Sinhgad Institutes

Sinhgad Technical Educational Society's

SINHGAD ACADEMY OF ENGINEERING

KONDHWA

Lab Manual

Lab Practice IV (STQA and CSDF)
BE Computer
AY:-2022-2023

Prepared By,

Ms. P . S . G a w a l i

Mr. V.K. Sambhar

Marking Scheme

50 marks-Term work

Study material provided by: Vishwajeet Londhe

Join Community by clicking below links

Telegram Channel

https://t.me/SPPU_TE_BE_COMP

(for all engineering Resources)

SPPU Engineering & Technical UPDATES
WhatsApp channel

WhatsApp Channel

(for all Engg & tech updates)

https://whatsapp.com/channel/0029ValjFriICVfpcV9HFc3b

@SPPU_ENGINEERING_UPDATE

Insta Page

(for all Engg & tech updates)

https://www.instagram.com/sppu_engineering_update

Sinhgad Institutes

Sinhgad Technical Educational Society's

SINHGAD ACADEMY OF ENGINEERING KONDHWA

CERTIFICATE

This is to certify that,

Mr/Ms....., of class .....Roll No..... has completed all the practical work in the subject "Cyber Security and Digital Forensics" satisfactorily in the Computer Engg. Department as prescribed by Savitribai Phule Pune University, in the the academic year 2022-23

Staff In-charge Head of the Department Principal

Savitribai Phule Pune

University Bachelor of Computer Engineering

Program Outcomes (POs)

Learners are expected to know and be able to–

PO1	**Engineering knowledge**	Apply the knowledge of mathematics, science, Engineering fundamentals, and an Engineeringspecialization to the solution of complex Engineering problems.
PO2	**Problem analysis**	Identify, formulate, review research literature, and analyze complex Engineering problems reaching substantiated conclusions using first principles of mathematics natural sciences, and Engineering sciences.
PO3	**Design / Development of Solutions**	Design solutions for complex Engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health andsafety, and the cultural, societal, and Environmental considerations.
PO4	**Conduct Investigations of Complex Problems**	Use research-based knowledge and research methods including design of experiments, analysisand interpretation of data, and synthesis of the information to provide valid conclusions.
PO5	**Modern Tool Usage**	Create, select, and apply appropriate techniques, resources, and modern Engineering and IT tools including prediction and modeling to complex Engineering activities with an understanding of the limitations.
PO6	**The Engineer and Society**	Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
PO7	**Environment and Sustainability**	Understand the impact of the professional Engineering solutions in societal and Environmentalcontexts, and demonstrate the knowledge of, and need for sustainable development.
PO8	**Ethics**	Apply ethical principles and commit to professional ethics and responsibilities and norms ofthe Engineering practice.
PO9	**Individual and Team Work**	Function effectively as an individual, and as a member or leader in diverse teams, and inmultidisciplinary settings.
PO10	**Communication Skills**	Communicate effectively on complex Engineering activities with the Engineering community and with society at large, such as, being able to comprehend and write effective reports anddesign documentation, make effective presentations, and give and receive clear instructions.
PO11	**Project Management and Finance**	Demonstrate knowledge and understanding of the Engineering and management principles andapply these to one‘s own work, as a member and leader in a team, to manage projects and in multidisciplinary Environments.
PO12	**Life-long Learning**	Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes (PSO)

PSO1	**Professional Skills**-The ability to understand, analyze and develop computer programs in the areas related to algorithms, system software, multimedia, web design, big data analytics, and networking for efficient design of computer-based systems of varying complexities.
PSO2	**Problem-Solving Skills**- The ability to apply standard practices and strategies in software project development usingopen-ended programming environments to deliver a quality product for business success.
PSO3	**Successful Career and Entrepreneurship**- The ability to employ modern computer languages, environments, andplatforms in creating innovative career paths to be an entrepreneur, and a zest for higher studies.

Course Outcomes:

At the end of the course, the students will be able to:

CO1: Analyze threats in order to protect or defend it in cyberspace from cyber-attacks.

CO2: Build appropriate security solutions against cyber-attacks.

CO3:Underline the need of digital forensic and role of digital evidences.

CO4: Explain rules and types of evidence collection

CO5: Analyze, validate and process crime scenes

CO6: Identify the methods to generate legal evidence and supporting investigation reports.

INDEX

Sr. No.	Name of Assignment	Page No.	Date	Remark
1.	Tracking Emails and crimes related to Emails			
2.	Generation and verification of CAPTCHA image			
3.	Recovery of permanent deleted files and deleted partitions			
4.	Log Capturing and event Correlation			
5.	Honey pot			

Assignment No.	**1**
Title	Tracking Emails and crimes related to Emails
Roll No.	
Class	B.E. Computer
Date	
Subject	**Cyber Security and Digital Forensics**
Signature	

Assignment No. 1

Title: Tracking Emails and crimes related to Emails

Aim: Write a program for tracking Emails and investigating Email Crimes.

Objectives: The objective of this lab is to provide expert knowledge on tracking emails, investigating email crimes, and other responsibilities

Hardware/Software requirements: PC

Theory:

Investigating email crimes is the process of tracing, collecting, analyzing, and investigating digital evidence and cyber trails. Digital evidence and cyber trails can relate to email spamming, mail bombing/mail storms, email spoofing, identity fraud/chain letters, phishing attacks, and email hijacking.

Email Architecture:

When a user sends an email to a recipient, this email does not travel directly into the recipient's mail server. Instead it passes through several servers. The MUA is the email program that is used to compose and read the email messages at the client end [1]. There are multiple MUAs available such as Outlook express, Gmail, and Lotus Notes. MTA is the server that receives the message sent from the MUA. Once the MTA receives a message it decodes the header information to determine where the message is going, and delivers the message to the corresponding MTA on the receiving machine [1]. Every time when the MTA receives the message, it modifies the header by adding data. When the last MTA receives the message, it decodes it and sends to the receiver's MUA, so the message can then be seen by the recipient. Therefore an email header has multiple pieces of server information, including IP addresses.

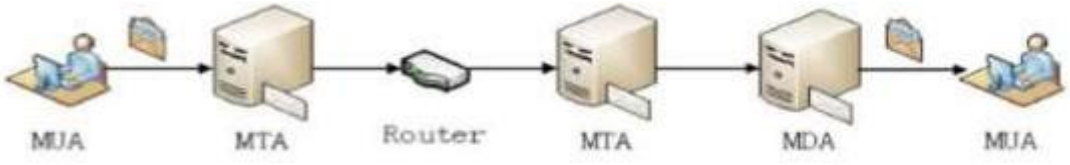


Figure 1: Email Architecture

Email Identities and Data:

The primary evidence in email investigations is the email header. The email header contains a considerable amount of information about the email. Email header analysis should start from bottom to top, because the bottom-most information is the information from the sender, and the top-most information is about the receiver. In the previous section it was shown that email travels through multiple MTAs. These details can be found in the email header. The following picture depicts a sample header.

```
Delivered-To: MrSmith@gmail.com
Received: by 10.36.81.3 with SMTP id e3cs239nzb;Tue, 29 Mar 2005 15:11:47
-0800 (PST)
Return-Path: MrJones@emailprovider.com
Received: from mail.emailprovider.com (mail.emailprovider.com
[111.111.11.111]) by mx.gmail.com with SMTP id h19si826631rnb; Tue, 29
Mar 2005 15:11:47 -0800 (PST)
Message-ID: <20050329231145.62086.mail@mail.emailprovider.com>
Received: from [11.11.111.111] by mail.emailprovider.com via HTTP; Tue,
29 Mar 2005 15:11:45 PST
Date: Tue, 29 Mar 2005 15:11:45 -0800 (PST)
From: Mr Jones
Subject: Hello
To: Mr Smith
```

Figure 2: E-mail header

In order to understand the header information, it is necessary to understand the structured set of fields available in the header. The following are some of the basic field names and descriptions.

Table 1: E-mail header fields and its description

Field Name	Description
From	E-mail address (sometimes names) of the author(s) of the e-mail
To	The e-mail address(es) (sometimes names) of the message recipient
Cc	Carbon Copy
Bcc	Blind Carbon Copy
Subject	A summary of the topic
Date	The local time and date when the message was written
Reply-to	Address that e-mail reply will redirected to
Message-ID	Globally unique message identification string generated when it is sent
References	Identifies other documents related to this message, such as other e-mail message
Received	Tracking information generated by mail servers that have previously handled a message, in reverse order

Email Forensic Investigation Techniques:

Email forensics refers to analyzing the source and content of emails as evidence. Investigation of email related crimes and incidents involves various approaches.

Header Analysis

Email header analysis is the primary analytical technique. This involves analyzing metadata in the email header. It is evident that analyzing headers helps to identify the majority of email-related crimes. Email spoofing, phishing, spam, scams and even internal data leakages can be identified by analyzing the header.

Server Investigation

This involves investigating copies of delivered emails and server logs. In some organizations they do provide separate email boxes for their employees by having internal mail servers. In this case, investigation involves the extraction of the entire email box related to the case and the server logs.

Network Device Investigation:

In some investigations, the investigator requires the logs maintained by the network devices such as routers, firewalls and switches to investigate the source of an email message. This is often a complex situation where the primary evidence is not percent (when the ISP or proxy does not maintain logs or lacks operation by ISP)
Software Embedded Analysis

Some information about the sender of the email, attached files or documents may be included with the message by the email software used by the sender for composing the email. This information may be included in the form of custom headers or in the form of MIME content as a Transport Neutral Encapsulation Format (TNEF)
Sender Mail Fingerprints

The "Received" field includes tracking information generated by mail servers that have previously handled a message, in reverse order. The "X-Mailer" or "User-Agent" field helps to identify email software. Analysing these fields helps to understand the software, and the version used by the sender.

```
X-Sender: chirath@aptt.com
User-Agent: Roundcube Webmail/1.0.6
```

Figure 3: Senders' E-mail software and version used in creating a message

Use of Email Trackers

In some situations, attackers use different techniques and locations to generate emails. In such situations it is important to find out the geographical location of the attacker. To get the exact location of the attacker, investigators often use email tracking software embedded into the body of an emaqil. When a recipient opens a message that has an email tracker attached, the investigator will be notified with the IP address and geographical location of the recipient. This technique is often used to identify suspects in murder or kidnapping cases, where the criminal communicates via email.
Volatile Memory Analysis

Recent research has been conducted in analyzing spoofed mails from volatile memory Since everything passes through volatile memory, it is possible to extract email related evidence (header information) from volatile memory
Attachment Analysis:

Most viruses and malware are sent through email attachments. Investigating attachments is crucial in any email-related investigation. Confidential information leakage is another important field of investigation. There are software tools available to recover email-

related data, such as attachments from computer hard discs. For the analysis of suspicious attachments, investigators can upload documents into an online sandbox such as Virus Total to check whether the file is malware or not. However, it is important to bear in mind that even if a file passes a test such as Virus Total's, this is not a guarantee that it is fully safe. If this happens, it is a good idea to investigate the file further in a sandbox environment such as Cuckoo.

Conclusion:

Thus the Email has been tracked with reference to the sender and the messages have been blocked.

Assignment No.	**2**
Title	Generation and verification of CAPTCHA image
Roll No.	
Class	B.E. Computer
Date	
Subject	**Cyber Security and Digital Forensics**
Signature	

Assignment No. 2

Title: Generation and verification of CAPTCHA image

Aim: Implement a program to generate and verify CAPTCHA image.

Objectives: The objective of this lab is to provide expert knowledge on tracking emails, investigating email crimes, and other responsibilities

Hardware/Software requirements: PC

Theory:

Captcha is a method used to protect websites against spam. The goal is to stop interactive websites from being spammed by filtering out automatically generated input. The acronym CAPTCHA stands for 'Completely Automated Public Turing test to tell Computers and Humans Apart'. As early on as the year 1950, the computer scientist Alan Turing suggested a method for testing the intellectual capacity of artificial intelligence. According to the computer pioneer, a machine is able to mimic the human mind when it manages to converse with people in a chat without then realizing it is a computer.

The Turing Test went down in the history of AI (artificial intelligence) research and was first passed by a computer program in 2014: As the first machine in the world, chatterbot *Eugene Goostman*, succeeded in deceiving more than 30 percent of an independent jury for at least 5 minutes. Eugene pretended to be a Ukrainian teenager with guinea pigs, who was also a big Eminem fan.

What is the purpose of captchas?

Captchas are usually used when web applications require user input. Imagine you are running an online store and want to give your customers the opportunity to write product reviews in a comments section. In this case, you want to ensure that the entries are actually from your customers or at least from human site visitors. You will often come across automatically generated **spam comments** – in the worst case linking to your competition.

You can reduce the risk of this happening by **protecting online forms with a captcha**, by making users verify that they are human before they can submit their comment. Captchas are now found in almost all sectors where human users need to be distinguished from bots. For example, this includes registration forms for e-mail services, newsletters, communities and social networks, as well as online surveys or web services, such as search engine services.

Over time, various methods have been developed to carry out Human Verification. In principle, however, **no established procedure offers 100% protection against spam** and the captcha technology is often associated with decreased user-friendliness.

What type of captchas are there?

The concept of captcha is based on the assumption that, despite the rapid advances in AI research, there are still differences between the mental capacities of a person and those of a computer program. Each captcha therefore needs to present a task that is easy for human users to solve, but not machines.

Captcha-based methods for Human Verification can be roughly divided into text and image-based captchas, audio captchas, mathematical captchas, logic captchas, and gamification captchas.

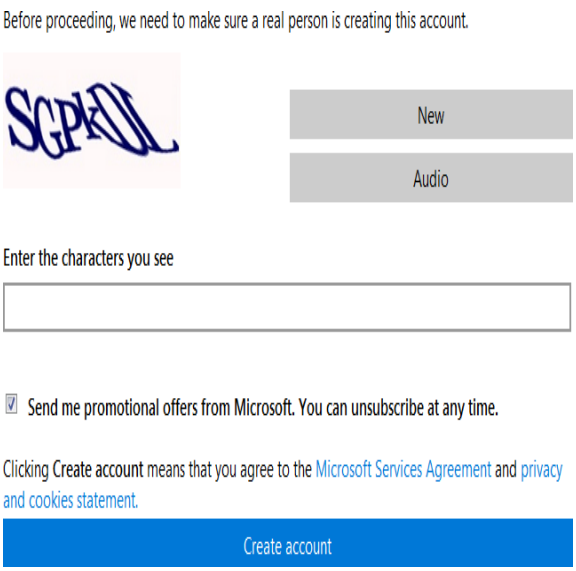
Text-based captchas:

The oldest form of Human Verification is the text-based captcha. Known words or random combinations of letters and digits are alienated. In order to continue, a user has to decipher the code represented in the captcha box and enter the solution into the text box. Classic techniques used to create text-based captchas are Gimpy, ez-Gimpy, Gimpy-r, and Simard's HIP.

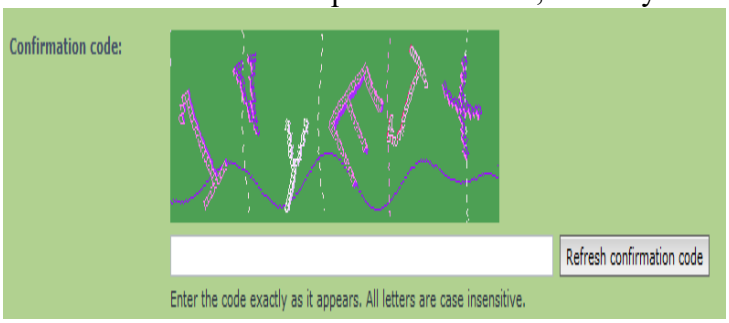
The alienation involves distorting, scaling, rotating, or curving the individual characters and even combining them with additional graphical elements, such as lines, arcs, dots, colors, or background noises. The following graphic shows a selection of possible text-based captchas that can be encountered online.

Text distortion and background noise should make it difficult for recognition systems to read
Text captchas only provide reliable protection against spam when the solution can't be cracked by programs with automatic text recognition. As a rule, however, this requires alienation, which also significantly limits readability for human users.

This can be demonstrated with the following examples. If you want to create a free account with Microsoft, you first have to enter letters in the box, so the user would write 'SGPKDL'. Spambots, on the other hand, wouldn't be able to recognize these contorted letters.



The distorted letters are difficult for spambots to read, but easy for human users



Conclusion: Thus we have studied and verified the generation and verification of CAPTCHA images.

Assignment No.	**3**
Title	Recovery of permanent deleted files and deleted partitions
Roll No.	
Class	B.E. Computer
Date	
Subject	**Cyber Security and Digital Forensics**
Signature	

Assignment No. 3

Title: Recovery of permanent deleted files and deleted partitions

Aim: Write a computer forensic application program for recovering permanent deleted files and deleted partitions.

Objectives: The objective of this lab is to provide expert knowledge on recovering permanent deleted files and deleted partitions.

Hardware/Software requirements: PC

Theory:

What is Digital Evidence?

Digital Evidence is any information that is stored or transmitted in the digital form that a party at court can use at the time of trial. Digital evidence can be Audio files, and voice recordings, Address books and contact lists, Backups to various programs, including backups to mobile devices, Browser history, Cookies, Database, Compressed archives (ZIP, RAR, etc.) including encrypted archives, etc.

Recovering deleted files:

When a file is deleted, the file system removes the file logically. That is, it removes all the meta-data and stamps related to the file. However, the file still resides in the disk as a physical entity until it is overwritten. These physical areas can be very easily explored and read and converted to a readable file using forensic application. It is observed that data resides on a computer for a very long time and are retrieved to a good extent.

Retrieving cached files:

One can find the webpage visited by the suspect or the victim by looking into the cache. The cache file of an application can be spread across in the system storage. We can confine only search by using typical keywords elated to the case or probable websites.

Retrieving files in unallocated space:

In general, a deleted file can be searched sequentially or structurally by looking for file headers or extensions. However, certain tools help us to scan and look for broken headers and use supplementary headers to retrieve data or at least retrieve blocks of a lost file for unallocated space. These retrieved blocks can later be studied and reformed using other tools to retrieve lost files to a great extent. This is also called as file carving.

Meta data of the files can be found from the applications used to create the files however there can be certain tools available to view the metadata of a files like Meta Viewer, Metadata Analysis, iscrub etc. lost file for unallocated space. These retrieved blocks can later be studied and reformed using other tools to retrieve lost files to a great extent. This is also called as file carving

File Carving:

File carving can be used to recover data from a hard disk where the metadata is missing or damaged, especially by professional data recovery companies. When a file is deleted, only the entry in the file system metadata is removed, while the actual data is still on the disk. After a format and even a repartitioning it might be that most of raw data is untouched and can be recovered using file carving. All file systems contain some metadata that describes the actual file system. At a minimum the following is stored: the hierarchy of folders and files, with names for each. For each file is also stored the physical address on the hard disk where the file is stored. As explained below, a file might be scattered in fragments at different physical addresses. File carving is the process of trying to recover files without

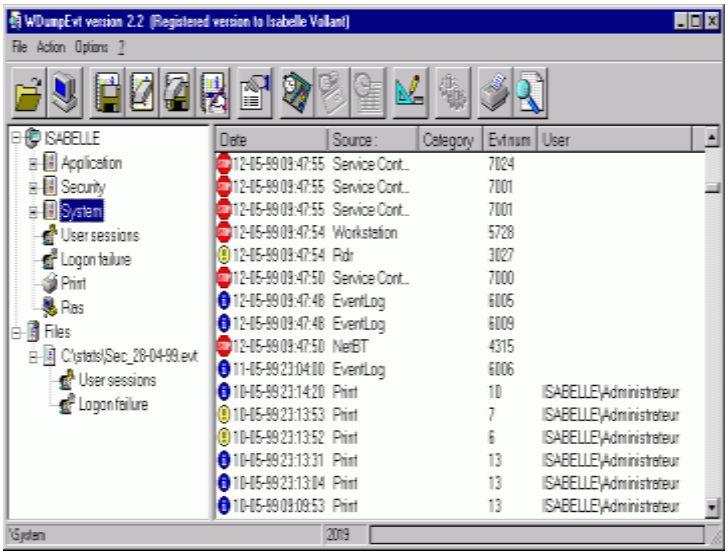
this metadata. This is done by analysing the raw data and identifying what it is (text, executable, png, mp3, etc.). This can be done in different ways, but the simplest is to look for headers. For instance, every Java class file has as its first four bytes the hexadecimal value CA FE BA BE. Some files contain footers as well, making it just as simple to identify the ending of the file.

Most file systems, such as FAT and UNIX Fast File System, work with the concept of clusters of an equal and fixed size. For example, a FAT32 file system might be broken into clusters of 4 KB each. Any file smaller than 4 KB fits into a single cluster, and there is never more than one file in each cluster. Files that take up more than 4 KB are allocated across many clusters. Sometimes these clusters are all contiguous, while other times they are scattered across two or potentially many more so called fragments, with each fragment containing a number of contiguous clusters storing one part of the file's data. Obviously large files are more likely to be fragmented. File carving is a highly complex task, with a potentially huge number of permutations to try. To make this task tractable, carving software typically makes extensive use of models and heuristics. This is necessary not only from a standpoint of execution time, but also for the accuracy of the results. State of the art file carving algorithms use statistical techniques like sequential hypothesis testing for determining fragmentation points.

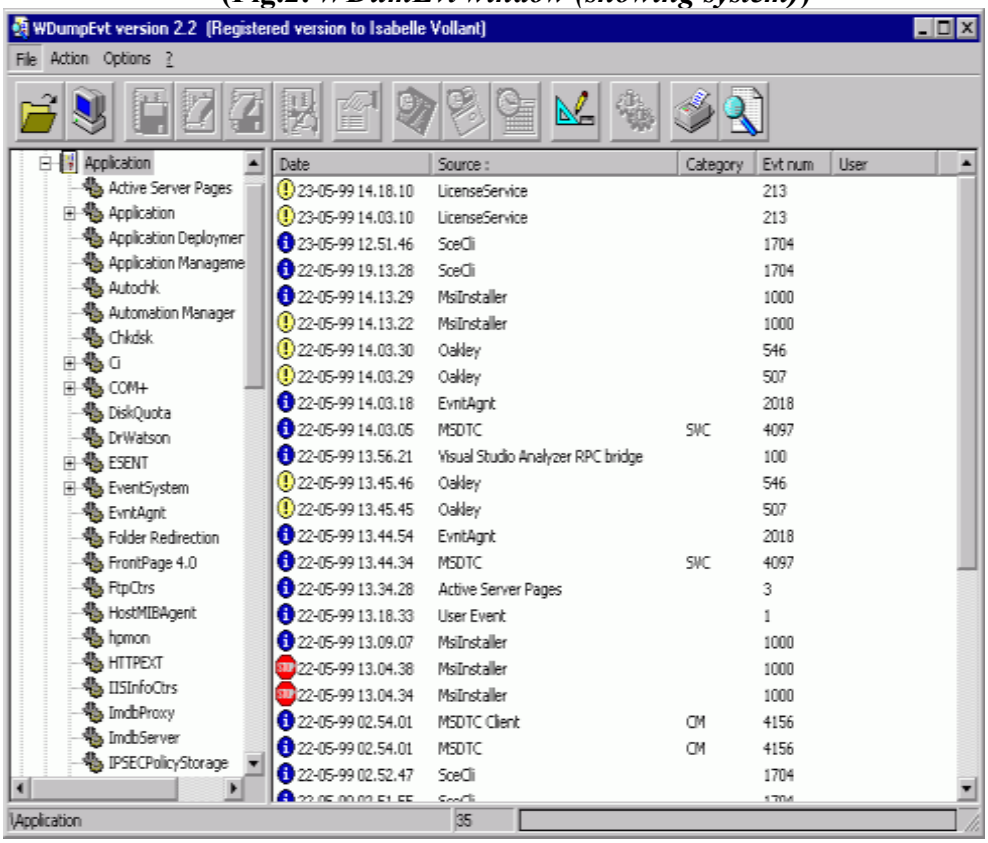
Event logs:

Event logs are stored in Metadata files. The entries in these files can be retrieved on a good way depending upon how refining is carried out by investigators. The victim or suspect system log entries change rapidly as the new events are recorded. The event logs can also be configured minimal to maximum events and durations. We can use tools like Ps log list and EVT to retrieve event records

(Fig.1: PsLoglist output)



(Fig.2: *WDumEvt window (showing system)*)



(Figure 3: WDumEvt window (showing apps))

Conclusion: Thus we have studied the recovery of permanent deleted files and deleted partitions.

Assignment No.	**4**
Title	Log Capturing and event Correlation
Roll No.	
Class	B.E. Computer
Date	
Subject	**Cyber Security and Digital Forensics**
Signature	

Assignment No. 4

Title: Log Capturing and event Correlation.

Aim: Write a program for Log Capturing and event Correlation.

Objectives: The objective of this lab is to provide expert knowledge on Log Capturing and event Correlation.

Hardware/Software requirements: PC

Theory:

What are logs?

Detailed list of an application information, system performance, or user activities. A log can be useful for keeping track of computer use, emergency recovery, and application improvement. } Each software program that is capable of creating a log has different methods of starting or stopping the log creation. } A log is a record of computer activity used for statistical purposes as well as backup and recovery. Log files are written by the operating system or other control program for such purposes as recording incoming dialogs, error and status messages and certain transaction details. Start and stop times of routine jobs may also be recorded.

What is log Capturing?

The logs generated by any system are stored in files which are called as log files. } We can retrieve this information using various system commands which is called 'Log Capturing'. } Linux systems have a very flexible and powerful logging system, which enables you to record almost anything you can imagine and then manipulate the logs to retrieve the information you require. } Linux uses a set of configuration files, directories, programs, commands and daemons to create, store and recycle these log messages.

Need of Log Capturing:

Linux system administrators often need to look at log files for troubleshooting purposes. } Knowing where the system keeps its log files and how to make use of related commands can therefore help save valuable time during troubleshooting. } It helps to monitor the system performance. } From security point of view the events logged in the log file are mainly: 1. Authentication attempts – both successful and failed. 2. All bad requests – which includes attempts for SQL injections and various hacking efforts. } Helps to monitor and check resource allocation and usage.

What is event Correlation?

Event correlation is a technique for making sense of a large number of events and pinpointing the few events that are really important in that mass of information. } Event correlation is performed by a special utility called as event correlator. } Event correlation can be decomposed into four steps: event filtering, event aggregation, event masking and root cause analysis.

Steps in Event Correlation:

- Event filtering:- Event filtering consists in discarding events that are deemed to be irrelevant by the event correlator.
- Event aggregation:- Event aggregation (also known as event de-duplication) consists in merging duplicates of the same event.
- Event masking:- Event masking consists in ignoring events pertaining to systems that are downstream of a failed system.
- Root cause analysis:- It consists in analyzing dependencies between events, based for instance on a model of the environment and dependency graphs, to detect whether some events can be explained by others.

Implementation in our Program:

- In the Log folder in Desktop there is a file ProcessLog which has the process related logs.
- To make sense out of the captured log we sort out the top 10 CPU intensive processes from rest of the large entries of the logs
- These processes may be running in the background or foreground.
- This has been achieved by using the 'sed' command from the coreutils package. 'sed' is a stream editor for filtering and transforming text.
- We have also made a provision to search how many times a particular process has been invoked, the process being provided by the user.

Conclusion: Thus we have studied and implemented the program for Log Capturing and event Correlation.

Assignment No.	**5**
Title	Honeypot
Roll No.	
Class	B.E. Computer
Date	
Subject	**Cyber Security and Digital Forensics**
Signature	

Assignment No. 5

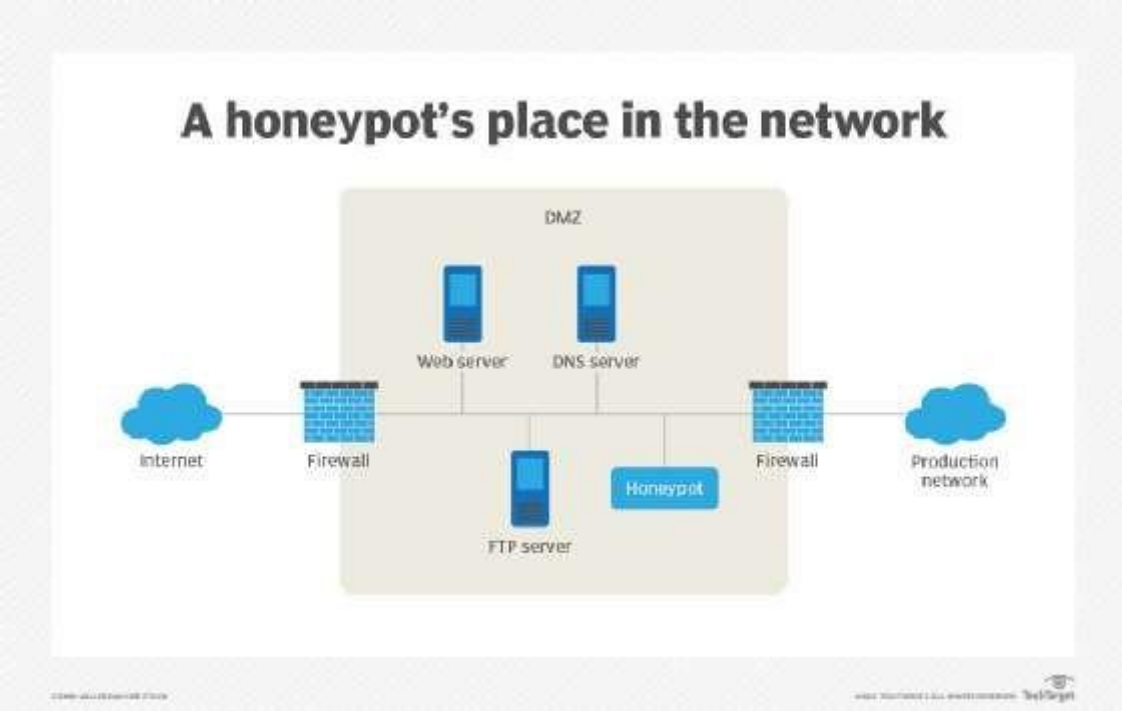
Title: Honey pot.
Aim: Study of honeypot.
Objectives: The objective of this lab is to study of honeypot.
Hardware/Software requirements: PC
Theory:
What is a honeypot?

A honeypot is a network-attached system set up as a decoy to lure cyber attackers and detect, deflect and study hacking attempts to gain unauthorized access to information systems. The function of a honeypot is to represent itself on the internet as a potential target for attackers -- usually, a server or other high-value asset -- and to gather information and notify defenders of any attempts to access the honeypot by unauthorized users.

Honeypot systems often use hardened operating systems (OSes) where extra security measures have been taken to minimize their exposure to threats. They are usually configured so they appear to offer attackers exploitable vulnerabilities. For example, a honeypot system might appear to respond to Server Message Block (SMB) protocol requests used by the WannaCry ransomware attack and represent itself as an enterprise database server storing consumer information.

Large enterprises and companies involved in cybersecurity research are common users of honeypots to identify and defend against attacks from advanced persistent threat (APT) actors. Honeypots are an important tool that large organizations use to mount an active defense against attackers or for cybersecurity researchers who want to learn more about the tools and techniques attackers use.

The cost of maintaining a honeypot can be high, in part because of the specialized skills required to implement and administer a system that *appears* to expose an organization's network resources, while still preventing attackers from gaining access to any production systems.



Honeypots are placed at a point in the network where they appear vulnerable and undefended, but they are actually isolated and monitored.

How do honeypots work?

Generally, a honeypot operation consists of a computer, applications and data that simulate the behavior of a real system that would be attractive to attackers, such as a financial system, internet of things (IoT) devices, or a public utility or transportation network. It appears as part of a network but is actually isolated and closely monitored. Because there is no reason for legitimate users to access a honeypot, any attempts to communicate with it are considered hostile.

Honeypots are often placed in a demilitarized zone (DMZ) on the network. That approach keeps it isolated from the main production network, while still being a part of it. In the DMZ, a honeypot can be monitored from a distance while attackers access it, minimizing the risk of the main network being breached.

Honeypots may also be put outside the external firewall, facing the internet, to detect attempts to enter the internal network. The exact placement of the honeypot varies depending on how elaborate it is, the traffic it aims to attract and how close it is to sensitive resources inside the corporate network. No matter the placement, it will always have some degree of isolation from the production environment.

Viewing and logging activity in the honeypot provides insight into the level and types of threats a network infrastructure faces while distracting attackers from assets of real value. Cybercriminals can hijack honeypots and use them against the organization deploying them. Cybercriminals have also been known to use honeypots to gather intelligence about researchers or organizations, act as decoys and spread misinformation.

Virtual machines (VMs) are often used to host honeypots. That way, if they are compromised by malware, for example, the honeypot can be quickly restored. Two or more honeypots on a network form a honeynet, while a honey farm is a centralized collection of honeypots and analysis tools.

Both open source and commercial offerings are available to help with deploying and administering honeypots. Products include standalone honeypot systems, as well as honeypots packaged with other security software and marketed as deception technology. GitHub has an extensive list of honeypot software that can help beginners get an idea of how honeypots are used.

What are honeypots used for?

Honeypots are used to capture information from unauthorized intruders that are tricked into accessing them because they appear to be a legitimate part of the network. Security teams deploy these traps as part of their network defense strategy. Honeypots are also used to research the behavior of cyber attackers and the ways they interact with networks.

Spam traps are also similar to honeypots. They are email addresses or other network functions set up to attract spam web traffic. Spam traps are used in Project Honey Pot, which is a web-based network of honeypots embedded in website software. Its purpose is to harvest and collect the Internet Protocol (IP) addresses, email addresses and related information on spammers so web administrators can minimize the amount of spam on their sites. The group's findings are used for research as well and by law enforcement to combat unsolicited bulk mailing offenses.

Honeypots aren't always used as a security measure. Anyone can use them for network reconnaissance, including hackers. For instance, a Wi-Fi Pineapple lets users create a Wi-Fi honeypot. Wi-Fi Pineapples are relatively cheap because consumer devices are used to create

a fake Wi-Fi network that mimics a real one in the vicinity. Unsuspecting individuals mistakenly connect to the fake Wi-Fi network, and the honeypot operator can then monitor their traffic. Wi-Fi Pineapples also have legitimate uses, such as for penetration testing (pen testing), where ethical -- or white hat -- hackers are hired to identify vulnerabilities in a network.

Types of honeypots

Based on design and deployment, there are two main types of honeypots: production and research.

1. **Research honeypots** perform close analysis of hacker activity and aim to discover how hackers develop and progress in order to learn how to better protect systems against them. Data placed in a honeypot with unique identifying properties can also help analysts track stolen data and identify connections between different participants in an attack.
2. **Production honeypots** are usually deployed inside production networks alongside production servers; the honeypot acts as a decoy, drawing intruders away from the production network as part of the intrusion detection system (IDS). A production honeypot is designed to appear as a real part of the production network and contains information to attract and occupy hackers to tie up their time and resources. This approach ultimately gives administrators time to assess the threat level and mitigate any vulnerabilities in their actual production systems.

Honeypots can be classified as pure, high-interaction or low-interaction:

1. **Pure honeypots** are full-fledged production systems that monitor a honeypot's link to the network. They are the most complex and difficult to maintain, but they also appear most realistic to attackers, complete with mock confidential files and user information.
2. **High-interaction honeypots** imitate the activities of the production systems, hosting a variety of services and capturing extensive information. The goal of a high-interaction honeypot is to entice an attacker to gain root -- or administrator-level -- access to the server and then monitor the attacker's activity.
3. **Low-interaction honeypots** simulate the most common attack vectors on the network: the ones services attackers frequently request. Therefore, they are less risky and easier to maintain. They do not point malicious users to the root system. The downside of this type

of honeypot is that it is more likely to look fake to an attacker. Low-interaction honeypots are good for detecting attacks from bots and malware. Honeyd is an open source virtual low-interaction honeypot.

Honeypots can be used to mimic several types of networks and technologies. A few examples are the following:

- enterprise databases;
- industrial and other control systems; and
- malware attack vectors and replication vectors, such as Universal Serial Bus (USB) drives.

There are several types of specialized honeypot technologies, such as the following:

- **Malware honeypots.** These are honeypots that mimic malware attack vectors -- places that malware attacks and replicates.
- **Spam honeypots.** These can detect the methods of spammers, monitor their activity and block spam.
- **Database honeypots.** These create decoy databases to mislead attackers using methods that are sometimes missed by firewalls, like Structured Query Language (SQL) injections.
- **Client honeypots.** These actively seek out malicious servers behind client attacks instead of passively waiting for connections. They use virtualization to establish themselves on the server and watch for suspicious modifications to the honeypot.

Benefits and risks of honeypots

Honeypots provide significant benefits, but they also come with disadvantages and risks.

Benefits

- **Real data collection.** Honeypots collect data from actual attacks and other unauthorized activities, providing analysts with a rich source of useful information.

- **Fewer false positives.** Ordinary cybersecurity detection technologies generate alerts that can include a significant volume of false positives, but a honeypot reduces the number of false positives because there is no reason for legitimate users to access the honeypot.
- **Cost-effectiveness.** Honeypots can be good investments because they only interact with malicious activities and do not require high-performance resources to process large volumes of network traffic looking for attacks.
- **Encryption circumvention.** Honeypots capture malicious activity, even if an attacker is using encryption.

Disadvantages

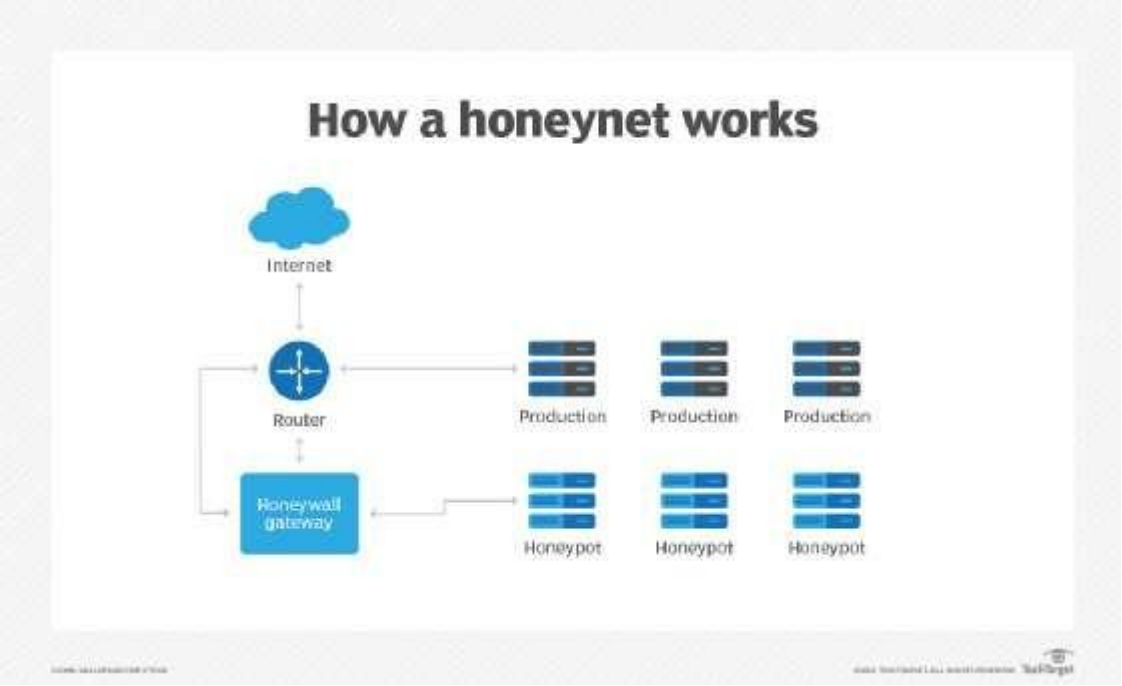
- **Limited data.** Honeypots only collect information when an attack occurs. Zero attempts to access the honeypot means there is no data to analyze.
- **Isolated network.** Malicious traffic that has been captured is only collected when an attack targets the honeypot network; if attackers suspect a network is a honeypot, they will avoid it.
- **Distinguishable.** Honeypots are often distinguishable from legitimate production systems, which means experienced hackers can often differentiate a production system from a honeypot system using system fingerprinting techniques.
- **Put production systems at risk.** Although they are isolated from the real network, they do eventually connect in some way to enable administrators to collect the information they contain. A high-interaction honeypot is generally considered riskier than a low-interaction one because it aims to entice hackers to gain root access.

Overall, honeypots help researchers understand threats in network systems, but production honeypots should not be a replacement for a standard IDS. If a honeypot is not configured correctly, it can be used to gain access to real production systems or as a launchpad for attacks against other target systems.

Honeynet

A honeynet consists of two or more honeypots on a network. Having an interconnected network of honeypots can be useful. It enables organizations to track how an attacker interacts with one resource or network point, and it also monitors how an intruder

moves among points on the network and interacts with multiple points at one time. The goal is to get hackers to believe that they have successfully breached the network, so having more fake network destinations makes the setup more convincing.



A honeynet creates an alternative network to lure in hackers. The honeywall directs intruders toward the honeypot instances where they can be monitored and controlled.

The term *deception technology* has been used to describe the more complex implementations of honeypots and honeynets, often packaged with other technology, such as next-generation firewalls (NGFWs), IDSes and secure web gateways. Deception technology includes automated features that let a honeypot respond in real time to potential attackers.

Cyber threats continue to evolve, and honeypots can help organizations keep up with the ever-changing threat landscape. Even though it's impossible to predict and prevent every attack, honeypots provide useful information to ensure an organization is prepared and are perhaps the best way to catch an attacker in the act. They are a good place for cybersecurity professionals to gather information as well.

Conclusion: Thus we have studied the honeypot.