilities. It is also called information gathering. In most cases, it precedes an actual access or DoS attack. The malicious intruder typically ping-sweeps the target network first to determine what IP addresses are alive. After this is accomplished, the intruder determines what services or ports are active on the live IP addresses. From this information, the intruder queries the ports to determine the application type and version as well as the type and version of the operating system running on the target host [14].The hacker surveys a network and collects data for a future attack. Important information that can be compiled during a reconnaissance attack includes the following:

- Ports open on a server
- Ports open on a firewall
- IP addresses on the host network
- Hostnames associated with the IP addresses

As with access attacks, there are four main subcategories or methods for gathering network data:

- Packet sniffers (also known as network monitors)
- Ping sweeps
- Port scans
- Information queries

2.4.2 Access Attack

Access is an all-encompassing term that refers to unauthorized data manipulation, system access or privilege escalation. Unauthorized data retrieval is simply reading, writing, copying, or moving files that are not intended to be accessible to the intruder. Sometimes this is as easy as finding shared folders in Windows 9X or NT, or NFS exported directories in UNIX systems with read or read-write access to everyone. The intruder has no problem getting to the files. More often than not, the easily accessible information is highly confidential and completely unprotected from prying eyes, especially if the attacker is already an internal user. System access is an intruder's ability to gain access to a machine that he is not allowed access. Entering or accessing systems that don't have access to usually involve running a hack, script, or tool that exploits a known vulnerability of the system or application being attacked.

Another form of access attacks involves privilege escalation. This is done by legitimate users who have a lower level of access privileges or intruders who have gained lower privileged access [15]. The intent is to get information or execute procedures that are authorized at the user's current level of access. In many cases this involves gaining root access in a UNIX system to install a sniffer to record network traffic, such as usernames and passwords that can be used to access another target.

2.4.3 Denial of Service Attack

DoS (Denial-of-Service) attacks are probably the nastiest, and most difficult to address. These are the nastiest, because they're very easy to launch, difficult to track, and it isn't easy to refuse the requests of the attacker, without also refusing legitimate requests for service. DoS attack technology involves tools that generates and send packets from a single source aimed at a single destination [16].

Distributed DoS (DDoS)

In distributed DoS, multiple systems are compromised to send a DoS attack to a specific target. The compromised systems are commonly called zombies or slaves.

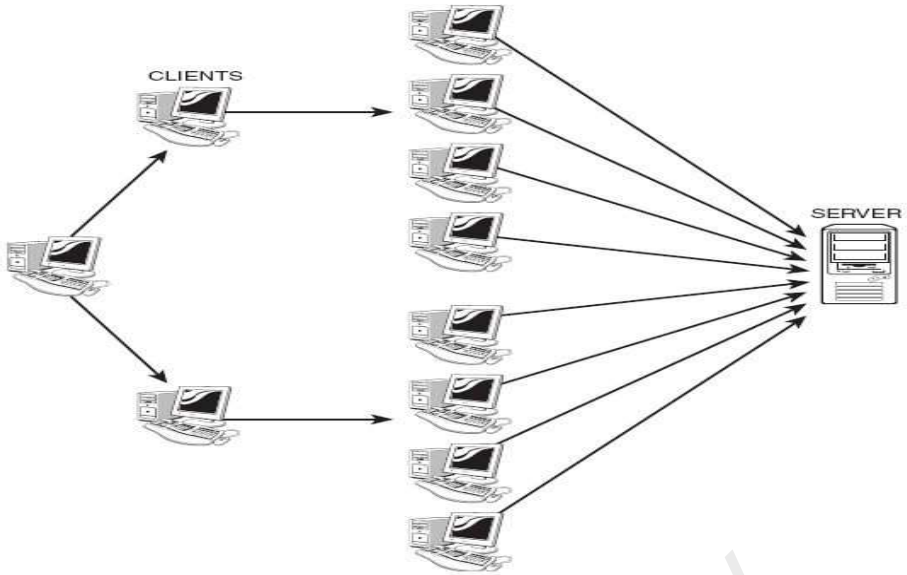


Figure 2.2: Distributed DoS attack.

TCP SYN Attack

A TCP session is established with the use of a three-way handshake in which a connection agreement segment is sent to the recipient, asking to synchronize systems. This step is associated with the bit name SYN.

The second and third segments acknowledge the request to connect and determine the rules of engagement. Sequencing synchronization is requested of the receiving device. A two-way connection is established. This step is associated with the bit name SYN-ACK.A final segment is sent as an acknowledgment that the rules have been accepted and a connection has been formed. This step is associated with the bit name ACK [17].

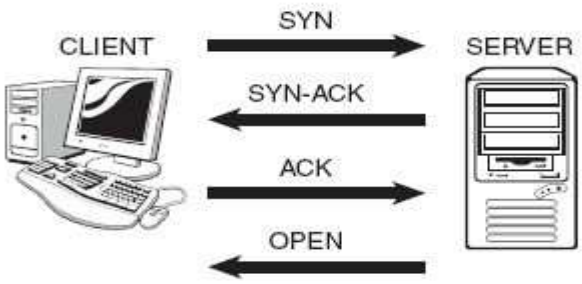


Figure 2.3: TCP SYN attack

2.4.4 Smurf Attack

Smurf technique is based on the use of broadcast servers to paralyze a network. A broadcast server is a server capable of duplicating a message and sending it to all machines present on the same network. The attacking machine sends a ping request to one or more broadcast servers while falsifying the source IP address and providing the IP address of a target machine, then the broadcast server passes on the request to the entire network after that all of the network's machines send a response to the broadcast server and the broadcast server redirects the responses to the target machine. As such, when the attacking machine sends a request to several broadcast servers located on different networks, all of the responses from computers on the various networks will be routed to the target machine. Smurf attack, multiple broadcast ping requests are sent to a single target from a spoofed IP address. Adding the no IP directed-broadcast command to a router might help mitigate a potential smurf attack.

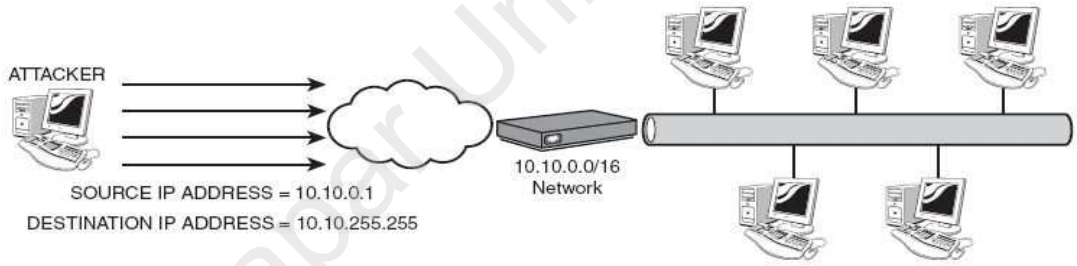


Figure 2.4: Smurf Attack.

2.4.5 Data Interception

Data transmitted via any type of network can be subject to interception by unauthorized parties. The intercepting perpetrators might eavesdrop on communications or even alter the data packets being transmitted. Perpetrators can use various methods to intercept data. IP spoofing, for example, entails posing as an authorized party in the data transmission by using the Internet Protocol (IP) address of

one of the data recipients. One of the most widespread methods of Data interception is "Man in the Middle Attack (MITM)".

In cryptography, a man in the middle attack (MITM*)* is an attack in which an attacker is able to read, insert and modify at will, messages between two parties without either party knowing that the link between them has been compromised. The attacker must be able to observe and intercept messages going between the two victims. Suppose 'A' wishes to communicate with 'B', and that 'C' wishes to eavesdrop on the conversation, or possibly deliver a false message to 'B'. To get started, 'A' must ask 'B' for his public key. If 'B' sends his public key to 'A', but 'C' is able to intercept it, a man in the middle attack can begin. 'C' can simply send 'A' a public key for which he has the private, matching, key. 'A', believing this public key to be 'B''s, then encrypts her message with 'C''s key and sends the enciphered message back to 'B'. 'C' again intercepts, deciphers the message, keeps a copy, and re enciphers it (after alteration if desired) using the public key 'B' originally sent to 'A'. When 'B' receives the newly enciphered message, he will believe it came from 'A'. A similar attack is possible, in principle, against any message sent using public key technology, including data packets carried on computer networks.

2.4.6 Social Engineering

Social Engineering is an attack method used by many attackers that takes advantage of trust and complacency at work. Humans by nature are very trusting and rarely question actions that are considered normal. Another forum that social engineering can expose is the Computer Conference. Computer conferences are great for obtaining information. Social engineering is referred to as an approach to gain access to information, primarily through misrepresentation, and often relies on the trusting nature of most individuals [18]. Most conferences stress openness, this within itself is not a bad idea but the problem occurs when people give too many details. Some of the information that attendee's and instructors give out could be used against them and their networks. Information about network configuration, types of firewalls and Intrusion Detection systems were just a few items commonly shared.

Dumpster diving, also known as trashing is another popular method of social engineering. A huge amount of information can be collected through company dumpsters. company phone books, organizational charts, memos, company policy manuals, calendars of meetings, events and vacations, system manuals, printouts of sensitive data or login names and passwords, printouts of source code, disks and tapes, company letterhead and memo forms, and outdated hardware are potential security leaks. Other major types of social engineering attacks used are social engineering by phone and persuasion etc.

2.4.7 Vandals

Web sites have come alive through the development of such software applications as ActiveX and Java Applets. These applications enable animation and other special effects to run, making web sites more attractive and interactive. However, the ease with which these applications can be downloaded and run has provided a new vehicle for inflicting damage. Vandals can take on the form of a software application or applet that causes destruction of various degrees. A vandal can destroy a single file or a major portion of a computer system.

2.5 Approaches to Network Security

2.5.1 Reactive Approach

Reactive approaches are those procedures that organizations use when some of systems have been compromised by an intruder or attack program. Reactive systems, such as intrusion detection or intrusion prevention systems depend on an attack, incident, or loss of some degree to occur before the start the information gathering and analysis that ultimately drives some form of automation or reporting. Reactive methods include Disaster Recovery Plans, use of private investigation services and loss recovery specialists, reinstallation of operating systems and applications on compromised systems, or switching to alternate systems in other locations. Having an appropriate set of reactive responses prepared and ready to implement is just as important as having proactive measures in place.

2.5.2 Proactive Approach

A proactive system constantly tests the organization's network for vulnerabilities and exposures. It then assesses and prioritizes those vulnerabilities and exposures and manages the process by which those vulnerabilities and exposures are addressed. All IP devices attached to the network are periodically or continuously scanned and profiled for changes, violations to policy, and vulnerabilities and exposures. Proactive network security is the act of managing the network security to get the most performance system with a vulnerability management system [19].

In contrast to reactive systems, proactive systems have the advantage of providing valuable intelligence about an organization's network and networked devices. Proactive systems work best when complemented with appropriate reactive systems.

2.6 Network Security Solutions

2.6.1 Firewall

A firewall is a system designed to prevent unauthorized access to or from a private network and restrict access to the services running on a firewall-enabled server. Firewalls use policies or set of rules that govern the flow of data packets to and from the outside world. This helps to prevent unauthorized Internet users from accessing private networks connected to the Internet [20].

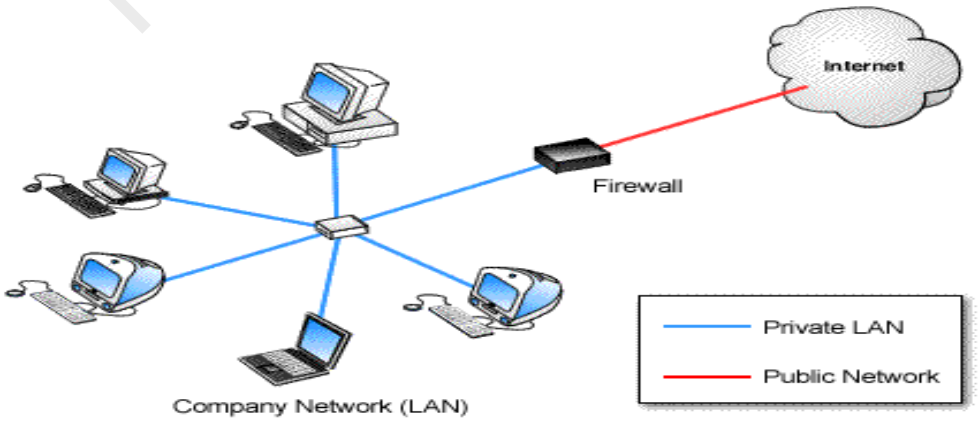


Figure 2.5: Firewall.

2.6.2 Intrusion Detection System

In Information Security, intrusion detection is the act of detecting actions that attempt to compromise the confidentiality, integrity or availability of a resource. A system that performs automated intrusion detection is called an Intrusion Detection System (IDS).An intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system.

Intrusion detection functions include:

- Monitoring and analyzing both user and system activities.
- Analyzing system configurations and vulnerabilities.
- Assessing system and file integrity.
- Ability to recognize patterns typical of attacks.
- Analysis of abnormal activity patterns.
- Tracking user policy violations.

There are two complementary trends in intrusion detection:

(1) Knowledge-based, to use the knowledge accumulated about attacks and look for evidence of the exploitation of these attacks.

(2) Behaviour-based, to build a reference model of the usual behaviour of the information system being monitored and look for deviations from the normal usage.

The behaviour on detection describes the response of the intrusion-detection system to attacks, reactive or passive. When it actively reacts to the attack by taking either or pro-active actions, then the intrusion-detection system is said to be active. If the intrusion-detection system merely generates alarms it is said to be passive [21].

IDS can be host-based, if it monitors system calls or logs, or network-based if it monitors the flow of network packets. There are also protocol-based IDS, application protocol-based IDS and hybrid one which combines two or more approaches.

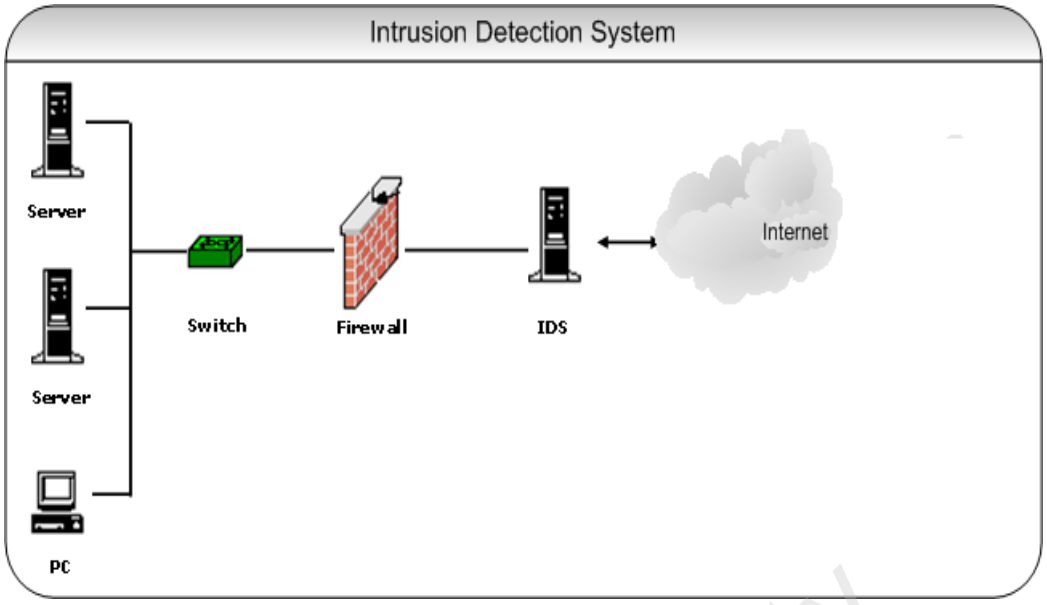


Figure 2.6: Intrusion Detection Systems

2.6.3 Intrusion Prevention System

IPS devices have gained popularity as security professionals increasingly focus on stopping potential intrusions before become a threat. An IPS can be thought of as a highly refined firewall, able to deny hostile traffic while allowing legitimate traffic to pass through. In contrast to IDS devices, IPS devices can actively intervene when potential attack or intrusion is detected. Intervention may result in flagging of traffic for further manual assessment, session termination, or some other action.

IPS mechanisms are designed to analyze traffic in the context of other network activities and because the rely on a wide range of intelligence and data, leveraged across multiple detection methodologies and detect intrusions more accurately than IDS devices. This capability reduces the number of false alarms and allows system administrators to focus on true threats [22].

Today's IPS devices are typically deployed in the same locations as firewalls and/or IDS devices. More broadly speaking, these locations can be thought of as the two edges of the network the outer perimeter and down to the desktop level. Neither of these solutions is perfect and many experts believe the technology won't reach full maturity until it is pushed into the actual network-switching fabric.

2.6.4 Honeypot

A honeypot is used in the area of computer and Internet security. It is a resource which is intended to be attacked and compromised to gain more information about the attacker and the tools used by attackers. It can also be deployed to attract and divert an attacker from their real targets.

L. Spitzner defines the term honeypot as follows:

"A honeypot is a resource whose value is being in attacked or compromised. This means, that a honeypot is expected to get probed, attacked and potentially exploited" [23]. Honeypots do not fix anything but provide additional valuable information. Honeypots do not help directly in increasing a computer network's security. On the contrary, they do attract intruders and can therefore attract some interest from the Blackhat community on the network where the honeypot is located. So the above definition can be modified as

"A honeypot is a resource which pretends to be a real target. A honeypot is expected to be attacked or compromised. The main goals are the distraction of an attacker and the gain of information about an attack and the attacker" There is a general definition which covers all the different manifestations of honeypots. A honeypot is an information system resource whose value lies in unauthorized or illicit use of that resource [24].

Honeyd

Honeyd is low interaction, freely available, open source pre-packaged virtual honeypot solution. The software was developed by Niels Provos. Since it is an Open source, the program is constantly developing and evolving with new features and functionalities from contributors from all around. The low interaction classification of honeyd will only allow emulating the services and doesn't allow attacker to interact with the operating system of the honeypot. Similar to KFSensor the services can be ran into any TCP port.

As shown in Figure 2.7 Honeyd is a daemon application which enables the setup of multiple virtual honeypots on a single machine. The main important difference with the KFSensor is that, personality feature. This feature or configuration will allow configuring the each production honeypot with a personality of OS IP stack and it binds a script to the emulated port to visualize the service.

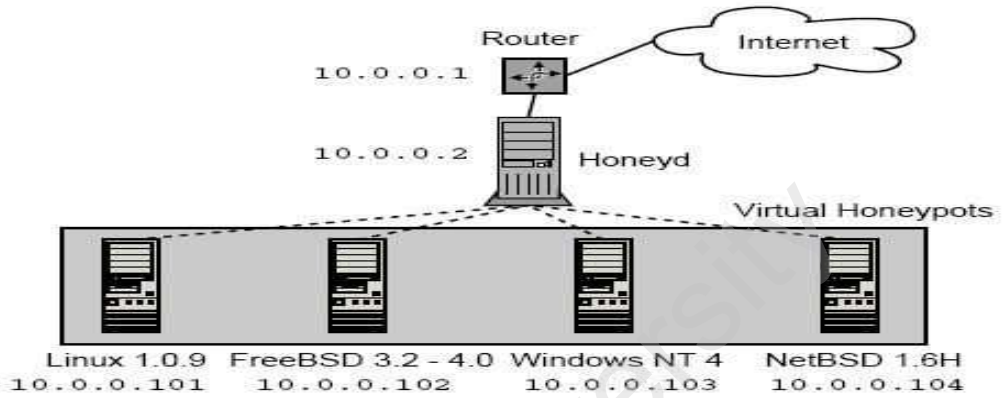


Figure 2.7: Honeyd is a daemon application.

Honeyd is primarily used for detecting attacks. It works by monitoring IP addresses that are unused, that have no system assigned to them. Whenever an attacker attempts to probe or attack a non-existant system, Honeyd assumes the IP address of the victim and then interacts with the attacker through emulated services. Honeypots are limited to detecting attacks only on the ports that have emulated services listening on. It detects and logs connections made to any port, regardless if there is a service listening [25].

2.7 Network Monitoring and Analysis

Network monitoring implies a series of sensors in and around the network. Every firewall produces a continuous stream of messages. Every other security product generates alarms in some way. It is a difficult and demanding task that is a vital part of a network administrator's job. Network administrators are constantly striving to maintain smooth operation of their networks. If a network were to be down even for a small period of time productivity within a company would decline, and in the case of

public service departments the ability to provide essential services would be compromised.

In order to be proactive rather than reactive, administrators need to monitor traffic movement and performance throughout the network and verify that security breaches do not occur within the network. The first step is intelligent alert. Network attacks can be subtle, and much depends on context. Software can filter the tens of megabytes of audit information a medium-sized network can generate in a day, but software is too easy for an attacker to fool. Intelligent alert requires people to:

• Analyze what the software finds suspicious.

• Delve deeper into suspicious events, determining what is really going on.

• Separate false alarms from real attacks.

• Understand context.

By itself, an alert is only marginally useful. More important is to know how to respond. This is the second step of good network monitoring. Software can only provide generic information; real understanding requires experts. Finally, the response must be integrated with organizational business needs. Efficient analysis tools are available, it could become possible.

- To detect the attacks and anomalies.
- To appropriately take action to mitigate the attacks before they have had time to propagate across the network or to cripple the infrastructure.

2.7.1 Active Monitoring

Active monitoring transmits probes into the network to collect measurements between at least two endpoints in the network. Active measurement systems deal with metrics such as:

- Availability.
- Routes.
- Packet Delay.

- Packet Reordering.
- Packet Loss.
- Packet Inter-arrival Jitter.
- Bandwidth Measurements.

Most commonly used tools such as ping, which measures delay and loss of packets, and traceroute which helps determine topology of the network, are examples of basic active measurement tools. They both send ICMP packets to a designated host and wait for the host to respond back to the sender [26]. In Figure 2.8 is an example of the ping command that uses active measurements by sending an Echo Request from the source host through the network to a specified destination. The destination then sends an Echo Response back to the source it received the request from.

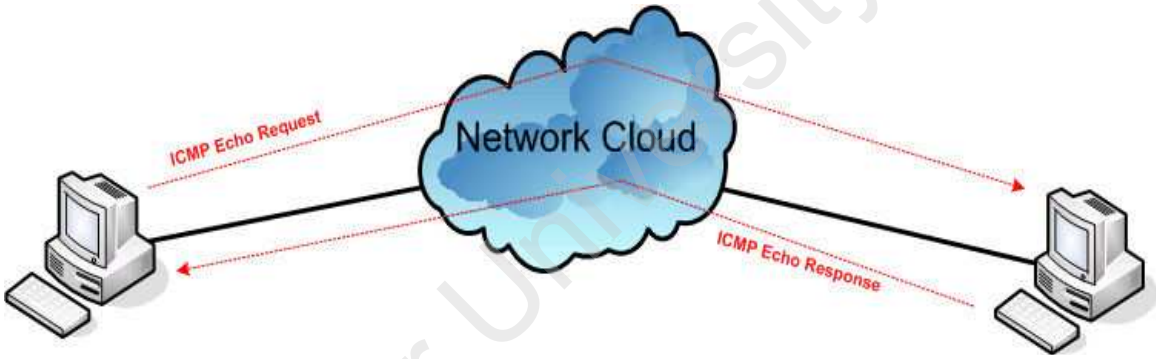


Figure 2.8: ICMP ping command (Active Measurement)

A person can collect the metrics above from active measurements and also determine the network topology.

The problem that exists with active monitoring is that introducing probes into the network can be interference to the normal traffic on the network. Often times the active probes are treated differently than normal traffic as well, which causes the validity of the information provided from these probes to be questioned. As a result of the information detailed above, active monitoring is very rarely implemented as a stand-alone method of monitoring as a good deal of overhead is introduced. On the other hand passive monitoring does not introduce much if any overhead into the network [27].

2.7.2 Passive Monitoring

Passive monitoring unlike active monitoring does not inject traffic into the network or modify the traffic that is already on the network. Also unlike active monitoring, passive monitoring collects information about only one point in the network that is being measured rather than between two endpoints as active monitoring measures.

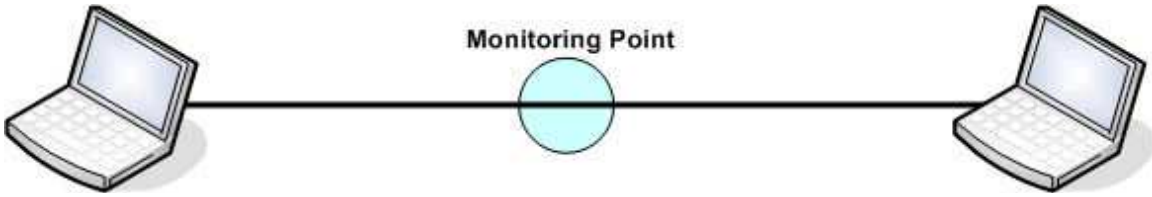


Figure 2.9: Setup of a passive monitoring.

Passive monitoring system where the monitor is placed on a single link between two endpoints and monitors traffic as it passes along the link as shown in Figure 2.9. In passive monitoring setup passive measurements deal with information such as traffic and protocol Although passive monitoring does not have the overhead that active monitoring has, it has its own set of downfalls.

With passive monitoring only off-line measurements can be analyzed. This creates another problem with processing the huge data sets that are collected. As one can see passive monitoring may be better than active monitoring in that overhead data is not added into the network but post-processing time can take a large amount of time [28].

2.8 Role of Log files in Security

Logs are generally more likely than others to record information that would be helpful for different situations, such as detecting attacks, fraud, and inappropriate usage. For each type of situation, certain logs are typically the most likely to contain detailed information on the activity. An intrusion detection system could record malicious commands issued to a server from an external host which is primary source of attack information. An incident handler could then review a firewall log to look for other connection attempts from the same source IP address which would be a secondary source of attack information

Administrators using logs should also be mindful of the trustworthiness of each log source. Log sources that are not properly secured, including insecure transport mechanisms, are more susceptible to log configuration changes and log alteration. Originally, logs were used primarily for troubleshooting problems, but logs now serve many functions within most organizations, such as optimizing system and network performance, recording the actions of users and providing data useful for investigating malicious activity. Logs have evolved to contain information related to many different types of events occurring within networks and systems. Within an organization, many logs contain records related to computer security.

2.9 Log Files

A log is a record of the events occurring within an organization's systems and networks. Logs are composed of log entries which contains information related to a specific event that has occurred within a system or network.

Log File is a file that lists actions that have occurred. A log file is simply a file that is used to keep track of all that is happening with sites. With log file analysis, good idea of where visitors are coming from and how often they return can be taken. It records the every time, a web page is requested, what incoming link a visitor followed to arrive at the site.

2.10 Need for Log Management

Log management can benefit an organization in many ways. It helps to ensure that computer security records are stored in sufficient detail for an appropriate period of time. Routine log reviews and analysis are beneficial for identifying security incidents, policy violations, fraudulent activity, and operational problems shortly after they have occurred, and for providing information useful for resolving such problems. Logs can also be useful for performing auditing and forensic analysis, supporting the organization's internal investigations, establishing baselines, and identifying operational trends and long-term problems [29].

Besides the inherent benefits of log management, a number of laws and regulations further compel organizations to store and review certain logs. The following is a listing of key regulations, standards, and guidelines that help define organizations' needs for log management:

2.11 Log Management

Log management is essential to ensure that computer security records are stored in sufficient detail in order to perform audits and forensic analysis. The benefits of routine log analysis, which are detailed in, are as follows:

- Performing auditing and forensic analysis.
- Supporting internal investigations.
- Establishing baselines, and identifying operational trends, security incidents, policy violations and fraudulent activity.

Functions

Log management infrastructures perform the functions that assist in the storage, analysis, and disposal of log data. These functions do not alter the original logs. The following items describe common log management infrastructure functions:

1. **Log Parsing**

   Log parsing is extracting data from a log so that the parsed values can be used as input for another logging process. A simple example of parsing is reading a text-based log file that contains 10 comma-separated values per line and extracting the 10 values from each line. Parsing is performed as part of many other logging functions, such as log conversion and log viewing.

2. **Event Filtering**

   Event filtering is the suppression of log entries from analysis, reporting, or long-term storage because their characteristics indicate that they are unlikely to contain information of interest. For example, duplicate entries and standard

informational entries might be filtered because they do not provide useful information to log analysts.

3. **Event Aggregation**

   In log files similar entries are consolidated into a single entry containing a count of the number of occurrences of the event. For example, a thousand entries that each record part of a scan could be aggregated into a single entry that indicates how many hosts were scanned.

4. **Log rotation**

   Log rotation is closing a log file and opening a new log file when the first file is considered to be complete. Log rotation is typically performed according to a schedule (e.g., hourly, daily, weekly) or when a log file reaches a certain size. The primary benefits of log rotation are preserving log entries and keeping the size of log files manageable. When a log file is rotated, the preserved log file can be compressed to save space.

5. **Log archival**

   Log archival is retaining logs for an extended period of time, typically on removable media, a storage area network (SAN), or a specialized log archival appliance or server. Logs often need to be preserved to meet legal or regulatory requirements. There are two types of log archival:

   - Retention: Log retention is archiving logs on a regular basis as part of standard operational activities.
   - Preservation: Log preservation is keeping logs that normally would be discarded, because they contain records of activity of particular interest.

6. **Log compression**

   Log compression is storing a log file in a way that reduces the amount of storage space needed for the file without altering the meaning of its contents. Log compression is often performed when logs are rotated or archived.

7. Log Reduction

Log reduction is removing unneeded entries from a log to create a new log that is smaller. A similar process is event reduction, which removes unneeded data fields from all log entries. Log and event reduction are often performed in conjunction with log archival so that only the log entries and data fields of interest are placed into long-term storage.

2.12 Architecture of Log Management

A log management infrastructure typically comprises the following three tiers:

Log Generation

The first tier contains the hosts that generate the log data. Some hosts run logging client applications or services that make their log data available through networks to log servers in the second tier. Other hosts make their logs available through other means, such as allowing the servers to authenticate to them and retrieve copies of the log files.

Log Analysis and Storage

The second tier is composed of one or more log servers that receive log data or copies of log data from the hosts in the first tier. The data is transferred to the servers either in a real-time or near-real-time manner or in occasional batches based on a schedule or the amount of log data waiting to be transferred. Servers that receive log data from multiple log generators are sometimes called collectors or aggregators. Log data may be stored on the log servers themselves or on separate database servers.

Log Monitoring

The third tier contains consoles that may be used to monitor and review log data and the results of automated analysis. Log monitoring consoles can also be used to generate reports. In some log management infrastructures, consoles can also be used to provide management for the log servers and clients. Also, console user privileges sometimes can be limited to only the necessary functions and data sources for each user [30].

The second tier log analysis and storage can vary greatly in complexity and structure. The simplest arrangement is a single log server that handles all log analysis and storage functions.

2.13 Limitations of Log Files

Log files are created and continuously updated by the system to provide a tracking mechanism of all the activities that transpired during the system uptime. These log file systems are constrained or limited by a non-existent log filtering mechanism.Critical events are not properly monitored and lack of a central log monitoring and reporting system make unavailability of a robust audit system.

A weak strategy of storage and preservation of log data and a non-existent log integrity preservation mechanism leads to the lack of an automated log analysis system.

Each system or application sends its log messages to a central server, which collects, normalizes, and archives them. The central system then searches for suspicious log messages or combinations of log messages and generates events.

2.14 Log Visualization

2.14.1 Visualization

"A picture is worth a thousand packets". In researchers opinion it would probably be more like "a picture is worth million packets". Visualization is the practice of mapping of data into visual form for exploration analysis and presentation. Visualization is about filtering data, mapping the remaining data to graphics primitives, and rendering the result for repeated playback and study Good visualizations reveal the hidden structure of the networks and amplify human understanding, thus leading to new insights, new findings and possible predictions for the future. Visualization can summarize many numbers neatly in one picture, highlight unusual parts of the data, and sometimes show meaningful patterns in the data that would have been difficult or impossible to find without visualization [31].

2.14.2 Need of visualization

When a network failure occurs, monitoring agents have to detect, isolate, and correct malfunctions in the network and possibly recover the failure. With the stable network, the administrators' jobs remain to monitor constantly if there is a threat from the network Task for administrator. The network flow information is very useful not only to understand network behaviour, to detect security holes, but also to make good decisions on network planning network traffic characteristics can detect security vulnerabilities.

Today's system administrators, under the burden of rapidly increasing network activity, need the ability to rapidly understand what is happening on their networks. While text-based systems are able to assist with awareness by alerting the user to potential problems. They are not good at helping the user form an accurate mental model of the situation. Because text-based systems cannot provide a dynamic overview of the whole network, network administrators need visualization tools [32].

2.14.3 Advantages of using Visualizations

- Visualizations can usually show much more information in the same viewing space.
- Allows understanding a greater amount of data in shorter time.
- Visualizations are possible to be processed very efficiently by human beings by making use of human perception capabilities Humans can process pictures, a parallel process, faster than text.
- Visualization helps the administrators gain comprehensive understanding of their networks and hence hopefully make more informative decisions and more timely responses during the network security management.
- Images are also easier to remember because humans think and learn visually.
- Helps analysts by scaling and visualizing data and facilitating the identification of patterns in the network in order to make decisions correctly [33].

2.14.4 Types of Visualization

Network event visualization can be categorized into offline and online visualization.

2.14.4.1 Online visualization

Online visualization means that the events are displayed while the simulation is running. Online visualization allows seeing the ongoing network behaviour right away. In case some undesired state occurs, the simulation can be terminated by the user directly. This can save precious time, since simulations are often complex and lengthy. With online visualization, one would have to redo the whole simulation again and again. Even when the simulation tool is capable of reproducing exactly the same event sequence which takes lots of time.

2.14.4.2 Offline visualization

With offline visualization, the events are recorded in a file and displayed only after the simulation is finished. Offline visualization allows some previous point of time to rewind. That way, specific events which caught the designer's eye but were too short to be grasped fully can be conveniently replayed as often as desired [34]. There are many flexible open-source visualization tools for producing the static visualizations as well such as Graphviz.

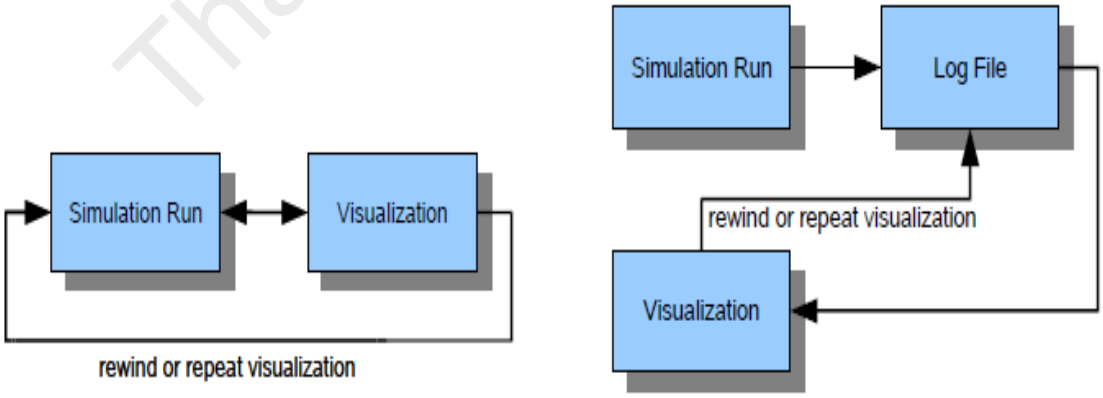


Figure 2.10: Online visualization and Offline visualization.

2.15 Log File Visualization

Log visualization is a very interesting way of representing large amount of data with this method some interesting data of the log files id obtained in a quick way. Log visualization is the key area in network security to analyze the log files. Logs plays a big role in modern computer security, but it has become a time consuming and daunting task due to the sheer amount of data involved. It is therefore necessary to make specialized tools to aid the investigation, so that the digital evidence can be extracted in a fast and efficient manner. Network security visualization aids the network administrator in the first two generalized tasks, monitoring and analyzing. The more specific tasks that a network security administrator must perform in the monitoring and analysis stages include detecting insecurities, detecting intrusion attempts, defending against network attacks, and detecting resource misuse. Network traffic visualization techniques aid an administrator in recognizing attacks and helps to take decision effectively.

2.16 Goals of Log File Visualization

System generates log files. These files contain all important information of all actions performed on the system like source addresses, destination addresses, and port number. So lots of data collected on server in huge amount. All information is text based. This takes lots of time and also cannot provide a dynamic overview of the whole network, so there is need of the ability to rapidly understand what is happening on the networks in short time. For administrator it is tedious task to understand the data properly and to take decision correctly. Visualization helps to detect expected and discover the unexpected, reduces analysis and response time and to make decision properly.

Three important goals were to select an approach that

(1) Can be efficiently used even by humans without a deep understanding of the systems that are analyzed.

(2) Can be applied to virtually all textual log files without any significant adaptation or configuration.

(3) Scales well up to large log files (e.g. log files with millions of log lines per day).

2.17 Stages of the visualization process:

Capture network traffic

Network traffic is very important because all activities can be identifying by capturing the data. Many tools are used to capture the traffic on the network for further analysis. Tcpdump is a network capturing tool which captures all the packets on the network.

Analysis data

The capture data consist of all the information residing on the network but network administrator is interested only on the specific data. by analyzing the data only required fields can be separated and redundancy of same data can be removed.

Visualization tool

The analyzed data is the data in which network administrator is interested. So this data can be used by the various tools to draw the pictures. Here the pictures can be customized with different colors and shapes.

2.18 Current visualization applications:

• AfterGlow (Christian & Raffy, 2007)

• TreeMap (HCIL, 2008)

• InteVis (Van Riel, 2007)

• Cytoscape (UCSD, 2008)

• GraphViz (Low, 2004)

• TNV (Goodall, 2007)

• NvisionIP (NCSA, 2004)

• Rumint (G. Conti, 2007)

• Nazar (Roth, 2008)

• Skyrails (Widjaja, 2007)

• Sequoia (Bruls, Geerlings, & Van Ham, 2002) [35].

2.18.1 MRTG Tool

The original MRTG program was a Perl script which used external utilities to do SNMP queries and to create GIF images for display on the HTML pages [36]. The feedback of this tool highlighted two key problem areas which are scalability and portability.

MRTG logged its data to an ASCII file, rewriting it every five minutes, constantly consolidating it, so that the log file would not grow over time. The log file did only store slightly more data than was needed to draw the graphs on the web page. The graphs were converted to GIF format by piping a graph in PNM format to the pnm to gif tool from the PBM package. This setup limited MRTG to monitor about 20 router ports from a workstation. A second obstacle for potential users was that MRTG required SNMP get from the SNMP package. This package proved to be rather difficult to compile on various platforms at that time.

2.18.2 Afterglow

Afterglow is a script based visualization tool built for showing the relationships among the entities of the data input. It supports mainly two types of visualizations, namely network graph and tree map [37]. Afterglow can be applied to various types of data sources such as network packet captures in pcap format, email logs, firewall logs, firewall rule-sets, web logs, IDS logs, OS logs, etc., but the main experience in its current usage is on network packet captures and email logs. It chooses a well known open source graph (network) visualization library named Graphviz as its

visualization back end, which is be very fast and scalable but does not directly support interaction.

2.18.3 RRD tool

The RRDTtool allows updating the log file at any time. It will automatically interpolate the value of the data-source at the latest official time-slot and write this value to the log [38].

Steps to use RRDtool for data graphing

1. Create an empty RRD database using rrdtool create.

2. Utilize a script and/or the cron to add data to the database using rrdtool update.

3. Generate, usually via script, custom output graphs using rrdtool graph.

Advantages of RRD Tool:

- Powerful,
- feature-rich

Disadvantages of RRD Tool:

- Complicated,
- database not scalable,
- database not platform independent

These visualization tools like MRTG, Afterglow and RRD tools are used to show data pictorially and help the administrator to take decision effectively.

Chapter 3

Problem statement

As per literature review carried out log files provides the whole information about all activities performing on the network. These files take huge storage which make difficult for the administrator to maintain and read the valuable information from it. The format of log files which are maintained by the systems consists of many fields. It contains redundant records and in very large record the critical information is mostly ignored by the administrator. The main problem is to find out records that contain the necessary information of events in which administrator is interested.

Objectives

- **Visualization of Log files:**

  The massive amount of data stored in log files is cumbersome for administrator to analyze the network traffic. Often, important details are overlooked and it is difficult to get an overall picture of what is occurring in the network by manually traversing the textual logs. The problem is to visualize the activities within a network. The log file presentation is an overview where administrator can get a general sense of network activities and easily detect anomalies. Visualizing log is faster and easier than analyzing text logs, and this technique makes effectively scale and display the data.

- **Analyzing Network traffic:**

  The second problem is how to understand and interpret the data from the log files in a more intuitive manner. Analyzing the log files helps the administrator with intuitive information on the dependant relationships, which may help many other important problems such as identifying compromised machines, security tracing and fault localization.

- **Identifying Compromised host:**

  The fundamental motivation of the system is for an administrator to know what activity is happening on network and which service cause to compromise others. Since log file contains all information about each activity but to identify the compromised system is a tedious job for the administrator. Visualization of log files provide the precise view in which the compromised host may have different color scheme or different graphical view from the regular images which can be easily detected by the human eye.

Chapter 4

Proposed Solution and Implementation

4.1 Proposed Framework for Log Visualization

In order to visualize the log file, following framework is proposed in which raw data on the network is captured through various network security solutions such as firewalls, Intrusion Detection System and honeypots etc. The log files maintained by these systems should be parsed in desired fields in which administrator is interested.

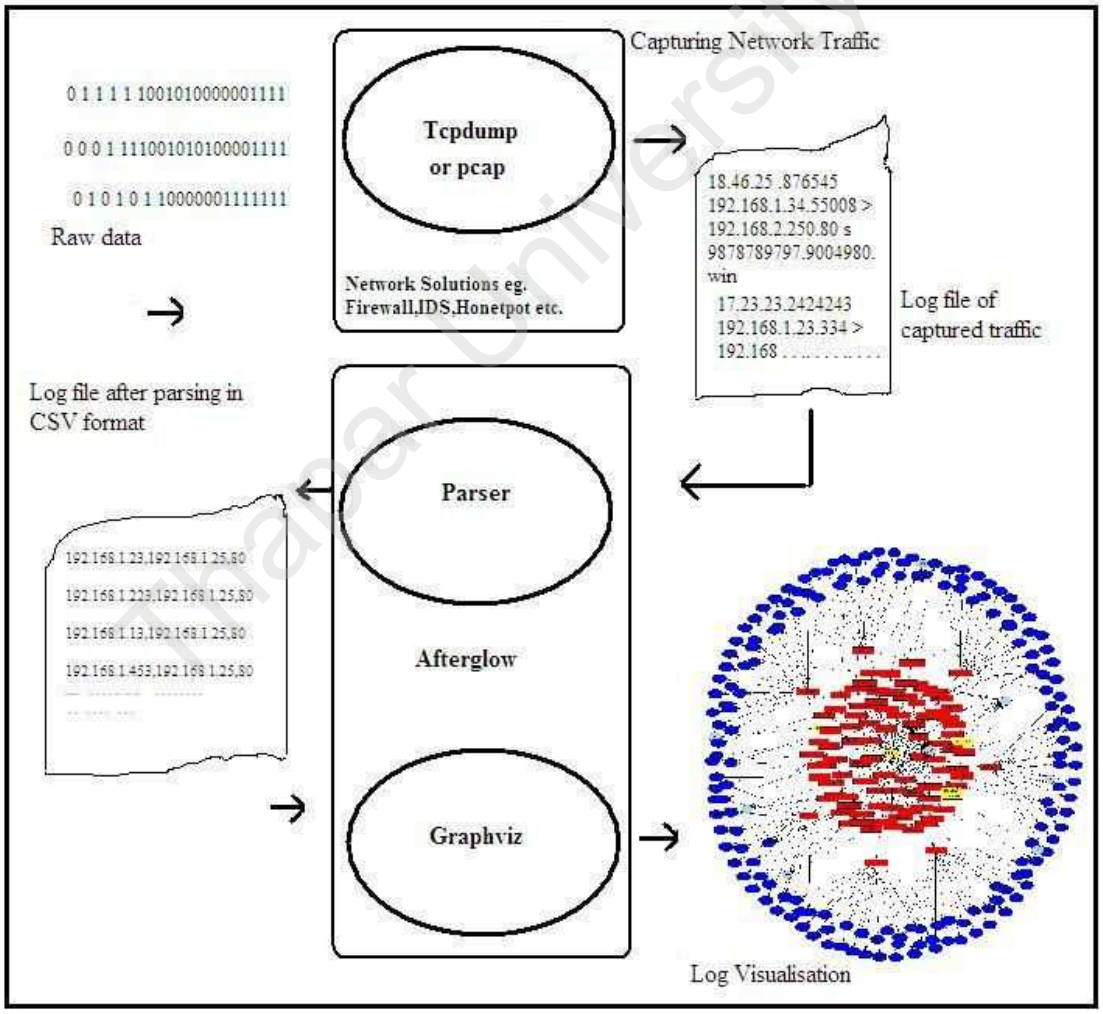


Figure 4.1: Framework for Log Visualization.

The visualization tool like afterglow is used to make a structure of digraph which is used by graphviz to make finally gif images. In these images the desired information like timestamps, source IP addresses, destination IP addresses, source ports, destination ports etc. are viewed in different color of nodes to extract the information of different events.

4.2 Implementation and Results

Network traffic visualization techniques aid an administrator in recognizing attacks in a pictorial form. Instead of reading huge text visualize the data with the graph using various data visualization tools. Visualizing the log by the administrator is very important to detect the compromised hosts on the network. The aim of Log visualization is to facilitate the administrators and executives of an organization can view the network for effective monitoring of the networks through the logs. The compromised host can be highlighted so that prevention actions can be taken. After Glow is a script based visualization tool built for showing the relationships among the Entities of the data input.

4.2.1 The Virtualization software

VMware Server software allows to optimize and manage the IT infrastructure from the desktop to the data center,by virtualizing computing, storage and networking systems. VMware products create enterprise-class virtual machines that increase server and other resource utilization, improve performance, increase security and minimizesystem downtime, reducing the cost and complexity ofdelivering enterprise services.VMware Server Console is used as a Virtualization software.Virtualization is a proven software technology that is rapidly transforming the IT landscape and fundamentally changing the way that people compute. Today's powerful x86 computer hardware was designed to run a single operating system and a single application. This leaves most machines vastly underutilized. Virtualization runs multiple virtual machines on a single physical machine, sharing the resources of that single computer across multiple environments. Different virtual machines can run different operating systems and multiple applications on the same physical computer [39]. The exact specifications are as follows:

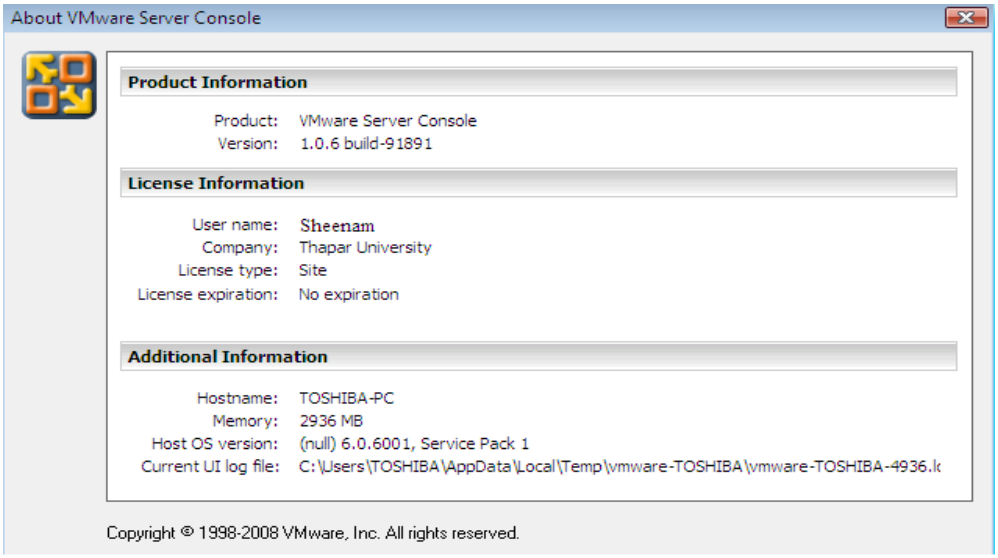


Figure 4.2: Virtualization Software.

4.2.2 Fedora 8 under VMware

Fedora is an open, innovative, forward looking operating system and platform, based on Linux, that is always free for anyone to use, modify and distribute, now and forever. It is developed by a large community of people who strive to provide and maintain the very best in free, open source software and standards. The Fedora Project is managed and directed by the Fedora Foundation and sponsored by Red Hat.

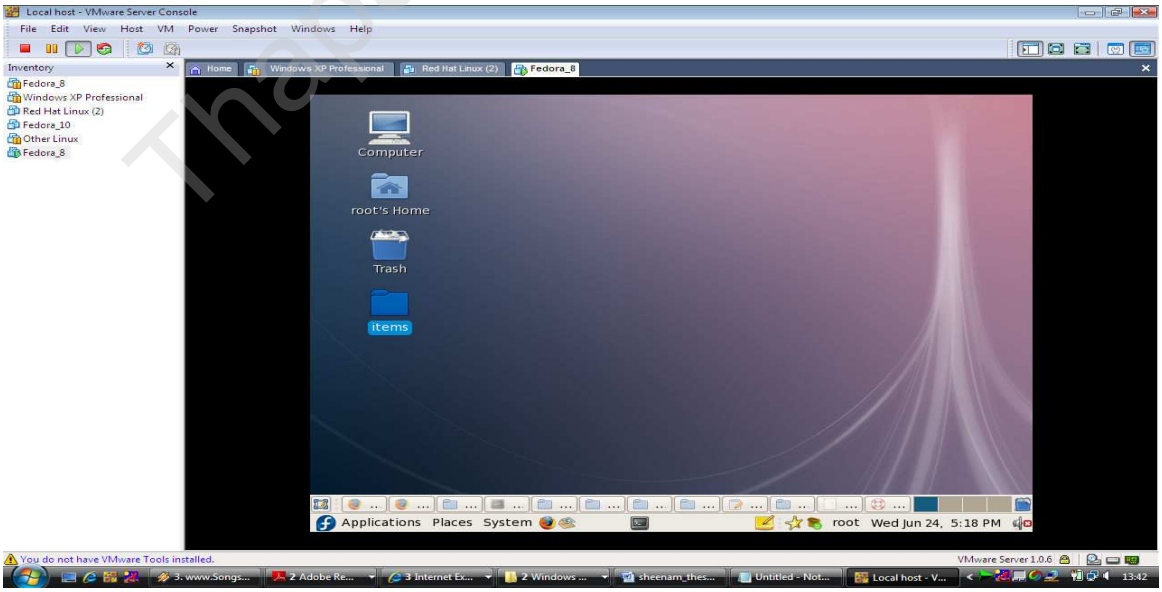


Figure 4.3: Fedora 8 under VMware.

4.2.3 Capturing Network traffic

In order to implement the visualization of log files the first step is to capture the network traffic. For capturing the packets from the network traffic various tools can be used like tcpdump, ethereal, wireshark.the captured data packets can be used for active and passive analysis. In active analysis the data packets are analyzed at the same time when they are captured where in passive analysis packets once stored and analyzed later in future. All of these tools are developed by using popular programming library called libpcap that provides a high level interface to packet capture [40].

Tcpdump

Tcpdump is one of the best data sources for a network security administrator is raw network traffic, which provides untainted, unfiltered network data. It is one of the oldest and most popular tools for obtaining raw network data, which uses the pcap (packet capture) library to obtain all data travelling through an interface, such as a network card. Problems can be detected via raw network traffic only with a thorough understanding of the data.

Using Tcpdump for network traffic capturing

Tcpdump provides the description of the contents of the packets on a network interface that match the Boolean expression.

tcpdump -vttttnneli eth2

When tcpdump finishes capturing packets, it will report of packet captured, Packets received by filter(the meaning of this depends on the operating system on which tcpdump is running and possible on the way the operating system was configured. If a filter was specified on the command line. On some operating system it counts packets captured and packets dropped by kernel. When it runs with – v flag it provides **-v** runs in the verbose output with time to live, identification, total length and options in an IP packet. Also enables additional packet integrity checks such as

verifying the IP and ICMP header checksum. -tttt print a timestamp in default format proceeded by date on each dump line.

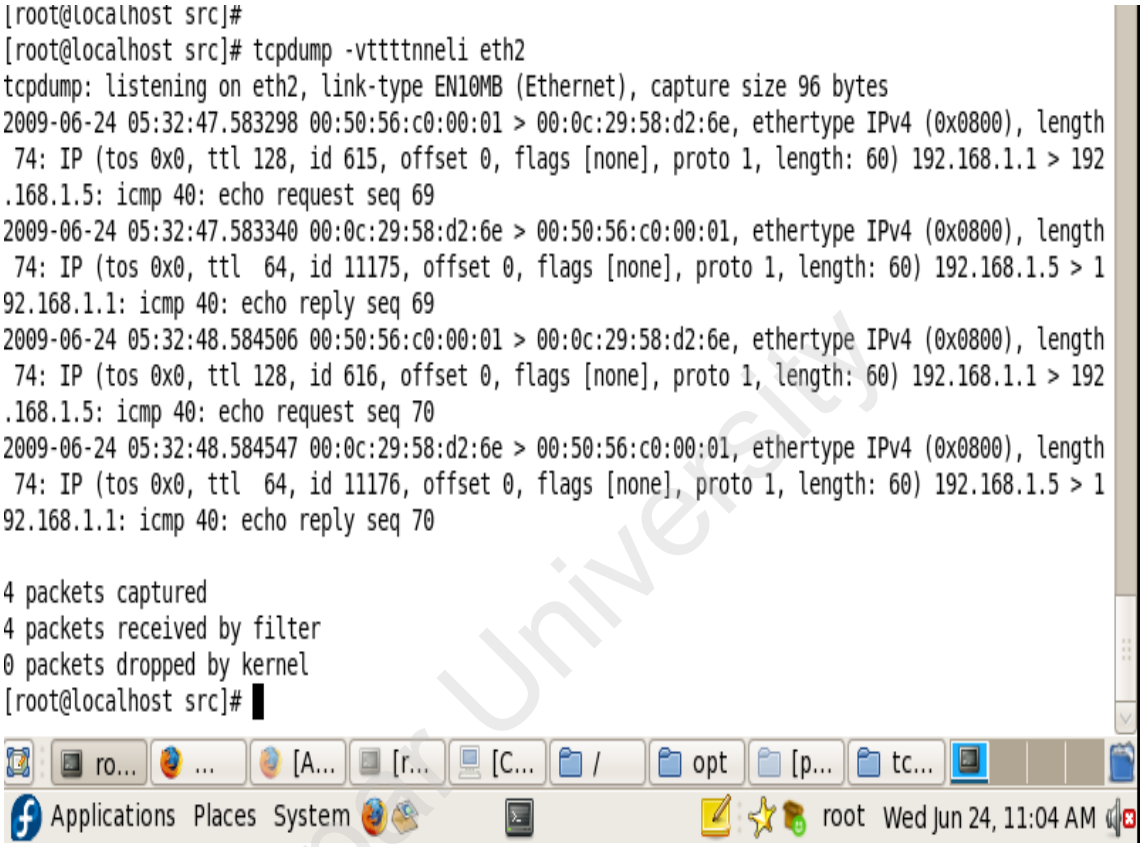


Figure 4.4: Capturing data using tcpdump.

Parsing the output of Tcpdump

The parser takes the tcpdump pcap file and parses it into a CSV output. The CSV file does not show the exact output of tcpdump it shows the source and destination inverted because the parser remembers the source of a communication and automatically inverts the responses to reflect that behaviour. It outputs the direction of the communication (client to server) and not the direction of the packets. This is very useful when visualizing network traffic.

tcpdump -vttttnneli eth2 | parsers/tcpdump2csv.pl "sip dip dport"

This command invokes tcpdump on interface eth2 and pipe the input through the parser. "sip dip dport " line tells the parser that the source IP (sip), the destination IP (dip) and the destination port (dport) are required. There are other fields available in the parser but the output of this command is a comma separate list of sip, dip, dport pairs.

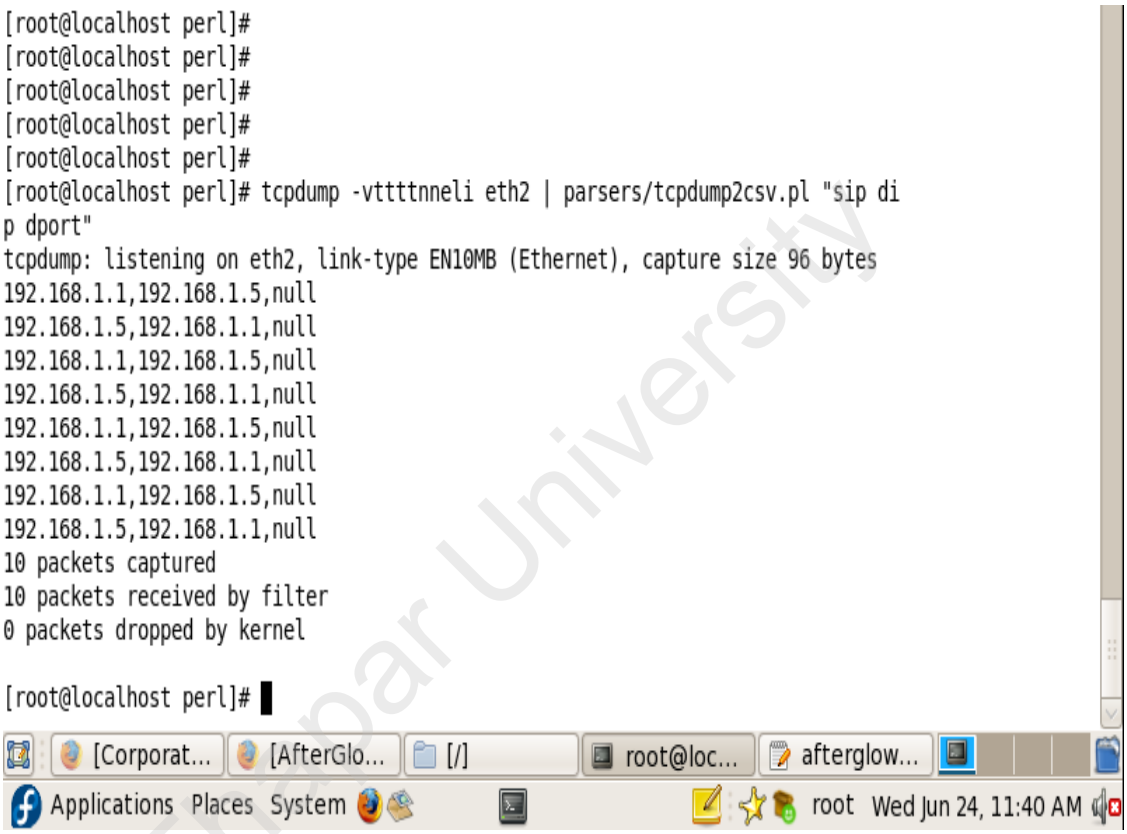


Figure 4.5: Parsing of data fields.

4.2.4 Visualizing logs

Graphviz is open source graph visualization software. It takes standard textual input and can automatically generate graphs depending on the input itself [41]. Graphviz and Afterglow provides a unique overview of the incoming and outgoing traffic and a much broader overview of what is coming in and out of immediate network. Afterglow is a collection of scripts which assist in converting csv and other table data into information which can be pushed off to Graphviz to generate graphs with.

To generate a dot graph file for graphviz, run the following command:

cat file.csv | perl afterglow.pl -c color.properties > file.dot

This command invokes afterglow.pl with color.properties which specified a color property to determine the colors of the edges and nodes in the graph.

digraph structs{

node [shapes=ellipse,style=filled, fontsize=10];

edge [len=3];

}

cat files.csv| neato -Tgif -o test.gif

neato is a program that makes layouts of undirected graphs following the filter model of DOT. AfterGlow expects two values on each line. Each line then represents two nodes and a connection between the nodes using the input, AfterGlow will produce output that can be passed on to one of the utilities from GraphViz [42].

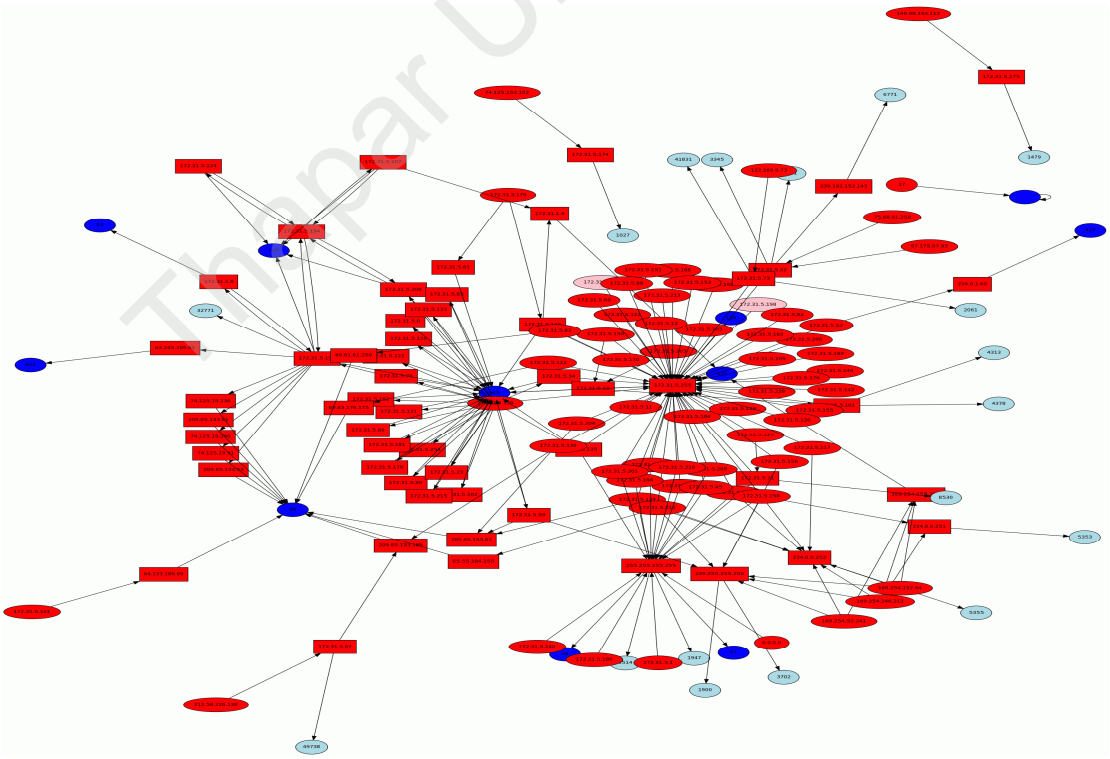


Figure 4.6: Connections on the server.

Visualising log files provide pictorial view of the machines communicating with the server.

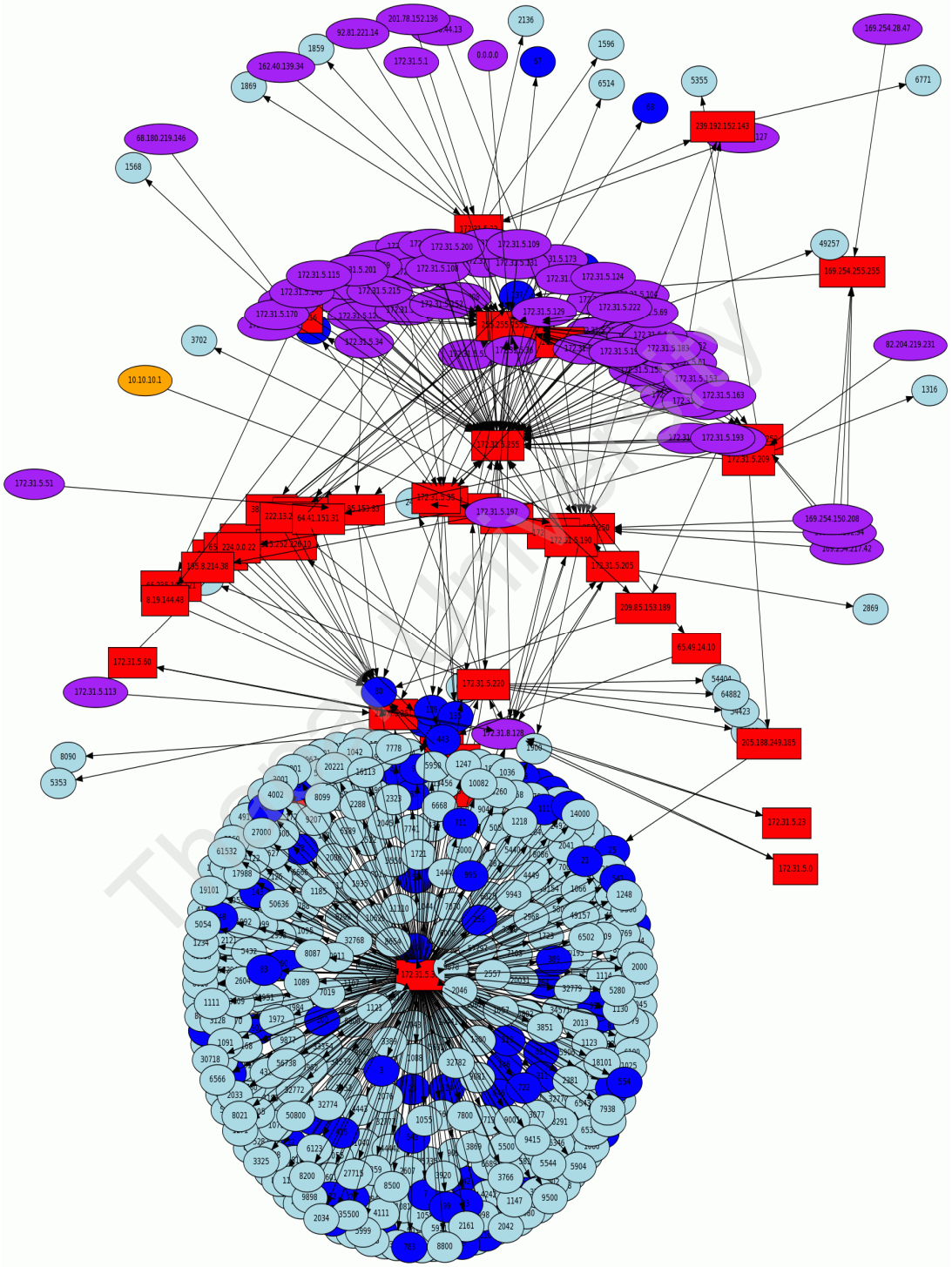


Figure 4.7: Graph showing the network traffic of Thapar University.

4.2.5 Identifying Compromised Host

A system seems to be compromised by different types of attacks, denial of service attack, smurf attack and flooding attacks send packets in the access amount so that the machine gets busy in response. In order to find out this type of attack from the log is a tedious job for the administrator because log contains thousands of packets and identifying some desired records is very time consuming.

Visualizing records of compromised host have a unique pictorial representation. As compromised host also response to the attacker so node contains both incoming and outgoing arcs. For example the following figures show that the machine 172.31.5.10 is compromised due to so many connections of different machines sending and receiving request.

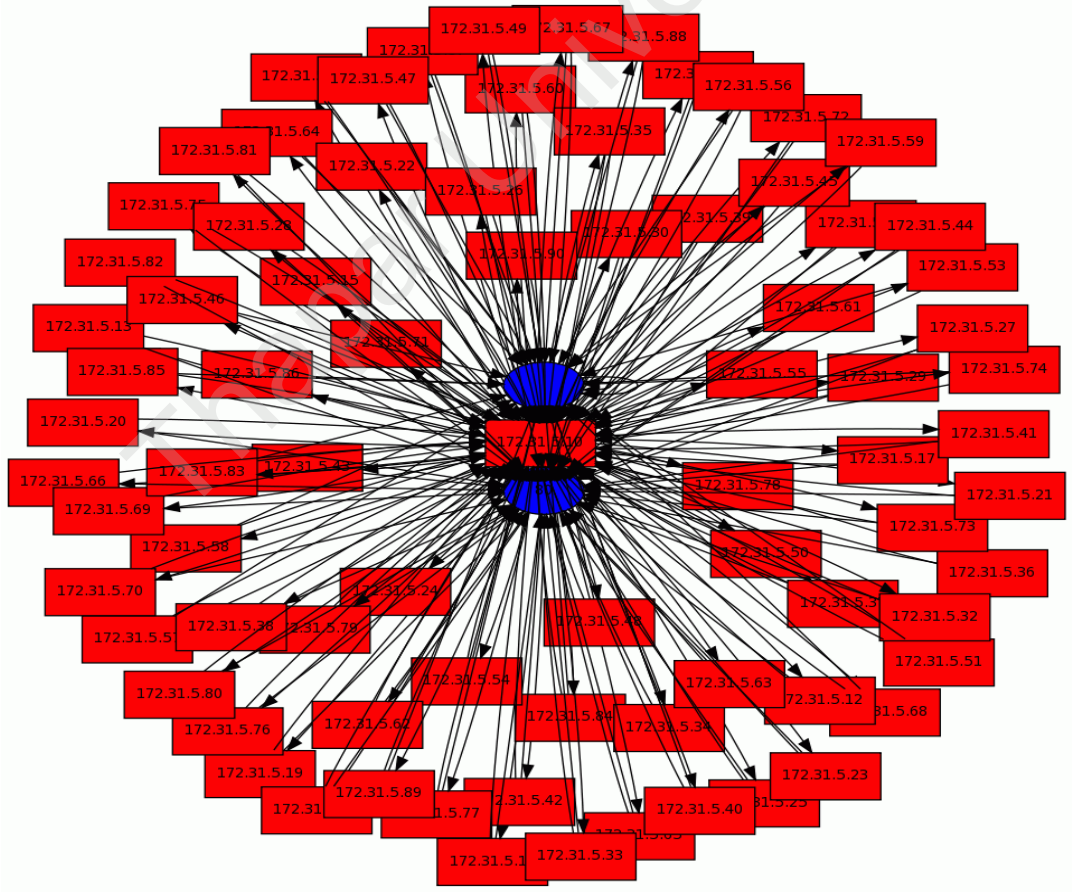


Figure 4.8: Compromised machines 172.31.5.10.

So the same pattern identified in the picture shows that the node is being compromised. Let there is a log of traffic captured in the college' network in which sources, events and targets are depicted with different color. In the following figure the rounded area shows that a node whose IP address is 172.31.5.99 is sending and receiving requests to some systems.

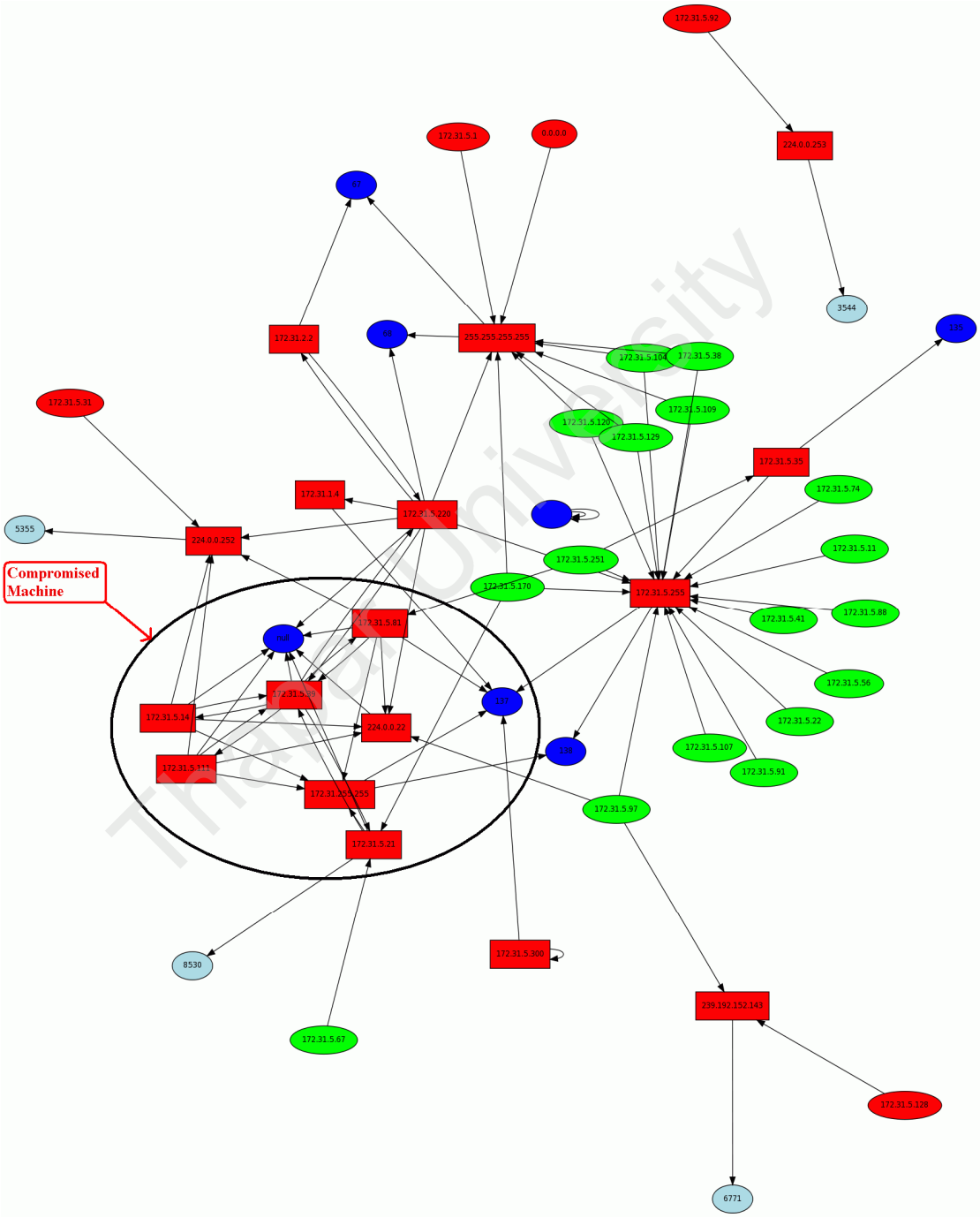


Figure 4.9: Showing compromised host.

By taking the closer view as shown in figure 4.10 there are six machines which are communicating with the machine 172.31.5.99.The observed patteren shows that the machine is production machine or a compromised machine which makes easy for the administrator to identify the malicious activities performing on the network.

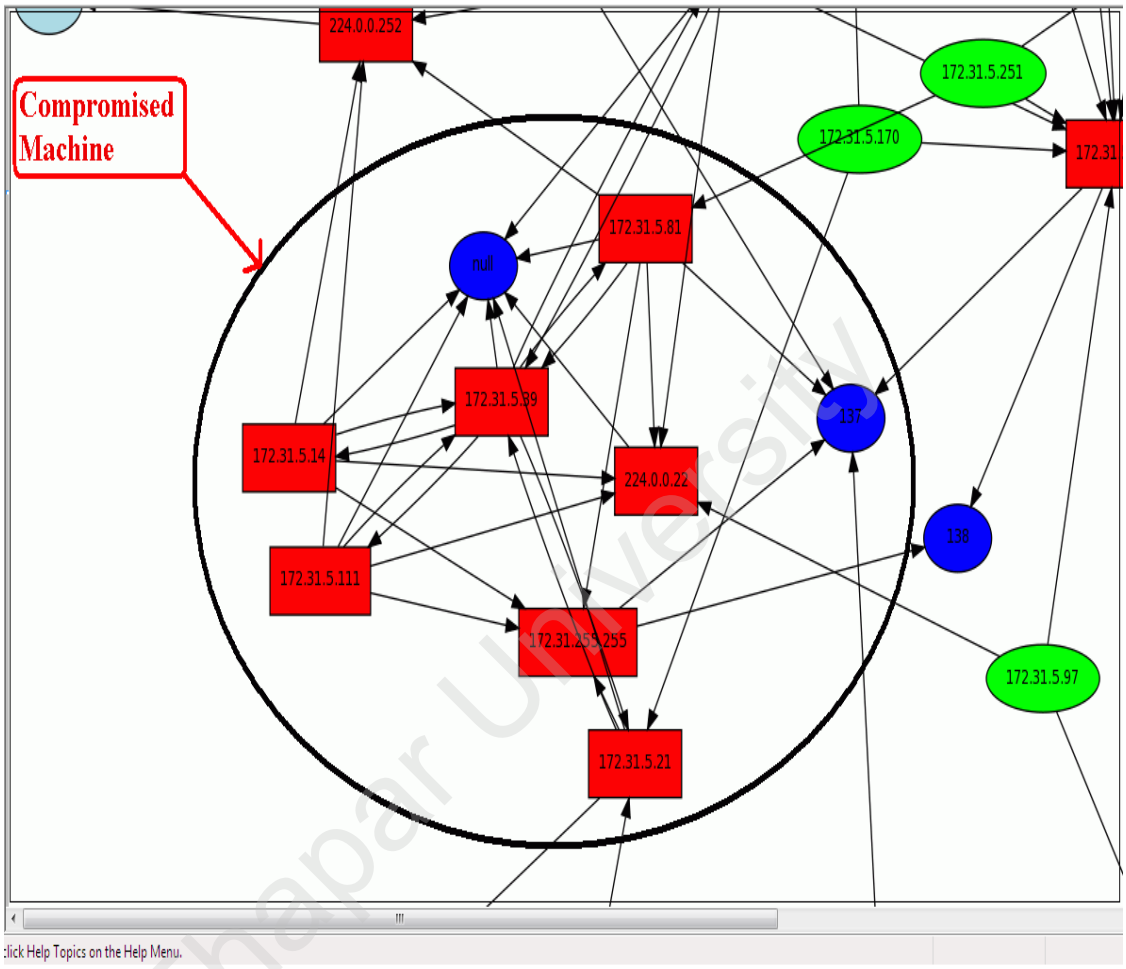


Figure 4.10 : Closer view of machine 172.31.5.99(compromised)

4.2.6 Visualizing Log for a particular source IP Address

The traffic of college' network has been captured by the tcpdump and parsed into two fields.

1. Source IP Address.

2. Destination IP Address.

Network traffic of 4339 packets are captured and no packet is dropped by the kernel. is captured The graph generated by the system is shown in figure

Figure 4.11: Packets captured for extracting a particular source IP address.

For example let an administrator wants to view the graph of only those IP addresses which are associated with a particular IP address 172.31.5.159.By restricting the source variable as 172.31.5.159 in afterglow.pl script will show the graph of desired IP addresses whose source IP address is 171.31.5.159.

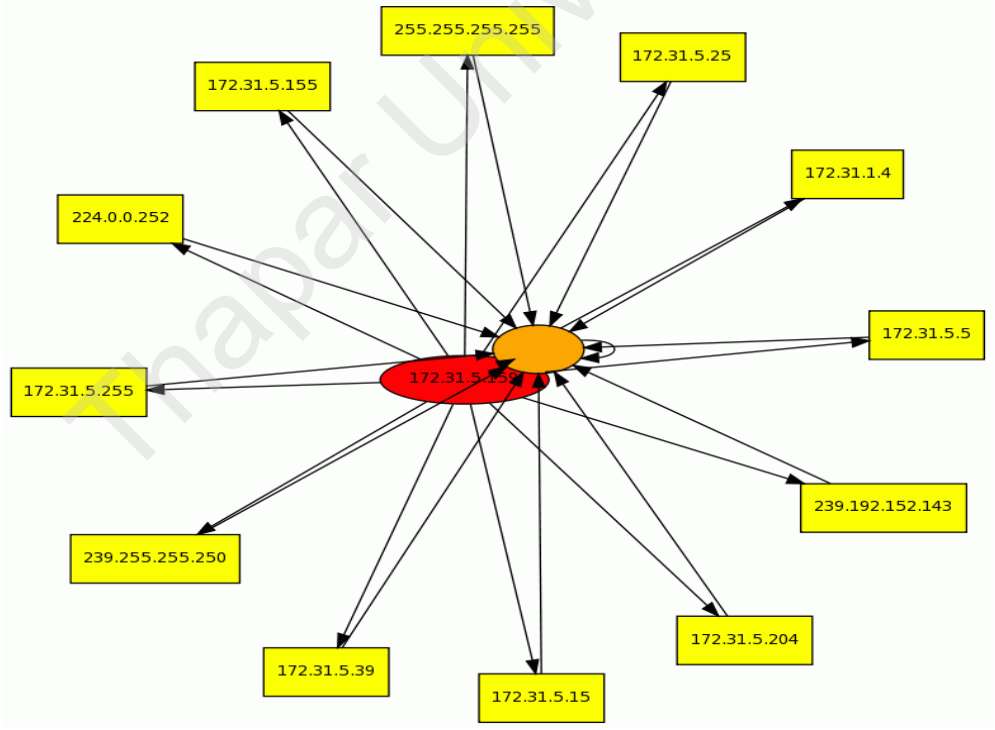


Figure 4.12: Machine 172.31.5.159 is communicating with 12 IP addresses.

The result shows that IP address 172.31.5.159 is associated with only 12 IP addresses as source machine out of 4339 packets.

Chapter 5

Conclusions and Future Scope

The work presented in this thesis shows the importance of Log visualisation in the world of Network Security. Visualising log files that contain the information of every events processed on the network provide ease to administrator to analyze, explore and understand the data so that decision can be taken accordingly. Visualization plays an important role in detection of malicious activities performing on the network by presenting the same in the pictorial form.

Network Security solutions not only prevent from the intruder' but also take the record of the every activity in the form of log files. This log contains the information of compromised machines and the malicious events performing on the same can be easily identified by visualizing the log files.

In this thesis the main focus is to identify the compromised hosts on the network through visualizing the log files which is a tedious job for the administrator in the case of text log files.

The robustness of a network depends on the dynamic, interactive visualization coupled with real-time active network statistics. The real time system will collect all the log files on sight and will provide different graphs according to current status of the network and helps the administrator to find the compromised host on the network and will also produce various visual information to an administrator for better decision making in time.

The framework proposed for log visualization takes the data from the database stored in a specific format. Since the visualization system takes the log files already stored by the network security solutions and make it offline visualization so more work is required to store the data in different formats. In future the work can be done to make

a realtime visualization system which produces the different types of graph instantly to help the administrator to take effective decision on the same time.

References

[1] Frederick. Scott Pinzon,"Producing your network security policy",Watchguard Technologies,July 2007.

[2] Whitepaper, “Understanding the evolution of Network Security”, January 2008.

[3] R. Deepak, Timothy A.,Gonsalves, Hema .A. Murthy and N. Usha "Network Security Management for a National ISP", IIT Madras, December 2004.

[4] Whitepaper"Analyzing Logs for Security Information Event Management", Zoho Corp, 2005.

[5] Thorsten Kisner, Alex Essoh and Firoz Kaderali"Visualisation of Network Traffic using Dynamic Co-occurrence Matrices" Department of Communication Systems,Faculty of athematics and Computer Science FernUniversit¨at in Hagen, Germany, April 17, 2007.

[6] Muhammad Ghanbari,"Scalability of visualization’s evaluation",Department of Computer Science, Alabama, Southeastcon, 2008. IEEE, april 2008.

[7] Kulsoom Abdullah,"Scaling and Visualizing Network Data to Facilitate in Intrusion Detection Tasks", Georgia Institute of Technology, May 2006.

[8] Paul Innella, “A Brief History of Network Security and the Need for Adherence to The software Process Model”, Tetrad Digital Integrity.

[9] Bruce Schneier,“Managed Security Monitoring: Network Security for the 21st Century”, Counterpane Internet Security, December, 2005.

[10] Blain R. Jones, “Network Security: An Open-Source Approach”, July, 2005.

[11] Prof P L Pradhan “VIRUSE A System Enemy “,The Infosec Writers Tex Library,August 2005.

[12] Chad Parks,"Viruses and Worms: The Best Defense is Awareness",Canaudit, Inc.,October 2003.

[13] Whitepaper, “Computer Viruses and Spyware Henson Computer Services”, Computer Viruses and Spyware, Henson Computer Services – December 21, 2004.

[14] Alex Noordergraaf ,“How Hackers Do It:Tricks, Tools, and Techniques”, Enterprise Server Products Sun BluePrint OnLine—May, 2002.

[15] D.W. Park,"A Study of Packet Analysis regarding a DoS Attack in WiBro

Environments",Hoseo Graduate School of VENTURE . IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.12, December 2008.

[16] Kevin J. Houle,George M. Weaver,Neil Long,Rob Thomas,"Trends in Denial of Service Attack Technology",Carnegie Mellon University,October 2001.

[17] David Minutella, Jeremy Cioara, Heather Stevenson,"General Network Security",CCENT Exam Prep,April 2008.

[18] Whitepaper, "Social Engineering – Are You at Risk?",DAS Information Security Office, May 2008.

[19] Gary Miliefsky, "A guide to proactive network security",ZDNet News,November 2004.

[20] Whitepaper, "Firewall Gateways" AT&T and Lumeta Corporation,1994.

[21] Lisong Pei, Jakob Schütte,"Intrusion detection systems ", Carlos Simon 2007.

[22] Whitepaper, "Intrusion Prevention A Proactive Approach to Network Security", VeriSign, May 2005.

[23] Reto Baumann,Christian Plattne"White Paper:Honeypots",March 2002.

[24] Lance Spitzner,"Honeypots,Definitions and Value of Honeypots",May, 2003.

[25] Nirbhay Gupta, "Is Honeyd Effective or Not?", Edith Cowan University, Western Australia,November 2003.

[26] Alisha Cecil,"A Summary of Network Traffic Monitoring and Analysis Techniques",http://www.cse.wustl.edu/~jain/cse567-06/net_monitoring.htm.

[27] Bjom Landfeldt,Pipat Sookavatana,Aruna Seneviratne,"The Case for a Hybrid PassivdActive Network Monitoring Scheme in the Wireless Internet",IEEE,2000.

[28] Whitepaper, "The Case for a Hybrid Passive/Active Network Monitoring Scheme in the Wireless Internet", Singapore, CS Digital Library, September 2008.

[29] M. Souppaya, K. Kent, "Guide to Computer Security Log Management", National Institute of Standards and Technology, Aithersburg, MD, 2006.

[30] KarenKent,Murugiah Souppaya,"Guide to Computer Security Log Management", NIST U.S. Department of commerce, September 2006.

[31] Guillermo Francia, Monica Trifas Dorothy Brown,Rahjima Francia,Chrissy Scott"Forensic Data Visualization System: Improving Security through Automation" Computer Security Conference,USA, April 2007.

[32] Seok Hee Hong "Network Analysis and Visualisation" National ICT Australia, School of Information Technologies, University of Sydney, Australia, January 2006.

[33] Seong Soo Kim and A. L. Narasimha Reddy, "NetViewer: A Network Traffic Visualization and Analysis Tool", Texas A&M University, Large Installation System Administration Conference,2005.

[34] Johannes Lessmann, Tales Heimfarth, "Flexible Offline-Visualization for MobileWireless Networks", Proceedings of the Tenth International Conference on Computer Modeling and Simulation,2008.

[35] Iain Swanson, "Malware, Viruses and Log Visualisation",Proceedings of the 6th Australian Digital Forensics Conference,2008.

[36] Tobias Oetiker, "MRTG The Multi Router Traffic Grapher",Proceedings of the Twelfth Systems Administration Conference (LISA '98),Boston, Massachusetts, December 6-11, 1998.

[37] Jaime Blasco Aitsec,Madrid, Spain, "An approach to malware collection visualization",OSSIM, August 2008.

[38] Ben Rockwood, "Getting Started with RRDtool" Cuddletech TekRef Series",www.cuddletech.com/articles/rrd/index.html,May 2004.

[39] http://www.vmware.com/technology/virtualization.html,VMwareServer Console.

[40] http://linux.softpedia.com/get/Programming/Libraries/tcpdump-169.shtml., About Tcpdump.

[41] http://www.graphviz.org/Download_linux_fedora.php.,Graphviz tool.

[42] http://afterglow.sourceforge.net/manual.html, Manual Afterglow Visualization Tool.

Papers Communicated

Sheenam Goyal, Dr. Maninder Singh, "Log Visualization to track Compromised Host on the Network" communicated in 3rd IEEE International Symposium on Advanced Networks and Telecommunication Systems, IEEE ANTS 2009 in India, New Delhi from 14-16 December 2009(**Communicated**).