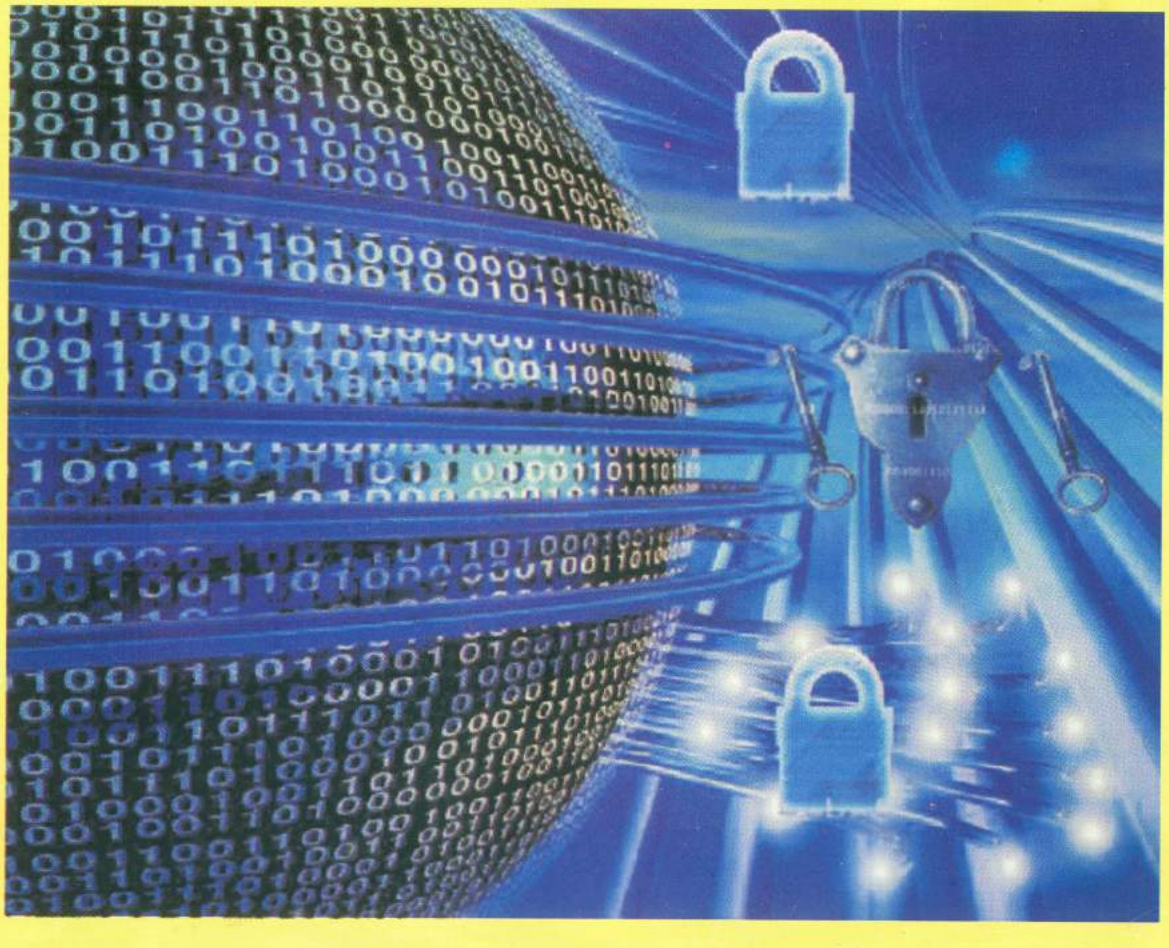
ignou
THE PEOPLE'S
UNIVERSITY

Indira Gandhi National Open University
School of Vocational Education and Training

MSEI-027
Digital Forensics



Digital Forensics: Tools and Techniques

2

"शिक्षा मानव को बन्धनों से मुक्त करती है और आज के युग में तो यह लोकतंत्र की भावना का आधार भी है। जन्म तथा अन्य कारणों से उत्पन्न जाति एवं वर्गगत विषमताओं को दूर करते हुए मनुष्य को इन सबसे ऊपर उठाती है।"

– *इन्दिरा गांधी*

"Education is a liberating force, and in our age it is also a democratising force, cutting across the barriers of caste and class, smoothing out inequalities imposed by birth and other circumstances."

—Indira Gandhi

ignou
THE PEOPLE'S
UNIVERSITY

Indira Gandhi National Open University
School of Vocational Education and Training

MSEI-027
Digital Forensics

Block

2

DIGITAL FORENSICS: TOOLS AND TECHNIQUES

Programme Expert/Design Committee of Post Graduate Diploma in Information Security (PGDIS)

Block Preparation

Unit Writer	Block Editors	Proof Reading
Mr. Sushil K Ocean Technocrats Noida (Unit 1, 2, 3 & 4)	Prof. Ajith Kumar R, Professor Indian Institute of Information Technology and Management-Kerala (IIITM-K) Trivandrum, Kerala Ms. Urshla Kant Assistant Professor, School of Vocational Education & Training, IGNOU	Ms. Urshla Kant Assistant Professor School of Vocational Education & Training IGNOU

Production

Mr. B. Natrajan Dy. Registrar (Pub.) MPDD, IGNOU, New Delhi	Mr. Jitender Sethi Asstt. Registrar (Pub.) MPDD, IGNOU, New Delhi	Mr. Hemant Parida Proof Reader MPDD, IGNOU, New Delhi

Feb, 2012

ISBN-978-81-266-5923-4

Further information about the School of Vocational Education and Training and the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110068. or the website of IGNOU www.ignou.ac.in

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi, by the Registrar, MPDD

Laser typeset by Mctronics Printographics, 27/3 Ward No. 1, Opp. Mother Dairy, Mehrauli, New Delhi-30

Printed at: Berry Art Press A-9, Mayapuri, Phase-I New Delhi-64

BLOCK INTRODUCTION

This block deals with the digital forensics. It is a branch of forensic science encompassing the recovery and investigation of material found in digital devices, often in relation to computer crime. It is the art and science of applying computer science to aid the legal process. It is more than the technological, systematic inspection of the computer system and its contents for evidence or supportive evidence of a civil wrong or a criminal act. Computer forensics requires specialized expertise and tools that goes above and beyond the normal data collection and preservation techniques available to end-users or system support personnel. This block comprises of four units and is designed in the following way:

The **Unit one** deals with the brief overview of digital investigation. Forensic practitioners need to understand and make regular use of the scientific method. The scientific method applied in conjunction with digital forensics methodologies and techniques enables us to adapt to differing circumstances and requirements and to ensure that conclusions reached are solidly based on fact. Familiarity with the limitations of forensic analysis of digital evidence will help investigators in apprehending modern white collar criminals and exculpate the innocent.

The **Unit two** provides an overview of the role evidence plays in a criminal case (particularly in a cybercrime case) and discusses standard procedures for dealing with digital evidence, as well as specific evidence location and examination techniques such as recovering supposedly deleted files, finding steganographic data, locating "forgotten" data and decrypting encrypted data. Procedures for documenting digital evidence are also outlined and examination some of the legal issues involved in evidence collection and handling will also be done.

The **Unit three** covers forensic examination of systems which includes various techniques of search which are group of techniques searches collects information to answer the question whether objects of given type, such as hacking tools or pictures of certain kind, are present in the collected information. According to the level of search automation, this dissertation classifies techniques into manual browsing and automated searches. Automated searches include keyword search, regular expression search, approximate matching search, custom searches and search of modifications. Different scientific methods followed in data search are also explained. It also describes about data recovery and various tools used in data recovery.

Unit four explains forensic examination of network devices which is required to compound the pressures faced by organizations in regards to implementing proper network forensics and log management techniques. Several organization have implemented policies that require organizations keep all network event data. In such situations there is no other choice but to implement expensive archiving equipment and analysis software to monitor and archive network security events. Intrusion detection systems help computer systems prepare for and deal with attacks. They accomplish this goal by collecting information from a variety of system and network sources and then analyzing the information for symptoms of security problems.

Hope you benefit from this block.

ACKNOWLEDGEMENT

The material we have used is purely for educational purposes. Every effort has been made to trace the copyright holders of material reproduced in this book. Should any infringement have occurred, the publishers and editors apologize and will be pleased to make the necessary corrections in future editions of this book.

UNIT 1 DIGITAL INVESTIGATION

Structure

1.0 INTRODUCTION

The information created and stored on computers is a double-edged weapon from a forensic perspective, providing compelling evidence in a wide variety of investigations but also introducing many complexities that can be very tricky to even very experienced practitioners. Digital evidence can be answer to many fundamental questions relating to a crime, including what happened, when it happened, who communicated with whom, who/which is the source of item connected to the crime and who was responsible for the crime.

At the same time, the complexity of computer systems requires appreciation that individual pieces of digital evidence may have multiple interpretations and corroborating information may be vital to reaching a correct conclusion. To make the most of digital evidence, forensic practitioners need to understand and make regular use of the scientific method. The scientific method applied in conjunction with digital forensics methodologies and techniques enables us to adapt to differing circumstances and requirements and to ensure that conclusions reached are solidly based on fact. Familiarity with the limitations of forensic analysis of digital evidence will help investigators in apprehending modern white collar criminals and exculpate the innocent.

1.1 OBJECTIVES

After studying this unit, you should be able to:

- understand various aspects of Digital Investigation;
- explain different scientific methods followed in investigation; and
- elucidate various tools used in digital forensics.

1.2 WHY INVESTIGATE?

First we will need to consider the complaint or the initial reason for conducting an investigation.

Some typical reasons that may warrant an investigation are the following:

- internet usage exceeds norm;
- using e-mail inappropriately;
- use of Internet, e-mail or PC in a non-work-related manner;
- theft of information;
- violation of security policies or procedures;
- intellectual property infractions; and
- electronic tampering.

1.2.1 Internet Usage Exceeds Norms

If the complaint is that someone's Internet usage is too high, we should first determine the basis for this complaint. It should also be determined whether the above normal Internet usage was identified through electronic monitoring or by personal observation. It is also appropriate to determine if the usage is out-of-line with organization standards for the type of job responsibilities held by the individual under investigation. Equally important is to determine how those standards were determined and developed.

There are different questions to be asked and answered, in order to investigate the claim, depending on the basis of the complaint.

If the usage was electronically monitored:

1) Did a firewall monitor the usage?

2) Was the usage monitored by Internet Protocol (IP) address or individual identification (ID)?

3) What exactly was monitored? (e.g. time, sites, keywords, etc.)

4) Can more than one person usc :his personal computer (PC) (or IP address)?

5) Can more than one person use this ID?

6) Can the usage times/dates be correlated to physical access by the individual under investigation? (If monitoring shows access was between 8 a.m. and 10 a.m., was the individual at work during this time?)

7) What was the pattern of access?

8) How does this compare with the individual's work schedule?

9) Could the individual have logged in and then not logged out? (i.e. get to an Internet site and then go to another task on the PC, thus leaving the Internet site up and running?)

10) Are there timeouts set on the Internet access? On the PC login?

11) Are there security cameras, login sheets, key card access logs or timecards that can verify that it was the individual who accessed the Internet via this PC?

12) Is there a pattern to the usage?

Once you obtain answers to these questions you will begin to see the outlines of a plan of the investigation forming. For example, if "X" is accused of exceeding Internet norms, based on a report generated from the firewall monitoring system, we can ask some additional questions to validate the concern/ complaint.

If the pattern of unusually high utilization was after-hours when "X" was not scheduled to be at work, then there might be a deeper issue that will require further investigating to uncover (i.e., who and how someone was using "X's" ID after-hours). However, if the case is simply that "X" is logging into the Internet first thing in the morning to check the latest news or stock quotes and not logging out, this is a case where the monitoring or rules might need to be adjusted to account for the high usage. Alternatively, "X" may simply need a refresher course on the organization's Internet usage policies.

On the other hand, if the usage concern was based on a person's observation of "X's" actions, there is another slightly different set of questions to ask such as:

1) Who made the observation?

2) Are logs available to support the observation? (e.g. login, logout, timecards, firewall access etc.)

3) Are there other witnesses to support the observation?

4) What exactly was the individual under investigation observed doing?

5) What is the pattern of usage?

6) Are there security cameras, login sheets, time cards or key card access logs that can verify the individual under investigation had access and was logged on to the Internet?

Again, once you obtain answers to these questions you will begin to formalize a plan of investigation. This plan will differ slightly from the plan based on electronic monitoring. With observation being the basis for a complaint, the ability to verify the usage is more difficult to substantiate – but not impossible.

There are a variety of tools, methods and techniques outlined in this text that will allow you to substantiate the claim, if there is any evidence. For example, there are several files located on the firewall and the PC that can be retrieved, displayed and reviewed in order to prove or disprove the above-normal access violation(s).

The above-normal utilization should prompt the investigator and management to inquire about the impact (financial, physical, operational etc.) of the so-called excessive usage. Several questions to help evaluate the impact include:

1) What damage (if any) did the excessive usage cause?

2) How can the damage be substantiated?

3) How can the damage be quantified?

4) Did the individual under investigation not meet his or her job responsibilities as a result of excessive Internet usage?

5) Did the individual under investigation interfere with another person's job performance as a result of the excessive utilization?

6) Was someone offended by the usage (e.g. inappropriate materials, games being played)?

7) Can you identify this person?

8) Is the person willing to state for the record that he or she was offended by the usage?

9) Did fraud occur in the form of falsified timesheets – hours of work reported or any other form, as a result?

10) The answers to these questions answers will not only help form the plan for this type of investigation, but will also help the investigator and management determine if the investigation should be (can be) pursued.

1.2.2 Inappropriate E-mail

Before performing any investigation on e-mail, you need to ensure that corporate policy allows it. IT Act 2000 protects the privacy of electronic communications. If corporate policy specifically states that all computers and data stored on them belong to the organization, then you are probably on safe ground. Be sure that there is such a policy and that the employee under investigation has read the policy before proceeding. Although this is one of the easiest investigations, this type of investigation should be done strictly by the book. If the corporate policy does not contain the rights to the employee's e-mail, then you and the organization could be subject to a lawsuit for invading the privacy of an employee. If the reason for an investigation is that there was inappropriate use of e-mail, either through the act of sending offensive material or for personal and non-work-related use, there is yet another set of questions that should be asked. These questions will help determine if there was inappropriate utilization of the organization's e-mail systems and if further investigative action is required.

1) What was sent?

2) Can you obtain a copy from the complainant or recipient?

3) Is a copy available from the automated e-mail archive system?

4) Was someone offended? (This could be a harassment issue and require HR involvement.)

5) Who if anyone else received the material?

6) Was the individual under investigation the originator of the e-mail or was it someone else?

7) How were you able to (or can you) validate this?

8) Could someone else have sent the e-mail, using the ID of the individual under investigation?

9) Are screen-saver passwords used?

10) Could someone else use the PC of the individual under investigation?

11) Was the time that the e-mail was sent during the time the individual under investigation had access to e-mail?

12) Is auto-forwarding of e-mail used? Available? Activated?

13) Was a group list used?

14) Are there patterns or history to the e-mail usage?

15) Have there been previous warnings to the individual under investigation about the e-mail usage?

16) If so, are these warnings documented?

17) What was the intent of the e-mail?

Some of the questions listed in the section on abnormal Internet utilization can also be applied to this type of investigation. The real issue with this type of investigation is to determine whether it is an issue of harassment or a case of violating organization e-mail policies/procedures.

Potential exposures to the organization, which can result from the lack of a proactive response by management to a harassment complaint, include a lawsuit filed against the organization by the complainant, as well as multiple instances of harassment that can lead to multiple lawsuits.

Furthermore, to make matters worse, the longer the organization waits to investigate, the more likely it is that lawyers will have a field day and turn this into the organization not caring and thus higher rewards to the complainant. To alleviate the appearance of a non-proactive response to harassment complaints, the organization should have anti-harassment policies and training programs. This training should be repeated annually for all employees. There should be documentation that is maintained in HR files stating that each employee has attended and signed a statement that he or she has read the organization's policies against harassment. This is also documentation that should be gathered during the investigation.

Non-Work-Related Usage of Organization Resources

If the reason for the investigation is about non-work-related use of organization resources (i.e. PC, e-mail or access to the Internet), the above questions apply, but there are additional questions that should be asked including:

1) What exactly occurred? (Was the individual under investigation using his or her PC to engage in "moonlighting" work, e-mail for personal use etc.?)

2) When did the incident occur?

3) How was it documented?

4) How often or how much does this happen?

5) Is the individual under investigation the only people engaged in this activity or are there others?

6) How can you determine this?

7) Is the action a widely accepted organization practice, albeit a violation of organization policy?

8) Did the individual under investigation take the action for personal financial gain?

9) Was the non-work-related usage for personal use?

10) Is there a liability to the organization due to the unauthorized use of organization property?

These more detailed questions will help frame the direction of the investigation more clearly. Thus, a more appropriate plan of action can be devised and carried out. The main issue with this type of investigation concerns the inappropriate use of organization property for personal gain and whether the inappropriate usage violated any standing organization policies.

1.2.3 Theft of Information

The theft of information raises the intensity and seriousness of an investigation to levels that may exceed those established in previously discussed scenarios. The intensity of an investigation into the theft of information will vary, depending on what type of information was stolen, its significance to the organization's ability to remain competitive, the nature and sensitivity of the information stolen and what was done with the stolen information.

Some of the previously mentioned questions can be applied to this type of investigation. However, there are additional questions that relate specifically to the theft of information including:

1) What type of information was stolen?

2) How has this been (or can this be) verified?

3) How much information was stolen?

4) How was the information stolen?

5) What is the impact or cost of the loss?

6) How can this loss be quantified?

7) How can this be substantiated?

8) Is the cost of the loss tangible or intangible (competitive information can be intangible)?

9) Has the goodwill of the organization been damaged as a result of the theft?

10) Has the organization lost a competitive edge due to the theft?

11) Was the information totally lost (e.g. copied and then erased or destroyed) or was it copied?

12) What was the level of security surrounding the information lost?

13) Who had access to the stolen information?

14) Can this be verified?

15) Are access logs available?

16) Are they free from potential, external tampering?

17) Were there procedures in place for the safe handling/accessing of the lost information?

18) Was the information proprietary, confidential or restricted?

19) How was this classification determined and communicated?

To determine exactly how the information was stolen, you might need to perform further security and access audits/reviews. For the purpose of planning and investigation, the investigator should develop a sense of how the information was stolen. One reason to quickly determine how the information may have been stolen is an attempt to prevent further information from being stolen in the same manner.

1.2.4 Violation of Security Parameters

Violation of security parameters can vary widely from an individual simply failing to properly log off when leaving work to covert hacking into secured files. Security parameters are not always those dramatic measures of using guards, secret codes, retinal scanners and IDs, but they do include the use of security cameras and passwords and following procedures for handling secure documents.

The violation or misuse of security parameters can lead to the theft or misuse of organization information or property or worse. Violation of security parameter complaints should begin with asking the following questions:

1) What security parameters or measures were violated?

   Note: Care must be exercised in both asking and documenting the response to this question. Some parameters may be proprietary while others may be highly sensitive and their disclosure might jeopardize the security of entire systems.

2) How were the parameters violated? (See note above.)

3) What was the result of the violation?

4) How can this be substantiated?

5) Were passwords compromised (hacked)?

6) Have new passwords been issued? Reset?

7) Were security measures disabled (e.g. security cameras unplugged, screen savers turned off etc.)?

8) Were security measures bypassed? If so, how? (See note in question 1 above.)

9) Was information falsified as part of the violation (e.g. fraud - pretending to be someone else)?

The violation of security parameters does not always result in the compromise of organization information. However, because the violation of security can lead to the compromise of information, it is important to investigate every violation.

The investigation can lead management to recognize the need to add more security measures or to improve existing measures to both secure and protect the organization's information.

1.2.5 Intellectual Property Infraction

Intellectual properties are those ideas, techniques, procedures or program codes that are considered proprietary and that belong to a specific organization. Companies usually have clauses in their employment contracts that state that any intellectual property developed during an employee's employment with the organization belongs to the organization and cannot be used outside the organization. Infractions of an organization's intellectual property policies usually involve former employees, contractors or consultants, using techniques or code that they created (or had access to) who are now at a new employer/competitor.

When investigating this type of infraction, the investigator may wish to begin by asking the following questions:

1) Does the organization require employees involved in or holding specific job responsibilities to sign an intellectual properties agreement/contract?

2) Are signed policies on file?

3) Does the organization have a viable intellectual properties policy?

4) Is it in force?

5) How can this be verified?

6) When the intellectual property in question was first created for the organization?

7) How can this be substantiated?

8) Who developed the intellectual property?

9) How can this be verified?

10) When was the intellectual property created or used outside the organization, violating the organization's (and previous employee's) intellectual property agreement?

11) Can this be substantiated?

12) Who is the original owner of the intellectual property?

13) Is this merely a case of plagiarism?

14) Are there copyrights involved?

15) Are there patents involved?

16) What proof is there that the intellectual property in question belongs to the organization?

With most intellectual property infractions, it is advisable to seek legal counsel in helping to design and plan the investigation. The major concern of management is the impact of the infraction. If the impact is minimal, management may decide that an investigation is not warranted. If, however, the infraction might place the organization at a competitive disadvantage, management may wish to proceed with the investigation.

Competent legal counsel may advise that any and all violations of a organization's intellectual properties policies be investigated and prosecuted to the fullest extent of the law. Failure to do so (or even to conduct an investigation) might be construed by the courts as indifference and thus weaken the organization's ability to prosecute future cases.

1.2.6 Electronic Tampering

Electronic tampering can involve fraud, mimicking someone or something (i.e. IP spoofing), masking or masquerading as someone (i.e. social engineering). The intent and result of the tampering is the primary reason to conduct an investigation.

Even if the intent of the tampering involves or can be linked to a non-competitive prank, there is still reason to investigate. If any tampering can occur, regardless of the reason, then it should be prevented to protect the organization's information assets.

When investigating electronic tampering, the following questions provide the investigator with a good starting point. Additionally, the questions listed in the section that addressed the "Violation of Security Parameters" should also be incorporated into the investigation plan.

1) What was tampered with?

2) How can this be verified?

3) Did the tampering result in the perpetration of a fraud?

4) What was the intent of the tampering?

5) How can this be verified?

6) How the tampering was carried out?

7) Who first noticed the tampering?

8) How was the tampering first identified?

9) Could the tampering have been undertaken in more than one way?

Because some forms of tampering can involve theft or fraud, which can be criminal offenses, legal counsel should be involved in planning this type of investigation.

We have reviewed some of the questions an investigator can ask and gathered some preliminary information. We now need to review the basis upon which the complaints were formulated, such as a violation of organization policies, procedures or legal statutes.

1.2.7 Establishing a Basis or Justification to Investigate

If there is a justification for a specific complaint or reason to investigate, there should also be rules or a baseline for which the complaint was filed, such as violating a standing organization policy or procedure. For example, if organization policies and procedures state that employees should only use e-mail for organization business, this would be the baseline for a complaint about an individual suspected of using the organization's e-mail system for non-work-related activities.

Baselines that guide many complaints (or a justification to investigate) often include a misuse or violation of:

- Organization policies and procedures;
- Legal statutes;
- Mandatory statutes; and
- Regulatory statutes.

The investigator will need to consult these baselines as appropriate and as part of the investigation to determine how the baseline(s) apply and if there are any documented penalties for violation of these baselines. For example, a policy and its associated penalty for sending inappropriate e-mail could result in the loss of employment for the individual found guilty of violating this policy.

First, consult the organization's policies and procedures. There are several different policies and procedures within the organization that should initially be reviewed. For example, Human Resources, Security and Employee policies are a good beginning and represent general, most often found standard policy types in force within most organizations.

As part of investigation planning, consider asking the following questions to learn more about the organization's current policies and procedures (these are especially relevant to investigation of harassment charges).

1) Are the policies and procedures published and available?

2) Are the policies and procedures current?

3) Are the policies and procedures available in hard or soft copy?

4) Is there mandatory training or orientation to acquaint all employees with these policies and procedures?

5) Are signatures of the employees gathered to verify that all employees have received and reviewed these policies?

6) Have there been audits or reviews to verify compliance with policies and procedures?

7) There are several additional items of documentation that should be considered for examination, depending on the type of investigation to be undertaken. The investigator should be prepared to ask for and examine:

   - *Contracts*, both with third-party suppliers and with external consultants;

   - *Non-disclosure agreements*, with third-party suppliers and with consultants.

In addition to organization policies, procedures and contracts, the investigator may find it appropriate to consult legal statutes if criminal activities are involved. Mandatory statutes refer to those contracts that the organization has with other companies or entities. Those contracts will also need to be examined in cases of loss of information (data) and the subsequent impact on the organization. Regulatory statutes will need to be examined when the investigation reveals a potential for the loss or disclosure of confidential information (data).

Another reason for consulting th: legal, mandatory and regulatory statutes is to determine the liabilities to the organization if no investigation is conducted, when a breach of organization security and or policy occurs. For example, if a fraud was to be perpetrated and the organization knew about it but does not pursue or report it, the organizati on could be held liable for damages resulting from the fraud.

The organization could also face penalties for not reporting the fraud.

1.2.8 Determine the Impact of Incident

Once both the reasons for an investigation and the baseline have been determined, we must now determine the impact of the incident. By understanding the impact, we can determine if it is feasible to continue on with the investigation. By their very nature, some incidents, regardless of the impact (Financial or otherwise) will need to be investigated.

Some items to keep in mind when determining the impact include, but are not limited to:

- Benefits to pursue such an investigation;

- Liabilities for not pursuing an investigation;

- Obligations to pursue or not to pursue (goodwill toward public, partners and other contracts); and

- Resources available (time, people, finances, tools etc.).

To perform an impact review, you will need to plan your investigation. First, you should review why you are conducting the investigation. For illustration, let us assume that the complaint was that someone was sending inappropriate e-mail. For planning sake, you know you will need to talk to several people; namely, the person who filed the complaint, the person under investigation, network administrator, supervisors and a representative from human resources (HR), as well as the legal department. So you will need to allot time to speak with each of these individuals. You will need to gather their statements as well as evidence along the way. While speaking with HR and with the supervisors, you will need to gather the organization policies and procedures on e-mail harassment and any training programs specifically aimed at workplace harassment. You might even need to talk with the auditors to see what kinds of audits they have conducted in the HR department and networking areas to date (reference audit checklists, list of questions to ask, samples of HR policies).

The second step now that you have planned interviews and gathered the appropriate documentation is to draw up a timeline. You should also consider what resources will be available to you in conducting this investigation. Those resources can include personnel in your group and those external to your department/group. In your consideration, you need to know if you have the authority to ask other departments to help you gather information. Having someone gather information for you will save you time and effort. This may, however, compromise the independence of your investigation. Be sure to verify your sources and check with the legal department on the appropriateness of using external third parties for this type of specialized work.

You will also need to consider what tools you can use to gather information. There are various tools aimed at providing the auditor with access to erased or "hidden" data that will be used to investigate.

As an example and as a means of further discussing the investigation model, we will operate under the assumption that your organization has (or you have as the auditor/security professional) several of the tools discussed in this text. We will use these tools to help gather evidence that the individual under investigation has sent an offensive e-mail. Because this example takes a look at e-mail as the source of policy infraction (or worse), we might not need to confiscate any equipment for examination. This is due to the fact that e-mail (for most organizations) runs on network servers and network or operations personnel maintain these servers. You must also keep in mind an important question: How do you determine that the e-mail in question (the one cited as being offensive) was not "planted"? A basic first step for reviewing e-mail includes gathering all the e-mails sent by the individual under investigation (or the e-mails sent during a certain time period).

You might need to trace back several days or weeks, looking at e-mail details to see if there is more than the one e-mail that may have started the complaint. Also, you might need to gather e-mail from and to other persons involved in the incident (if others were involved or affected). You can determine whose e-mail you will need to examine by reviewing a copy of the original e-mail that initiated the complaint. There may be a list of who else was copied on the e-mail. This list is where you start gathering e-mails. How do you gather the e-mails? Ask the network/ operations personnel to recover the needed e-mails and have them copied to a file you can access. Most e-mails can be copied to MS Word files in text format.

This step takes a critical "leap of faith" on the part of the auditor/security professional: that the organization does indeed maintain an e-mail log/file and also archives all e-mail traffic.

Additionally, the issue of independence is raised here once again. If someone other than the auditor/security professional retrieves these e-mails it is imperative that control be maintained over the retrieval process to ensure accountability and authenticity of the e-mails retrieved from the log file.

Now that you have your list of whom to interview, who will help and how you plan to gather the evidence, you should be able to put together a timeline with an estimate of hours required for this part of the investigation. The second half of the investigation previously discussed will depend on what you find. This is the ambiguous part, the part where you have to hypothesize, theorize and even guess for planning purposes.

Once again, let us assume the worst-case scenario; you have found one e-mail that contains inappropriate material. Also assume there are several people involved that one individual initially sent the inappropriate e-mail and several additional individuals passed (e-mailed/forwarded) it around. You will need to review the policies and procedures on the distribution of inappropriate e-mail (if they exist). You will also need to review the personnel files of the people involved (and everyone who forwarded the e-mail(s) in question) and determine if they have been through any organization-sponsored harassment training.

The evidence you gatherèd (for this scenario) will most probably be turned over to HR department personnel so they can follow through on any disciplinary actions (if warranted) or establish new policies, procedures and controls. Do not forget to plan time to document and summarize your findings. In the worst case, your evidence may be used in a legal case and not only how you carry out your investigation but also what you document and how you document evidence could be critical.

A major consideration in your documentation efforts is to record who you talked to, when you talked to them, what they said, what evidence you gathered and how you gathered that evidence; and then to draw your conclusions, all without interjecting personal opinions. Just gather and document the facts.

1.2.9 Whom to Call/Contact

Whom to call/contact depends on the type investigation that will be conducted (e.g. fraud, misuse of company assets, etc.). In most companies, the department with the most experience in conducting investigations is the internal audit department (although in some organizations an independent fraud investigation unit may be in existence and can be called upon for assistance). This might be the first contact for help with any investigation. In most cases, you will probably need to contact:

- **Internal audit:** expertise in conducting investigations, past audit results;
- **Network/operations:** for tracking IP addresses, e-mails, log files, backup files, monitoring logs, incident reports;
- **Data security:** policies and procedures, password usage, security reports, log files, access requests and reports;
- **Physical security:** policies and procedures, after-hours logs to work areas, security camera tapes, key card access logs, access requests, incident reports;
- **Human resources:** policies and procedures, employee information, complaint reports;
- **Legal:** contracts, legal assistance, mandatory, statutory, legal requirements;
- **External consultants:** in area of needed expertise. And only after serious reflection and consultation with your legal counsel; and
- External law enforcement personnel, agencies or departments.

1.2.1o If You Are the Auditor/Investigator

As the auditor/investigator, you will need to determine the following to conduct an effective investigation:

1) Available resources;

2) Authority;

3) Obligations/goals;

4) Reporting hierarchy;

5) Escalation procedures;

6) Time frame;

7) Procedures to follow;

8) Precedence of past investigations; and

9) Independence.

Each of these items is discussed as to its importance in conducting an effective investigation.

- **Resources**

You need to know if you are working alone or with a team. A team might be faster but not warranted for small investigations. With a team you will also need to define who the leader is and who is responsible for each aspect of the investigation.

- **Authority**

You will need to know what your level of authority is in conducting the investigation. Do you have access to the information, areas and resources (not just personnel) you need to effectively conduct your investigation? Do you have to file requests to gain access? Also, do you have the authority to quarantine files and equipment? You might need to take someone's PC into your possession for investigation so you will need to know if you can take it without impacting ongoing operations. You might also need to know if you have the right to request certain people's time in order to interview them. If the person you wish to interview is an hourly employee, you might need prior permission from his or her supervisor.

- **Obligations/Goals**

You need to know what your obligation is both ethically and professionally in this company and with this investigation. If you find that your superior is the individual under investigation, you will need to know how to handle this investigation. You will need to consider what to do if you uncover some illegal activity/ action that involves the company, something that if made public could damage the company's reputation but if not reported could hurt someone else (as well as being illegal and misrepresentative). Hopefully, you will not find yourself in an Erin Brokovich "situation."

- **Reporting Hierarchy**

You will need to identify to whom you must report your findings to and how to report them. This reporting hierarchy is important in obtaining the go/no-go decision to conduct the investigation.

Also, you might need to know who to ask for help in getting the cooperation needed to conduct your investigation.

- **Escalation Procedures**

If you have problems obtaining cooperation or in reporting findings of an urgent matter, you will need to know the escalation procedures for your investigation.

1.3 . UNDERSTANDING DIGITAL FORENSICS

Digital Forensics includes preserving, collecting, confirming, identifying, analyzing, recording and presenting crime scene information.

A **digital investigation** is a process where we develop and test hypotheses that answer questions about digital events. This is done using the scientific method where we develop a hypothesis using evidence that we find and then test the hypothesis by looking for additional evidence that shows the hypothesis is impossible. *Digital evidence* is a digital object that contains reliable information that supports or refutes a hypothesis.

Another way to define **digital investigation** could be that it is a process to answer questions about digital states and events. The basic digital investigation process frequently occurs by all computer users when they, for example, search for a file on their computer. They are trying to answer the question "what is the full address of the file named important.doc?" In general, digital investigations may try to answer questions such as "does file X exist?", "was program Y run?" or "was the user Z account compromised?"

A **digital forensic investigation** is a special case of a digital investigation where the procedures and techniques that are used will allow the results to be entered into a court of law. For example, an investigation may be started to answer a question about whether or not contraband digital images exist on a computer. An average Microsoft Windows user may be able to answer this question by booting the computer and using the Find Files function but these results may not be court admissible because steps were not taken to preserve the state of the computer or use trusted tools.

A digital forensic investigation is a process that uses science and technology to analyze digital objects and that develops and tests theories, which can be entered into a court of law to answer questions about events that occurred. In other words, a digital forensic investigation is a more restricted form of digital investigation.

The digital investigation process involves formulating and testing hypotheses about the state of a computer. Hypotheses are to be formulated because digital events and digital states cannot be observed directly and the facts are not known. Tools are used to observe the state of digital data, which makes them indirect observations. This is similar to being told about something instead of seeing it by you. With digital investigations, the confidence is based on the trust of the hardware and software used to collect and analyze the data. The methods used to formulate and test the hypotheses can make the investigation process a scientific one.

Digital evidence is data that supports or refutes a hypothesis that was formulated during the investigation. This is a general notion of evidence and may include data that may not be court admissible because it was not properly or legally acquired.

The forensic analysis process involves taking factual observations from available evidence, forming and testing possible explanations for what caused the evidence and ultimately developing deeper understanding of a particular item of evidence or the crime as a whole. Put another way, elements of digital forensic analysis include separating particular items for individual study, determining their significance and considering how they relate to the entire corpus of evidence. This process often involves experimentation and research and may lead to additional information that must be synthesized into the overall process. For instance, analysis can suggest additional keywords that forensic practitioners use to find additional information that adds to the analysis process. As such, the process is cyclic requiring multiple passes through the hypothesis formation and testing phases until a solid conclusion is reached.

In the simple case of an incriminating file on a hard drive, analysis of the contents and metadata of the file can reveal how the file came to be on the hard drive and can uncover distinctive characteristics to search all available media for related artifacts and file fragments. As another example, forensic analysis of SMS messages on a murder victim's mobile device may lead digital investigators to a prime suspect. Then, by obtaining usage details for the suspect's cell phone and analyzing the timing, location and content of both the victim's and suspect's mobile devices immediately prior to the offense, digital investigators can place the suspect at the crime scene.

More generally, forensic analysis involves objectively and critically assessing digital evidence to gain an understanding of and reach conclusions about the crime. This process can involve evaluating the source of digital objects, exploring unfamiliar file formats to extract usable information, developing timelines to identify sequences and patterns in time of events, performing functional analysis to ascertain what was possible and impossible and relational analysis to determine the relationships and interaction between components of a crime.

In essence, forensic analysts attempt to answer the fundamental questions in an investigation of what happened, where, when, how, who was involved and why. In addition, forensic analysts may be directed to address specific questions relevant to the investigation or to develop a list of other potential sources of evidence like e-mail accounts and removable storage media.

Check Your Progress 1

Note: a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of this Unit.

What is Electronic Tempering?

....

....

....

....

....

1.4 APPLYING SCIENTIFIC METHODS TO DIGITAL FORENSICS

Forensic analysis of digital evidence depends on the case context and largely relies on the knowlèdge, experience, expertise, thoroughness and in some cases the curiosity of the practitioner performing the work. Although every forensic analysis will have differing aspects based on the dataset, objectives, resources and other factors, the underlying process remains fundamentally the same.

1) Gather information and make observations

This phase is sometimes referred to as forensic examination and involves verifying the integrity and authenticity of the evidence performing a survey of all evidence to determine how to proceed most effectively and doing some pre-processing to salvage deleted data, handle special files, filter out irrelevant data and extract embedded metadata. This phase may include keyword searching to focus on certain items and a preliminary review of system configuration and usage. This phase needs not be limited to digital evidence and can be augmented by interviews, witness statements and other materials or intelligence.

2) Form a hypothesis to explain observations

While forensic practitioners are gathering information about the crime under investigation, we develop possible explanations for what we are seeing in the digital evidence. Although such conjecture is often influenced by the knowledge and experience of a forensic practitioner, we must guard against preconceived notions that are based on personal prejudice rather than facts.

3) **Evaluate the hypothesis**

Various predictions will flow naturally from any hypothesis (if the hypothesis is true, then we would expect to find X in the evidence) and it is our job as forensic practitioners to determine whether such expectations are borne out by the evidence. The success of a forensic analysis hinges on how thoroughly an initial hypothesis is attacked. Therefore, it is crucial to consider other plausible explanations and include tests that attempt to disprove the hypothesis (if the hypothesis is false, then we would expect to find Y). If experiments and observations do not support the initial hypothesis, we revise our hypothesis and perform further tests.

4) **Draw conclusions and communicate findings**

Once a likely explanation of events relating to a crime has been established, forensic practitioners must convey their work to decision makers. Observe that the scientific method is cyclic, potentially requiring forensic analysts to repeat these steps until a conclusion can be made. If experiments disprove the initial hypothesis, a new one must be formed and evaluated. Even when some experiments support the hypothesis, new information often emerges that must be considered and tested to determine whether the hypothesis still holds.

1.4.1 Digital Investigation and Evidence

The focus of a digital investigation is going to be some type of digital device that has been involved in an incident or crime. The digital device was either used to commit a physical crime or it executed a digital event that violated a policy or law. An example of the first case is if a suspect used the Internet to conduct research about a physical crime. Examples of the latter case are when an attacker gains unauthorized access to a computer, a user downloads contraband material or a user sends a threatening e-mail. When the violation is detected, an investigation is started to answer questions such as why the violation occurred and who or what caused it to occur.

Consider a server that has been compromised. We start an investigation to determine how it occurred and who did it. During the investigation, we find data that were created by events related to the incident. We recover deleted log entries from the server, find attack tools and find numerous vulnerabilities that existed on the server. Using this data and more, we develop hypotheses about which vulnerability the attacker used to gain access and what she did afterwards. Later, we examine the firewall configuration and logs and determine that some of the scenarios in our hypotheses are impossible because that type of network traffic could not have existed and we do not find the necessary log entries. Therefore, we have found evidence that refutes one or more hypotheses.

The term evidence is used in the investigative context. Evidence has both legal and investigative uses. The definition that we defined earlier was for the investigative uses of evidence and there could be situations where not all of it can be entered into a court of law.

Because the legal admissibility requirements vary by country and state and because one may not have a legal background, let us focus on the general concept of evidence and you can make the adjustments needed within the jurisdiction. In fact, there are no legal requirements that are specific to file systems.

So far, you may have noticed that We have not used the term "forensic" during the discussion about a digital investigation. The American Heritage Dictionary defines forensic as an adjective and "relating to the use of science or technology in the investigation and establishment of facts or evidence in a court of law". The nature of digital evidence requires us to use technology during an investigation, so the main difference between a digital investigation and a digital forensic investigation is the introduction of legal requirements.

1.4.2 Digital Crime Scene Investigation Process

There is no single way to conduct an investigation. If you ask five people to find the person who drank the last cup of coffee without starting a new pot, you will probably see five different approaches. One person may dust the pot for fingerprints, another may ask for security camera tapes of the break room and another may look for the person with the hottest cup of coffee. As long as we find the right person and do not break any laws in the process, it does not matter which process is used although some are more efficient than others.

The approach that we use for a digital investigation is based on the physical crime scene investigation process. In this case, we have a digital crime scene that includes the digital environment created by software and hardware. The process has three major phases, which are system preservation, evidence searching and event reconstruction.

These phases do not need to occur one after another and the flow is shown in Fig. 1.

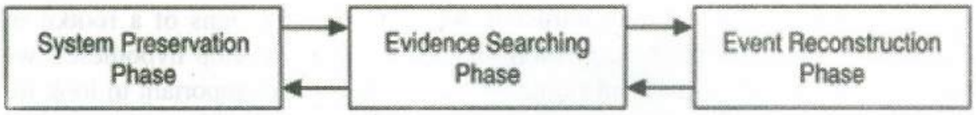


Fig. 1: The three major phases of a digital crime scene investigation

This process can be used when investigating both live and dead systems. A live *analysis* occurs when you use the operating system or other resources of the system being investigated to find evidence. *A dead analysis* occurs when you are running trusted applications in a trusted operating system to find evidence. With a live analysis, you risk getting false information because the software could maliciously hide or falsify data. A dead analysis is more ideal, but is not possible in all circumstances.

- **System Preservation Phase**

The first phase in the investigation process is the *System Preservation Phase* where we try to preserve the state of the digital crime scene. The actions that are taken in this phase vary depending on the legal, business or operational requirements of the investigation. For example, legal requirements may cause you to unplug the system and make a full copy of all data. On the other extreme could be a case involving a spyware infection or a honeypot (A honeypot is "*an information resource whose value lies in unauthorized or illicit use of that resource*") no preservation is performed. Most investigations in a corporate or military setting that will not go to court use techniques in between these two extremes.

The purpose of this phase is to reduce the amount of evidence that may be overwritten. This process continues after data has been acquired from the system because we need to preserve the data for future analysis.

- **Preservation Techniques**

The goal of this phase is to reduce the amount of evidence that is overwritten, so we want to limit the number processes that can write to our storage devices. For a dead analysis, we will terminate all processes by turning the system off and we will make duplicate copies of all data. Write blockers can be used to prevent evidence from being overwritten. For a live analysis, suspect processes can be killed or suspended. The network connection can be unplugged (plug the system into an empty hub or switch to prevent log messages about a dead link) or network filters can be applied so that the perpetrator cannot connect from a remote system and delete data. Important data should be copied from the system in case it is overwritten while searching for evidence. For example, if you are going to be reading files, then you can save the temporal data for each file so that you have a copy of the last access times before you cause them to be updated.

When important data are saved during a dead or live analysis, a cryptographic hash should be calculated to later show that the data have not changed. A cryptographic hash, such as MD5, SHA-1 and SHA-256 is a mathematical formula that generates a very big number based on input data. If any bit of the input data changes, the output number changes dramatically. The algorithms are designed such that it is extremely difficult to find two inputs that generate the same output. Therefore, if the hash values of important data changes, then you know that the data has been modified.

- **Evidence Searching Phase**

After we have taken steps to preserve the data we need to search them for evidence. Recall that we are looking for data that support or refute hypotheses about the incident. This process typically starts with a survey of common locations based on the type of incident, if one is known. For example, if we are investigating Web-browsing habits, we will look at the Web browser cache, history file and bookmarks. If we are investigating a Linux intrusion, we may look for signs of a rootkit or new user accounts. As the investigation proceeds and we develop hypotheses, we will search for evidence that will refute or support them. It is important to look for evidence that refutes your hypothesis instead of only looking for evidence that supports your hypothesis.

The theory behind the searching process is fairly simple. We define the general characteristics of the object for which we are searching and then look for that object in a collection of data. For example, if we want all files with the JPG extension, we will loosk at each file name and identify the ones that end with the characters ".JPG." The two key steps are determining for what we are looking and where we expect to find it.

- **Search Techniques**

Most searching for evidence is done in a file system and inside files. A common search technique is to search for files based on their names or patterns in their names. Another common search technique is to search for files based on a keyword in their content. We can also search for files based on their temporal data, such as the last accessed or written time.

We can search for known files by comparing the MD5 or SHA-1 hash of a file's content with a hash database such as the National Software Reference Library (NSRL) (http://www.nsrl.nist.gov). Hash databases can be used to find files that are known to be bad or good. Another common method of searching is to search for files based on signatures in their content. This allows us to find all files of a given type even if someone has changed their name. When analyzing network data, we may search for all packets from a specific source address or all packets going to a specific port. We also may want to find packets that have a certain keyword in them.

- **Event Reconstruction Phase**

The last phase of the investigation is to use the evidence that we found and determine what events occurred in the system. Our definition of an investigation was that we are trying to answer questions about digital events in the system. During the Evidence Searching Phase, we might have found several files that violate a corporate policy or law, but that does not answer questions about events. One of the files may have been the effect of an event that downloaded it, but we should also try to determine which application downloaded it. Is there evidence that a Web browser downloaded them or could it be from malware? (Several cases have used malware as a defence when contraband or other digital evidence has been found) After the digital event reconstruction phase, we may be able to correlate the digital events with physical events. Event reconstruction requires knowledge about the applications and the OS that are installed on the system so that you can create hypotheses based on their capabilities. For example, different events can occur in Windows 95 than Windows XP and different versions of the Mozilla Web browser can cause different events.

1.4.3 General Guidelines

Not every investigation will use the same procedures and there could be situations where you need to develop a new procedure. There are some techniques that have not been implemented, so you may have to improvise to find the evidence.

The first guideline is *preservation* of the system being investigated. The motivation behind this guideline is that you do not want to modify any data that could have been evidence and you do not want to be in a courtroom where the other side tries to convince the jury that you may have overwritten exculpatory evidence. This is what we saw in the Preservation Phase of the investigation process. Some examples of how the preservation guideline is implemented are:

- Copy important data, put the original in a safe place and analyze the copy so that you can restore the original if the data is modified.
- Calculate MD5 or SHA hashes of important data so that you can later prove that the data has not changed.
- Use a write-blocking device during procedures that could write to the suspect data.
- Minimize the number of files created during a live analysis because they could overwrite evidence in unallocated space.
- Be careful when opening files on the suspect system during a live analysis because you could be modifying data, such as the last access time.

The second guideline is to *isolate* the analysis environment from both the suspect data and the outside world. You want to isolate yourself from the suspect data because you do not know what it might do. Running an executable from the suspect system could delete all files on your computer or it could communicate with a remote system. Opening an HTML file from the suspect system could cause your Web browser to execute scripts and download files from a remote server. Both of these are potentially dangerous and caution should be taken.

Isolation from the suspect data is implemented by viewing data in applications that have limited functionality or in a virtual environment, such as VMW are (http://www.vmware.com) that can be easily rebuilt if it is destroyed. You should isolate yourself from the outside world so that no tampering can occur and so that you do not transmit anything that you did not want to. For example, the previous paragraph described how something as simple as an HTML page could cause you to connect to a remote server. Isolation from the outside world is typically implemented using an analysis network that is not connected to the outside world or that is connected using a firewall that allows only limited connectivity.

Note that isolation is difficult with live analysis. By definition, you are not isolated from the suspect data because you are analyzing a system using its OS, which is suspect code. Every action you take involves suspect data. Further, it is difficult to isolate the system from the outside world because that requires removing network connectivity and live analysis typically occurs because the system must remain active.

The third guideline is to *correlate* data with other independent sources. This helps reduce the risk of forged data. For example, we will later see that timestamps can be easily changed in most systems. Therefore, if time is very important in your investigation, you should try to find log entries, network traffic or other events that can confirm the file activity times.

The final guideline is to log and document your actions. This helps identify what searches you have not yet conducted and what your results were. When doing a live analysis or performing techniques that will modify data, it is important to document what you do so that you can later document what changes in the system were because of your actions.

1.4.4 Data Analysis

In the previous section, we were going to search for digital evidence, which is a rather general statement because evidence can be found almost anywhere. In this section, we will narrow down the different places where we can search for digital evidence and identify them. We will also discuss which data we can trust more than others.

- **Analysis Types**

When analyzing digital data, we are looking at an object that has been designed by people. Further, the storage systems of most digital devices have been designed to be scalable and flexible and they have a layered design. I will use this layered design to define the different analysis types.

If we start at the bottom of the design layers, there are two independent analysis areas. One is based on storage devices and the other is based on communication devices. This book is going to focus on the analysis of storage devices, specifically non-volatile devices such as hard disks.

Fig. 2 shows the different analysis areas. The bottom layer is Physical Storage Media Analysis and involves the analysis of the physical storage medium. Examples of physical store mediums include hard disks, memory chips and CD-ROMs. Analysis of this area might involve reading magnetic data from in between tracks or other techniques that require a clean room. We are going to assume that we have a reliable method of reading data from the physical storage medium and so we have a stream 1s and 0s that were previously written to the storage device.

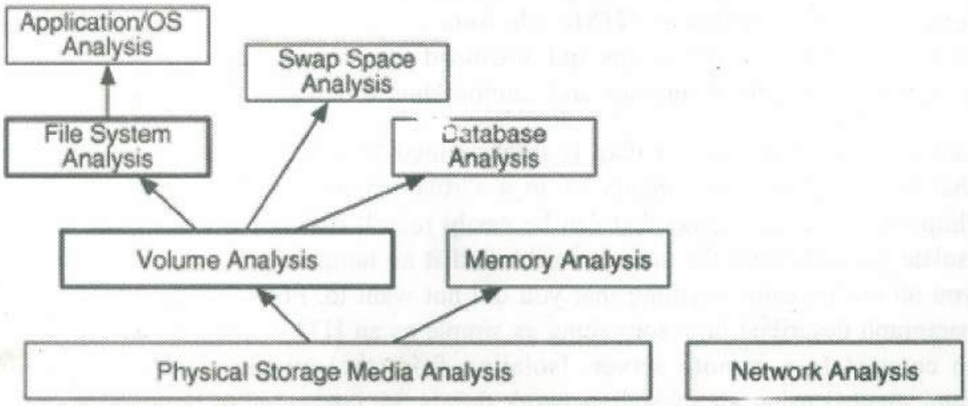


Fig. 2: Layers of analysis based on the design of digital data

We now analyze the 1s and 0s from the physical medium. Memory is typically organized by processes. We will focus on non-volatile storage, such as hard disks and flash cards.

Storage devices that are used for non-volatile storage are typically organized into volumes. A *volume* is a collection of storage locations that a user or application can write to and read from. There are two major concepts in this layer. One is partitioning, where we divide a single volume into multiple smaller volumes and the other is assembly, where we combine multiple volumes into one larger volume, which may later be partitioned. Examples of this category include DOS partition tables, Apple partitions and RAID arrays. Some media, such as floppy disks, do not have any data in this layer and the entire disk is a volume. We will need to analyze data at the volume level to determine where the file system or other data are located and to determine where we may find hidden data.

Inside each volume can be any type of data, but the most common contents are file systems. Other volumes may contain a database or be used as a temporary swap space (similar to the Windows pagefile). File Systems is a collection of data structures that allow an application to create, read and write files. We analyze a file system to find files, to recover deleted files and to find hidden data. The result of file system analysis could be file content, data fragments and metadata associated with files.

To understand what is inside of a file, we need to jump to the application layer. The structure of each file is based on the application or OS that created the file. For example, from the file system perspective, a Windows registry file is no different from an HTML page because they are both files. Internally, they have very different structures and different tools are needed to analyze each. Application analysis is very important and it is here where we would analyze configuration files to determine what programs were running or to determine what a JPEG picture is of.

We can see the analysis process in Fig. 3. This shows a disk that is analyzed to produce a stream of bytes, which are analyzed at the volume layer to produce volumes. The volumes are analyzed at the file system layer to produce a file. The file is then analyzed at the application layer.

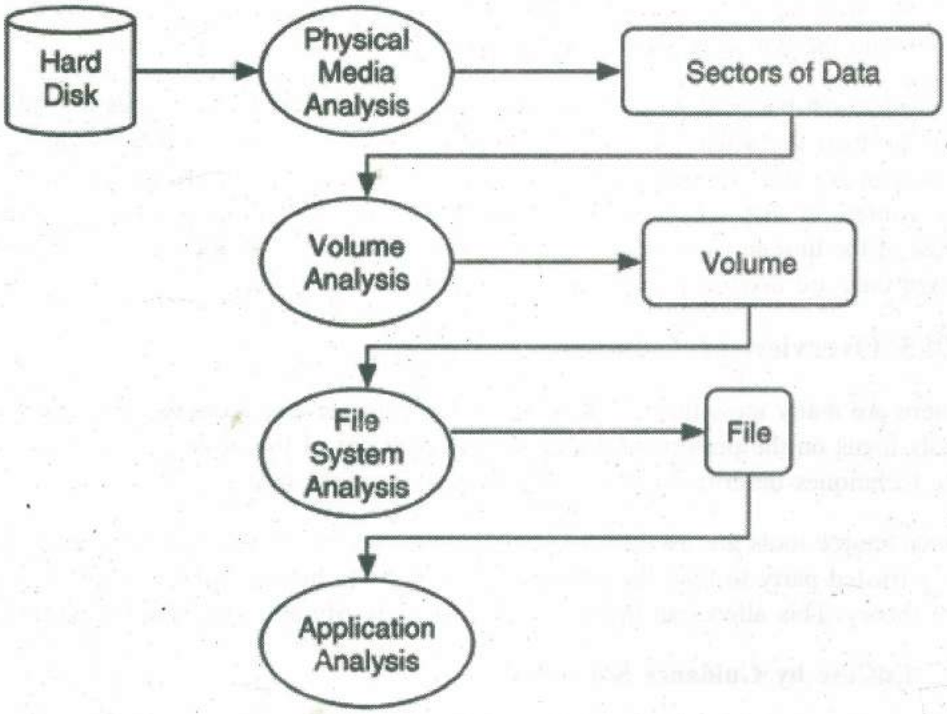


Fig. 3: Process of analyzing data at the physical level to the application level

- **Essential and Nonessential Data**

All data in the layers previously discussed have some structure but not all structure is necessary for the layer to serve its core purpose. For example, the purpose of the file system layer is to organize an empty volume so that we can store data and later retrieve them. The file system is required to correlate a file name with file content. Therefore, the name is essential and the on-disk location of the file content is essential. We can see this in Fig. 4 where we have a file named miracle.txt and its content is located at address 345. If either the name or the address were incorrect or missing, then the file content could not be read. For example, if the address were set to 344, then the file would have different content.

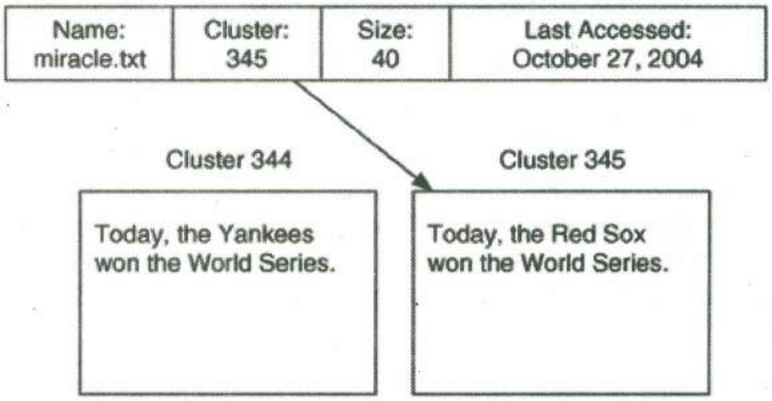

Name: miracle.txt
Cluster: 345
Size: 40
Last Accessed: October 27, 2004
Cluster 344
Cluster 345
Today, the Yankees won the World Series.
Today, the Red Sox won the World Series.

Fig. 4: To find and read this file, it is essential for the name, size and content location to be accurate, but it is not essential for the last accessed time to be accurate

Fig. 4 also shows that the file has a last accessed time. This value is not essential to the purpose of the file system and if it were changed, missing or incorrectly set, it would not affect the process of reading or writing file content.

The concepts of essential and nonessential data are introduced because we can trust essential data but we may not be able to trust nonessential data. We can trust that the file content address in a file is accurate because otherwise the person who used the system would not have been able to read the data. The last access time may or may not be accurate. The OS may not have updated it after the last access, the user may have changed the time or the OS clock could have been off by three hours and the wrong time was stored.

Note that just because we trust the number for the content address does not mean that we trust the actual content at that address. For example, the address value in a deleted file may be accurate, but the data unit could have been reallocated and the content at that address is for a new file. Non-essential data may be correct most of the time but you should try to find additional data sources to support them when they are used in an incident hypothesis.

1.4.5 Overview of Toolkits

There are many tools that can help an investigator analyze a digital system. Most tools focus on the preservation and searching phases of the investigation. Most of the techniques described can be performed using these tools.

Open source tools are useful for investigations because they allow an investigator or a trusted party to read the source code and verify how a tool has implemented the theory. This allows an investigator to better testify about the digital evidence.

- **EnCase by Guidance Software**

There are no official numbers on the topic but it is generally accepted that *EnCase* (http://www.encase.com) is the most widely used computer investigation software.

EnCase is Windows-based and can acquire and analyze data using the local or network-based versions of the tool. EnCase can analyze many file system formats, including FAT, NTFS, HFS+, UFS, Ext2/3, Reiser, JFS, CD-ROMs and DVDs. EnCase also supports Microsoft Windows dynamic disks and AIX LVM.

EnCase allows you to list the files and directories, recover deleted files, conduct keyword searches, view all graphic images, make timelines of file activity and use hash databases to identify known files. It also has its own scripting language called EnScript, which allows you to automate many tasks. Add-on modules support the decryption of NTFS encrypted files and allow you to mount the suspect data as though it were a local disk.

- **Forensic Toolkit by AccessData**

The *Forensic Toolkit* (FTK) is Windows-based and can acquire and analyze disk, file system and application data (http://www.accessdata.com). FTK supports FAT, NTFS and Ext2/3 file systems, but is best known for its searching abilities and application-level analysis support. FTK creates a sorted index of the words in a file system so that individual searches are much faster. FTK also has many viewers for different file formats and supports many e-mail formats.

FTK allows you to view the files and directories in the file system, recover deleted files, conduct keyword searches, view all graphic images, search on various file characteristics and use hash databases to identify known files. AccessData also has tools for decrypting files and recovering passwords.

- **ProDiscover by Technology Pathways**

ProDiscover (http://www.techpathways.com) is a Windows-based acquisition and analysis tool that comes in both local and network-based versions. ProDiscover can analyze FAT, NTFS, Ext2/3 and UFS file systems and Windows dynamic disks. When searching, it provides the basic options to list the files and directories, recover deleted files, search for keywords and use hash databases to identify known files. ProDiscover is available with a license that includes the source code so that an investigator or lab can verify the tool's actions.

- **SMART by ASR Data**

SMART (http://www.asrdata.com) is a Linux-based acquisition and analysis tool. Andy Rosen, who was the original developer for Expert Witness (which is now called EnCase), developed SMART. SMART takes advantage of the large number of file systems that Linux supports and can analyze FAT, NTFS, Ext2/3, UFS, HFS+, JFS, Reiser, CD-ROMs and more. To search for evidence, it allows you to list and filter the files and directories in the image, recover deleted files, conduct keyword searches, view all graphic images and use hash databases to identify known files.

- **The Sleuth Kit/Autopsy**

The Sleuth Kit (TSK) is a collection of Unix-based command line analysis tools and Autopsy is a graphical interface for TSK (http://www.sleuthkit.org). The file system tools in TSK are based on *The Coroner's Toolkit* (TCT) (http://www.porcupine.org), which was written by Dan Farmer and Wietse Venema. TSK and Autopsy can analyze FAT, NTFS, Ext2/3 and UFS file systems and can list files and directories, recover deleted files, make timelines of file activity, perform keyword searches and use hash databases.

Check Your Progress 2

Note: a) Space is given below for writing your answers.

b) Compare your answers with the one given at the end of this Unit.

1) When a complaint on information theft is received, how should one start investigation?

....

....

....

....

2) As an investigator, what are different things that have to be determined for conducting an effective investigation?

....

....

....

....

3) How can one define digital investigations and digital evidence?

....

....

....

....

4) What is the general process for conducting a digital investigation?

....

....

....

....

1.5 LET US SUM UP

There is no single way to conduct an investigation and we studied a brief overview of one approach that one can take. It has only three major phases and is based on a physical crime scene investigation procedure. We have also looked at the major investigation types and a summary of the available toolkits. In the next two units, we will look at the computer fundamentals and how to acquire data during the Preservation Phase of an investigation.

1.6 CHECK YOUR PROGRESS – THE KEY

Check Your Progress 1

Electronic Tampering: Electronic tampering can involve fraud, mimicking someone or something (i.e. IP spoofing), masking or masquerading as someone (i.e. social engineering). The intent and result of the tampering is the primary reason to conduct an investigation.

Even if the intent of the tampering involves or can be linked to a non-competitive prank, there is still reason to investigate. If any tampering can occur regardless of the reason, then it should be prevented to protect the organization's information assets.

Check Your Progress 2

1) On receipt of a complaint, one has to first establish a basis or justification for investigation. Once justification for investigation is established, then relevant rules or a baseline for which the complaint has been lodged which could be company policies or procedures needs to be checked for violation. These would then guide the investigator to consult the appropriate baselines and look out for any document penalties. Organizations may have many policies and procedures and each will have to be looked into. Planning of investigation would largely depend on this examination. Besides the policies and procedures, the investigator should also keep in mind the legal statues and law of land with respect to such crimes.

2) As an investigator one requires to look into various components that require attention and these might provides vital evidences for effective investigation of the complaint. Resources at the disposal of the investigator are very important component for planning the investigation. What is the level of authority has been bestowed upon the investigator would provide the clue to which level of security access that has been provided to the investigator. Then the investigator needs to know what are obligations and goal both ethically and professionally so that the complaint can be given due respect. Investigator also needs to understand the reporting hierarchy as this is an important area where the jurisdiction of investigation depends. This would then lead the investigator to the escalation procedures in case of any problem/cooperation requirements. The investigator should also know the time frame for investigation so that the planning can properly be done. Before actually starting the process of investigation, knowledge about previous precedence would be very handy.

3) A digital investigation is a process where we develop and test hypotheses that answer questions about digital events. This is done using the scientific method where we develop a hypothesis using evidence that we find and then test the hypothesis by looking for additional evidence that shows the hypothesis is impossible. It is a process to answer questions about digital states and events. The basic digital investigation process frequently occurs by all computer users when they, for example, search for a file on their computer. They are trying to answer the question "what is the full address of the file named important.doc?" In general, digital investigations may try to answer questions such as "does file X exist?", "was program Y run?" or "was the user Z account compromised?"

   A digital investigation is a special case of a digital investigation where the procedures and techniques that are used will allow the results to be entered into a court of law. For example, an investigation may be started to answer a question about whether or not contraband digital images exist on a computer. An average Microsoft Windows user may be able to answer this question by booting the computer and using the Find Files function but these results may not be court admissible because steps were not taken to preserve the state of the computer or use trusted tools.

   The digital investigation process involves formulating and testing hypotheses about the state of a computer. We must formulate hypotheses because we can not directly observe digital events and states and therefore we do not know facts. We must use tools to observe the state of digital data, which makes them indirect observations. This is similar to being told about something instead of seeing it for yourself. The amount that you believe what you are told is based on how much you trust the person. With digital investigations, the confidence is based on the trust of the hardware and software used to collect and analyze the data. The methods used to formulate and test the hypotheses can make the investigation process a scientific one.

Digital evidence is data that supports or refutes a hypothesis that was formulated during the investigation. This is a general notion of evidence and may include data that may not be court admissible because it was not properly or legally acquired. It is defined as any data stored or transmitted using a computer that support or refute a theory of how an offence occurred or address critical elements of the offence or as any data stored or transmitted, using a computer, that supports or refutes a theory of how an offence occurred or addresses critical elements of the offence or prove a relevant aspect of the facts at issue. The term evidence is always used in the context of investigation. This has both legal and investigative uses.

4) There is no single procedure for conducting an investigation. We find that an intuitive procedure is to apply the same basic phases that are used by police at a physical crime scene, where we instead have a digital crime scene.

The first step is preservation, where we attempt to preserve the crime scene so that the evidence is not lost. In the physical world, yellow tape is wrapped around the scene. In a digital world, we make a copy of memory, power the computer off and make a copy of the hard disk. In some cases, the computer cannot be powered off and instead suspicious processes are killed and steps are taken to ensure that known evidence is copied and preserved.

The second step is to survey the crime scene for the obvious evidence. The "obvious" evidence is the evidence that typically exists with investigations of this type. For example, at a physical crime scene where a violent crime has occurred, then the "obvious" evidence may have blood on it or be damaged. In a digital crime scene, the obvious evidence may be found based on file types, keywords and other characteristics.

After the obvious evidence has been found, then more exhaustive searches are conducted to start filling in the holes. With each piece of evidence that is found, there could be questions about how it got there. Questions such as "which application created it?" or "what user caused it to be created?". If so, then event reconstruction techniques are needed to determine which application-level event occurred. This is similar to reconstructing where a bullet was shot from.

1.7 SUGGESTED READINGS

- Bejtlich, Richard. The Tao of *Network Security Monitoring: Beyond Intrusion Detection*.Boston: Addison Wesley, 2005.
- Brenner, Susan, Brian Carrier and Jef Henninger. "The Trojan Defense in Cybercrime Cases." *Santa Clara Computer and High Technology Law Journal*, 21(1), 2004.
- Carrier, Brian and Eugene H. Spafford. "Getting Physical with the Digital Investigation Process." *International Journal of Digital Evidence*, Fall 2003. http://www.ijde.org.

- Carrier, Brian. "Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers." *International Journal of Digital Evidence*, Winter 2003a. http://www.ijde.org.
- Carrier, Brian. "Open Source Digital Forensic Tools: The Legal Argument." Fall 2003b. http://www.digital-evidence.org.
- Casey, Eoghan. *Digital Evidence and Computer Crime*. 2nd ed. London: Academic Press, 2004.
- Clifford, Ralph, ed. *Cybercrime: The Investigation, Prosecution and Defense of a Computer-Related Crime*. Durham: Carolina Academic Press, 2001.
- George, Esther. "UK Computer Misuse Act-The Trojan Virus Defense." *Journal of Digital Investigation*, 1(2), 2004.
- Houghton Mifflin Company. *The American Heritage Dictionary*. 4th ed. Boston: Houghton Mifflin, 2000.
- Mandia, Kevin, Chris Prosise and Matt Pepe. *Incident Response and Computer Forensics*. 2nd ed. Emeryville: McGraw Hill/Osborne, 2003.
- Schneier, Bruce. *Applied Cryptography*. 2nd ed. New York: Wiley Publishing, 1995.
- The Honeynet Project. *Know Your Enemy*. 2nd ed. Boston: Addison-Wesley, 2004.

UNIT 2 DATA ACQUISITION AND INFORMATION GATHERING

Structure

2.0 INTRODUCTION

The field of computer forensics involves identifying, extracting, documenting and preserving information that is stored or transmitted in electronic or magnetic form (i.e. digital evidence). Like fingerprints, digital evidence can be visible (such as files stored on disk that can be accessed via the normal directory structure using standard file management tools such as Windows Explorer) or it can be latent (not readily visible or accessible, requiring some sort of processing-via special software or techniques-to locate and identify it). An important aspect of computer forensics involves finding and evaluating this "hidden data" for its evidentiary value.

Computer forensics standards have been developed that apply to the collection and preservation of digital evidence, which differs in nature from most other types of evidence and thus requires different methods of handling. Following procedures that are proper, accepted and in some cases, prescribed by law in dealing with evidence is vital to the successful prosecution of a cybercrime case. The proper handling of these procedures comes into play at two different points in a trial:

- If the evidence is not collected and handled according to the proper standards, the judge may deem the evidence inadmissible when it is presented (usually based on the opposing attorney's *motion to suppress*) and the jury members will never get a chance to evaluate it or consider it in making their decision.
- If the evidence is admitted, the opposing attorney will attack its credibility during questioning of the witnesses who testify regarding it. Such an attack can create doubt in jury members' minds that will cause them to disregard the evidence in making their decision-and perhaps even taint the credibility of the entire case.

The entire investigation will be of little value if the evidence that shows the defendant's guilt is not allowed into the trial or if the jury gives it no weight.

Thus proper handling of evidence is one of the most important issues facing all criminal investigators and because of the intangible nature of digital evidence, cyber crime investigators in particular. Because this is such an important topic-not only for investigators, but for prosecutors, judges and justice system professionals involved in cybercrime cases-many organizations and publications are devoted solely to issues concerning digital evidence. The International Organization of Computer Evidence (IOCE) was established in 1995 to provide a forum for law enforcement agencies across the world to exchange information about computer forensics issues.

The International Association of Computer Investigative Specialists (IACIS; www.cops.org) is a non-profit organization that is dedicated to educating law enforcement professionals in the area of computer forensics. The *International Journal of Digital Evidence* (www.ijde.org) is an online publication devoted to discussions of the theory and practice of handling digital evidence. *Computer Forensics Magazine* (www.forensic-computing.com) is published by DIBS, a maker of computer forensics equipment. *Computer Forensics Online* (www.shk-dplc.com/cfo) is a webzine that is run by attorneys and technical professionals specializing in computer law. Many other similar resources that focus on computer forensics are available and more broad-based organizations such as the American Academy of Forensic Sciences (www.aafs.org) address computer crimes and digital evidence along with other forensics topics.

A glance at any of these resources will reveal that digital evidence handling is a huge topic that could easily fill several books (and already has). It is far beyond the scope of this unit to cover every aspect of collecting and preserving digital evidence. This unit provides an overview of the role evidence plays in a criminal case (particularly in a cybercrime case) and discusses standard procedures for dealing with digital evidence as well as specific evidence location and examination techniques such as recovering supposedly deleted files, finding steganographic data, locating "forgotten" data and decrypting encrypted data. Procedures for documenting digital evidence are also outlined and examination some of the legal issues involved in evidence collection and handling will also be done. Finally, we provide many excellent online resources that furnish detailed instructions for performing the tasks described in this unit and provide information about commercial services and equipment that can aid in the evidence recovery process.

2.1 OBJECTIVES

After studying this unit, you should be able to:

- understand various aspects of Electronic evidence;
- explain different types of evidence;
- understand and explain about data acquisition and information gathering; and
- elucidate various methods of collection.

2.2 DATA ACQUISITION

2.2.1 Why Collect Evidence?

Electronic evidence can be very expensive to collect. The processes are strict and exhaustive, the systems affected may be unavailable for regular use for a long period of time and analysis of the data collected must be performed. So why bother collecting the evidence in the first place? There are two simple reasons: future prevention and responsibility.

- **Future Prevention**

Without knowing what happened, you have no hope of ever being able to stop someone else (or even the original attacker) from doing it again. It would be analogous to not fixing the lock on your door after someone broke in. Even though the cost of collection can be high, the cost of repeatedly recovering from compromises is much higher, both in monetary and corporate image terms.

- **Responsibility**

There are two responsible parties after an attack: the attacker and the victim. The attacker is responsible for the damage done and the only way to bring him to justice (and to seek recompense) is with adequate evidence to prove his actions.

The victim on the other hand, has a responsibility to the community. Information gathered after a compromise can be examined and used by others to prevent further attacks. The victim may also have a legal obligation to perform an analysis of evidence collected, for instance if the attack on their system was part of a larger attack.

2.2.2 Collection Options

Once a compromise has been detected, you have two options: pull the system off the network and begin collecting evidence or leave it online and attempt to monitor the intruder. Both have their pros and cons. In the case of monitoring, you may accidentally alert the intruder while monitoring and cause him to wipe his tracks any way necessary destroying evidence as he goes. You also leave yourself open to possible liability issues if the attacker launches further attacks at other systems from your own network system. If you disconnect the system from the network, you may find that you have insufficient evidence or worse that the attacker left a *dead man* switch that destroys any evidence once the system detects that it is offline. What you choose to do should be based on the situation. The "Collection and Archiving" section later in the unit contains information on what to do for either case.

2.2.3 Obstacles

Electronic crime is difficult to investigate and prosecute. Investigators have to build their case purely on any records left after the transactions have been completed. Add to this the fact that electronic records are extremely (and sometimes transparently) malleable and that electronic transactions currently have fewer limitations than their paper-based counterparts and you get a collection nightmare.

Computer transactions are fast. They can be conducted from anywhere (through anywhere, to anywhere) can be encrypted or anonymous and have no intrinsic identifying features such as handwriting and signatures to identify those responsible.

Any *paper trail* of computer records they may leave can be easily modified or destroyed or may be only temporary. Worse still, auditing programs may automatically destroy the records left when computer transactions are finished with them. Because of this, even if the details of the transactions can be restored through analysis, it is very difficult to tie the transaction to a person. *Identifying* information such as passwords or PIN numbers (or any other electronic identifier) does not prove who was responsible for the transaction. Such information merely shows that whoever did it either knew or could get past those identifiers.

Even though technology is constantly evolving, investigating electronic crimes will always be difficult because of the ease of altering the data and the fact that transactions may be done anonymously. The best you can do is to follow the rules of evidence collection and be as assiduous as possible.

2.2.4 Types of Evidence

Before you start collecting evidence, it is important to know the different types of evidence categories. Without taking these into consideration, you may find that the evidence you've spent several weeks and quite a bit of money collecting is useless.

Real evidence is any evidence that speaks for itself without relying on anything else. In electronic terms, this can be a log produced by an audit function-provided that the log can be shown to be free from contamination.

Testimonial Evidence

Testimonial evidence is any evidence supplied by a witness. This type of evidence is subject to the perceived reliability of the witness, but as long as the witness can be considered reliable, testimonial evidence can be almost as powerful as real evidence.

Word processor documents written by a witness may be considered testimonial- as long as the author is willing to state that he wrote it.

Hearsay

Hearsay is any evidence presented by a person who was not a direct witness. Word processor documents written by someone without direct knowledge of the incident are hearsay. Hearsay is generally inadmissible in court and should be avoided.

2.2.5 The Rules of Evidence

There are five rules of collecting electronic evidence. These relate to five properties that evidence must have to be useful.

1) Admissible

2) Authentic

3) Complete

4) Reliable

5) Believable

- **Admissible**

Admissible is the most basic rule. The evidence must be able to be used in court or otherwise. Failure to comply with this rule is equivalent to not collecting the evidence in the first place, except the cost is higher.

- **Authentic**

If you can't tie the evidence positively to the incident, you can't use it to prove anything. You must be able to show that the evidence relates to the incident in a relevant way.

- **Complete**

It's not enough to collect evidence that just shows one perspective of the incident. You collect not only evidence that can prove the attacker's actions but also evidence that could prove their innocence. For instance, if you can show the attacker was logged in at the time of the incident, you also need to show who else was logged in and why you think they didn't do it. This is called *exculpatory evidence* and is an important part of proving a case.

- **Reliable**

The evidence you collect must be reliable. Your evidence collection and analysis procedures must not cast doubt on the evidence's authenticity and veracity.

- **Believable**

The evidence you present should be clearly understandable and believable to the investigating agency. There's no point presenting a binary dump of process memory if the jury has no idea what it all means. Similarly, if you present them with a formatted, not possible for a normal human to understand, you must be able to show the relationship to the original binary, otherwise there's no way for the jury to know whether you've faked it. Using the preceding five rules, you can derive some basic dos and don'ts:

- Minimize handling and corruption of original data.
- Account for any changes and keep detailed logs of your actions.
- Comply with the five rules of evidence.
- Do not exceed your knowledge.
- Follow your local security policy.
- Capture as accurate an image of the system as possible.
- Be prepared to testify.
- Work fast.
- Proceed from volatile to persistent evidence.
- Don't shutdown before collecting evidence.
- Don't run any programs on the affected system.

Minimize Handling and Corruption of Original Data

Once you've created a master copy of the original data, don't touch it or the original. Always handle secondary copies. Any changes made to the originals will affect the outcomes of any analysis later done to copies. You should make sure you don't run any programs that modify the access times of all files (such as tar and xcopy).

You should also remove any external avenues for change and in general, analyze the evidence after it has been collected.

Account for Any Changes and Keep Detailed Logs of Your Actions

Sometimes evidence alteration is unavoidable. In these cases, it is absolutely essential that the nature, extent and reasons for the changes be documented. Any changes at all should be accounted for-not only data alteration but also physical alteration of the originals (the removal of hardware components).

Comply with the Five Rules of Evidence

The five rules are there for a reason. If you don't follow them, you are probably wasting your time and money. Following these rules is essential to guaranteeing successful evidence collection.

Do Not Exceed Your Knowledge

If you don't understand what you are doing, you can't account for any changes you make and you can't describe what exactly you did. If you ever find yourself "out of your depth" either go or learn more before continuing (if time is available) or find someone who knows the territory. Never soldier on regardless. You'll just damage your case.

Capture as Accurate an Image of the System as Possible

Capturing an accurate image of the system is related to minimizing the handling or corruption of original data. Differences between the original system and the master copy count as a change to the data. You must be able to account for the differences.

Be Prepared to Testify

If you're not willing to testify to the evidence you have collected, you might as well stop before you start. Without the collector of the evidence being there to validate the documents created during the evidence-collection process, the evidence becomes hearsay which is inadmissible. Remember that you may need to testify at a later time. No one is going to believe you if they can't replicate your actions and reach the same results. This also means that your plan of action shouldn't be based on trial-and-error.

Work Fast

The faster you work, the less likely the data is going to change. Volatile evidence may vanish entirely if you don't collect it in time. This is not to say that you should rush. You must still collect accurate data. If multiple systems are involved, work on them in parallel (a team of investigators would be handy here) but each single system should still be worked on methodically. Automation of certain tasks makes collection proceed even faster.

Proceed from Volatile to Persistent Evidence

Some electronic evidence (discussed later) is more volatile than others are. Because of this, you should always try to collect the most volatile evidence first.

Don't Shutdown before Collecting Evidence

You should never, ever shutdown a system before you collect the evidence. Not only do you lose any volatile evidence but also the attacker may have trojaned (via a trojan horse) the startup and shutdown scripts, plug-and-play devices may alter the system configuration and temporary file systems may be wiped out. Rebooting is even worse and should be avoided at all costs. As a general rule, until the compromised disk is finished with and restored. It should never be used as a boot disk.

Don't Run Any Programs on the Affected System

Because the attacker may have left trojaned programs and libraries on the system, you may inadvertently trigger something that could change or destroy the evidence you're looking for. Any programs you use should be on read-only media (such as a CD-ROM or a write-protected floppy disk) and should be statically linked.

2.2.6 Volatile Evidence

Not all the evidence on a system is going to last very long. Some evidence resides in storage that requires a consistent power supply; other evidence may be stored in information that is continuously changing. When collecting evidence, you should always try to proceed from the most volatile to the least. Of course, you should still take the individual circumstances into account. You shouldn't waste time extracting information from an unimportant or unaffected machine's main memory when an important or affected machine's secondary memory hasn't been examined.

To determine what evidence to collect first, you should draw up an order of volatility-a list of evidence sources ordered by relative volatility. An example of an order of volatility would be:

1) Registers and cache
2) Routing tables
3) Arp cache
4) Process table
5) Kernel statistics and modules
6) Main memory
7) Temporary file systems
8) Secondary memory
9) Router configuration
10) Network topology

Once you have collected the raw data from volatile sources you may be able to shut down the system.

2.2.7 General Procedure

When collecting and analyzing evidence, there is a general four-step procedure you should follow. Note that this is a very general outline. You should customize the details to suit your situation.

- **Identification of Evidence**

You must be able to distinguish between evidence and junk data. For this purpose, you should know what the data is, where it is located and how it is stored. Once this is done, you will be able to work out the best way to retrieve and store any evidence you find.

- **Preservation of Evidence**

The evidence you find must be preserved as close as possible to its original state. Any changes made during this phase must be documented and justified.

- **Analysis of Evidence**

The stored evidence must then be analyzed to extract the relevant information and recreate the chain of events. Analysis requires in-depth knowledge of what you are looking for and how to get it. Always be sure that the person or people who are analyzing the evidence are fully qualified to do so.

- **Presentation of Evidence**

Communicating the meaning of your evidence is vitally important-otherwise you can't do anything with it. The manner of presentation is important and it must be understandable by a layman to be effective. It should remain technically correct and credible. A good presenter can help in this respect.

Check Your Progress 1

Note: a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of this Unit.

What general procedures should be followed while collecting electronic evidence?

....

....

....

....

2.3 COLLECTING AND ARCHIVING

Once you've developed a plan of attack and identified the evidence that needs to be collected, it's time to start the actual process of capturing the data. Storage of that data is also important, as it can affect how the data is perceived.

Logs and Logging

You should run some kind of system logging function. It is important to keep these logs secure and to back them up periodically. Because logs are usually automatically timestamped, a simple copy should suffice, although you should digitally sign and encrypt any logs that are important to protect them from contamination. Remember, if the logs are kept locally on the compromised machine, they are susceptible to either alteration or deletion by an attacker. Having a remote syslog server and storing logs in a sticky directory can reduce this risk, although it is still possible for an attacker to add decoy or junk entries into the logs.

Regular auditing and accounting of your system is useful not only for detecting intruders but also as a form of evidence. Messages and logs from programs can be used to show what damage an attacker did. Of course, you need a clean snapshot for these to work, so there's no use trying it after the compromise.

Monitoring

Monitoring network traffic can be useful for many reasons-you can gather statistics, watch out for irregular activity (and possibly stop an intrusion before it happens) and trace where an attacker is coming from and what he is doing.

Monitoring logs as they are created can often show you important information you might have missed had you seen them separately. This doesn't mean you should ignore logs later-it may be what's missing from the logs that are suspicious.

Information gathered while monitoring network traffic can be compiled into statistics to define normal behaviour for your system. These statistics can be used as an early warning of an attacker's actions. You can also monitor the actions of your users. This can once again act as an early warning system. Unusual activity or the sudden appearance of unknown users should be considered definite cause for closer inspection.

No matter the type of monitoring done, you should be very careful. There are plenty of laws you could inadvertently break. In general, you should limit your monitoring to traffic or user information and leave the content unmonitored unless the situation necessitates it. You should also display a disclaimer stating what monitoring is done when users log on. The content of this should be worked out in conjunction with your lawyer.

2.3.1 Methods of Collection

There are two basic forms of collection: *freezing the scene* and *honeypotting*. The two aren't mutually exclusive. You can collect frozen information after or during any honeypotting.

Freezing the scene involves taking a snapshot of the system in its compromised state. The necessary authorities should be notified (the police and your incident response and legal teams), but you shouldn't go out and tell the world just yet. You should then start to collect whatever data is important onto removable non-volatile media in a standard format. Make sure the programs and utilities used to collect the data are also collected onto the same media as the data. All data collected should have a cryptographic message digest created and those digests should be compared to the originals for verification.

Honeypotting is the process of creating a replica system and luring the attacker into it for further monitoring. A related method (sandboxing) involves limiting what the attacker can do while still on the compromised system, so he can be monitored without (much) further damage. The placement of misleading information and the attacker's response to it is a good method for determining the attacker's motives. You must make sure that any data on the system related to the attacker's detection and actions is either removed or encrypted; otherwise they can cover their tracks by destroying it. Honeypotting and sandboxing are extremely resource intensive, so they may be infeasible to perform.

2.3.2 Artifacts

Whenever a system is compromised, there is almost always something left behind by the attacker-be it code fragments, trojaned programs, running processes or sniffer log files. These are known as *artifacts*. They are one of the important things you should collect but you must be careful. You should never attempt to analyze an artifact on the compromised system. Artifacts are capable of anything and you want to make sure their effects are controlled.

Artifacts may be difficult to find; trojaned programs may be identical in all obvious ways to the originals (file size, medium access control [MAC] times, etc.). Use of cryptographic checksums may be necessary, so you may need to know the original file's checksum. If you are performing regular file integrity assessments, this shouldn't be a problem. Analysis of artifacts can be useful in finding other systems the attacker (or his tools) has broken into.

2.3.3 Collection Steps

You now have enough information to build a step-by-step guide for the collection of the evidence. Once again, this is only a guide. You should customize it to your specific situation. You should perform the following collection steps:

1) Find the evidence.

2) Find the relevant data.

3) Create an order of volatility.

4) Remove external avenues of change.

5) Collect the evidence.

6) Document everything.

- **Find the Evidence**

Determine where the evidence you are looking for is stored. Use a checklist. Not only does it help you to collect evidence but it also can be used to double-check that everything you are looking for is there.

- **Find the Relevant Data**

Once you've found the evidence, you must figure out what part of it is relevant to the case. You must remember that you have to work fast. Don't spend hours collecting information that is obviously useless.

- **Create an Order of Volatility**

Now that you know exactly what to gather, work out the best order in which to gather it. The order of volatility for your system is a good guide and ensures that you minimize loss of uncorrupted evidence.

- **Remove External Avenues of Change**

It is essential that you avoid alterations to the original data and prevention is always better than a cure. Preventing anyone from tampering with the evidence helps you create as exact an image as possible. However, you have to be careful. The attacker may have been smart and left a dead-man switch. In the end, you should try to do as much as possible to prevent changes.

- **Collect the Evidence**

You can now start to collect the evidence using the appropriate tools for the job. As you go, re-evaluate the evidence you've already collected. You may find that you missed something important. Now is the time to make sure you get it.

- **Document Everything**

Your collection procedures may be questioned later, so it is important that you document everything you do. Timestamps, digital signatures and signed statements are all important. Don't leave anything out.

After briefly understanding the nuances of information collection let us start with identifying potential data sources, for which the analyst needs to acquire the data from the sources. Data acquisition should be performed using a three-step process: developing a plan to acquire the data, acquiring the data and verifying the integrity of the acquired data.

1) **Develop a plan to acquire the data.** Developing a plan is an important first step in most cases because there are multiple potential data sources. The analyst should create a plan that prioritizes the sources, establishing the order in which the data should be acquired. Important factors for prioritization include the following:

   Likely Value. Based on the analysts understanding of the situation and previous experience in similar situations, the analyst should be able to estimate the relative likely value of each potential data source.

   Volatility. Volatile data refers to data on a live system that is lost after a computer is powered down or due to the passage of time. Volatile data may also be lost as a result of other actions performed on the system. In many cases, acquiring volatile data should be given priority over non-volatile data. However, non-volatile data may also be somewhat dynamic in nature (e.g. log files that are overwritten as new events occur).

Amount of Effort Required

The amount of effort required to acquire different data sources may vary widely. The effort involves not only the time spent by analysts and others within the organization (including legal advisors) but also the cost of equipment and services (e.g. outside experts). For example, acquiring data from a network router would probably require much less effort than acquiring data from an ISP. By considering these three factors for each potential data source, analysts can make informed decisions regarding the prioritization of data source acquisition, as well as determining which data sources to acquire. In some cases, there are so many possible data sources that it is not practical to acquire them all. Organizations should carefully consider the complexities of prioritizing data source acquisition and develop written plans, guidelines and procedures that can help analysts perform prioritization effectively.

2) **Acquire the data.** If the data has not already been acquired by security tools, analysis tools or other means, the general process for acquiring data involves using forensic tools to collect volatile data, duplicating non-volatile data sources to collect their data and securing the original non-volatile data sources. Data acquisition can be performed either locally or over a network. Although it is generally preferable to acquire data locally because there is greater control over the system and data, local data collection is not always feasible (e.g. system in locked room, system in another location). When acquiring data over a network, decisions should be made regarding the type of data to be collected and the amount of effort to use. For instance, it might be necessary to acquire data from several systems through different network connections or it might be sufficient to copy a logical volume from just one system.

3) **Verify the integrity of the data.** After the data has been acquired, its integrity should be verified. It is particularly important for an analyst to prove that the data has not been tampered with if it might be needed for legal reasons. Data integrity verification typically consists of using tools to compute the message digest of the original and copied data, then comparing the digests to make sure that they are the same.

Before the analyst begins to collect any data, a decision should be made by the analyst or management (in accordance with the organizations policies and legal advisors) on the need to collect and preserve evidence in a way that supports its use in future legal or internal disciplinary proceedings. In such situations, a clearly defined chain of custody should be followed to avoid allegations of mishandling or tampering of evidence. This involves keeping a log of every person who had physical custody of the evidence, documenting the actions that they performed on the evidence and at what time, storing the evidence in a secure location when it is not being used, making a copy of the evidence and performing examination and analysis using only the copied evidence and verifying the integrity of the original and copied evidence. If it is unclear whether or not evidence needs to be preserved by default it generally should be preserved.

In addition, several other steps should be taken. Throughout the process, a detailed log should be kept of every step that was taken to collect the data, including information about each tool used in the process. The documentation allows other analysts to repeat the process later if needed. Additionally, evidence should be photographed to provide visual reminders of the computer setup and peripheral devices. In addition, before actually touching a system, the analyst should make a note or photograph of any pictures, documents, running programs and other relevant information displayed on the monitor. If a screen saver is active that should be documented as well since it may be password-protected. If possible,. one person on the scene should be designated the evidence custodian and given the sole responsibility to photograph, document and label every item that is collected and record every action that was taken along with who performed the action, where it was performed and at what time. Since the evidence may not be needed for legal proceedings for an extended time, proper documentation enables an analyst to remember exactly what was done to collect data and can be used to refute claims of mishandling.

To assist the analyst with evidence collection, the necessary resources such as forensic workstations, backup devices, blank media and evidence handling supplies (e.g. hard-bound notebooks, chain of custody forms, evidence storage bags and tags, evidence tape, digital cameras) should be prepared beforehand. In some cases, it may be necessary to ensure that the scene is physically secured to prevent unauthorized access and alteration of the evidence. This may be as simple as having a physical security staff member guard a room. There also may be situations where a law enforcement representative should handle the data collection for legal reasons. This includes, but is not limited to obtaining ISP records and collecting data from external computer systems and unusual devices and media. Based on guidance from legal advisors organizations should determine in advance what types of data are best collected by law enforcement officials.

Analysis should take into account what will be done with the collected data and plan for the potential ramifications. In some cases, the data may be turned over to a law enforcement agency or another external party for examination and analysis. This could result in the collected hardware being unavailable for an extended period of time. If the original media needs to be kept secured for legal proceedings, it could be unavailable for years. Another concern is that sensitive information unrelated to the investigation (e.g. medical records, financial information) might be inadvertently captured along with the desired data.

The guide is a first responder's reference to different types of electronic evidence and includes procedures that can be used to safely handle them. The guide is oriented towards those who respond to the physical crime scene, so emphasis is placed on those requirements. Little attention is paid to the analysis of the system. The following phases are given:

- **Preparation:** Prepare equipment and tools to perform needed tasks during an investigation.
- **Collection:** Search for and collect electronic evidence.
- **Secure and Evaluate the Scene:** Secure the scene to ensure the safety of people and the integrity of evidence. Potential evidence should be identified in this phase.
- **Document the Scene:** Document the physical attributes of the scene including photos of the computer.
- **Evidence Collection:** Collect the physical system or make a copy of the data on the system.
- **Examination:** A technical review of the system for evidence.
- **Analysis:** The Investigation team reviews the examination results for their value in the case.
- **Reporting:** Examination notes are created after each case.

Before attempting to collect evidence from a computer, it is important to have a solid understanding of how forensic science is applied to computers. It is also critical to follow a standard procedure when collecting evidence to ensure consistency and avoid mistakes or oversights.

There are two key issues when it comes to actually collecting digital evidence: authenticity and integrity. You need to be able to demonstrate that the evidence is what you say it is, came from where you say it came from and has not been modified since you obtained it. How you document evidence to demonstrate that it is authentic and reliable depends heavily on the circumstances and the computer systems you are dealing with.

Software is available to preserve evidence stored on a standalone personal computer. Also, *The SleuthKit* is a free tool for examining digital evidence. Another free tool called The Coroner's Toolkit (TCT) is being developed to collect both static and volatile evidence from a computer.

Logical Data Acquisition

Logical data acquisition refers to data extraction techniques without direct access to raw memory data. Logical data acquisition often uses common, high level protocols to extract information elements that can be interpreted by a user without additional decoding. Most forensic phone software use different methods to extract phone book entries, SMS messages and call related data from mobile phones. Some other popular protocols for logical data acquisition are SyncML, OBEX, IrMC, ActiveSync, Fbus, ISO7816-3. Another class of forensic logical data acquisition tools runs a kind of software agent on the exhibit, using the original operating system to transfer file system data to an examination host. Logical data acquisition methods do not generally capture deleted data.

Commercial, non-forensic, backup and synchronization tools that can be used for logical data acquisition are available for many embedded systems, including marine and handheld GPS (Global Positioning System) units. In contrast to navigation systems in vehicles, handheld GPS units generally store position data rather than complete routes. Three types of position data can be distinguished:

- **Track-log:** A FIFO (first in/first out) buffer in which the current position of the GPS unit is continuously stored as soon as this differs from the original position.

- **Routes:** The series of points from the track-log where the alteration in course took place. This information is stored on the initiative of the user.

- **Waypoints:** Autonomous positions kept by the user that can often be provided with a brief text (Home, Pub etc.). A standardized interface exists for getting live position data from GPS equipment (NMEA). This interface is not very useful for forensic acquisition but data formatted according to NMEA specifications can sometimes be found in volatile or non-volatile memories of GPS enabled devices. Various software are available for reading post-mortem position data and visualizing it on a map.

A vehicle electronic control unit (ECU) is another example of an embedded system that can be acquired logically. For vehicle systems that incorporate computer control, the assembly containing the microprocessor is called ECU. ECUs control vehicle functionality like antilock breaking, air bags and seat-belt tensions. ECUs get data from sensors measuring vehicle speed, wheel speed, deceleration among others. They transmit control signals to actuators like valves, air bag igniters and dashboard indicators. Within an ECU, non-volatile data is saved in EEPROM or flash memory. This information usually includes diagnostic trouble codes (DTCs) and optional parametric crash data. Data obtained from one or more ECUs are often called *black box data* and might be useful for forensic investigations. Connectors and connection protocols for vehicle diagnostics are highly standardized and a lot of commercial scanners exist. Most of these scanners can extract only the diagnostic error codes.

Proprietary scanners are needed to extract manufacturer-specific data, which is often the most interesting from a forensic perspective.

Physical Data Acquisition

Physical data acquisition refers to data extraction techniques with direct access to real memory locations. This can be compared with bit-stream images of computer hard drives. Most built-in embedded system storage cannot be accessed with generic connection interfaces like ATA or SCSI for personal computers. The most generic connection interface for embedded system memories is the use of the physical connection points on the memory chips. Several chip package technologies exist (DIP, SIP, TSOP, PGA, LGA, BGA) and for each package technology a large number of variations are used in the number, the spacing and the size of connection points. This diversity makes the use of physical connection points for forensic data acquisition expensive and impractical for generic use. Therefore, several less ideal methods of acquiring physical memory are commonly used.

Software Agents

Software agents are pieces of software running on the exhibit device, assisting with or responsible for the physical data acquisition. These agents run on the normal operating system of the device and use Application Programming Interface (API) calls for low-level memory access or they use a dedicated operating system for data acquisition. For this approach to work, the system needs to be accessible and must allow the execution of custom software. For example, the Symbian OS has a low-level API function called RRawDisk that enables direct disk access. On Windows CE-based devices a similar approach can be used to read data from RAM and flash memory using ActiveSync and remote API calls. For iPhone and iTouch devices a commonly used examination method puts commands on the system partition to remotely login to the device and copy the user partition via Wi-Fi to an examination machine.

Because most software agents run on the normal operating system, precautions are needed for locked files or other processes changing the target data. Software methods can also be used to acquire RAM data but because the agent itself also uses RAM space it potentially overwrites interesting evidence in unallocated RAM areas.

Boot Loaders

As described earlier in this unit boot loaders on embedded systems can contain functionality for direct access to memory locations. The interactive boot loader of a lot of HTC devices running Windows Mobile for example, can be activated by holding down the camera and power button and resetting the device. The boot loader's d2s command can be used on some devices to copy the contents of internal memory onto an inserted multimedia card or onto an examination system via the USB connector. Instead of directly using the built-in boot loader functionality another approach uses the primary boot loader to transfer custom executable code to one of the writable device memories and to start executing that code. Embedded systems often have this flash loader functionality for in-field upgrading of firmware.

Additional boot loader functionality is widely used in the mobile phone world by both manufactures and hackers. Manufacturers use this method for bugging and repair and for in-field firmware upgrades. Apple, for example, distributes firmware upgrades for their iPhone and iPod devices via their iTunes software. iPhones can be switched into different upgrade modes; the most low-level one is used to update the boot loader itself and could also be used for forensic purposes. Hackers also use boot loader functionality to attack device security functions, installing custom firmware or changing normal device behavior.10 Flasher box is the common name for interface devices to connect mobile phones with computer systems for managing additional boot loader functionality. Flasher tools mostly contain a flasher box, control software and a large number of cables to connect different phone models. Great care should be taken when using these tools for forensic examinations. Besides memory acquisition functionality, these tools sometimes have other options that are devastating in a forensic context like writing or erasing memory, changing serial numbers or adding functionality.

Before usage on an exhibit, a flasher tool needs to be tested on a similar device: once thoroughly to check the functionality and preferably before each individual examination to train the examiner in using only the forensically sound options of such a tool.

Forensic tools like XACT and FTS Hex use flash loader techniques for forensic acquisition of data from mobile phones and PDAs.

General Acquisition Procedure

The general and intuitive procedure for acquiring a storage device is to copy one byte from the original storage device (the source) to a destination storage device and repeat the process. This is analogous to copying a document by hand and reading a letter, punctuation mark or space from the original and writing it to the duplicate. While this works, most of us do not copy documents this way because we can remember entire words and it is more efficient to transfer one or more words at a time. Computers do the same thing and copy data from the suspect systems in chunks of data, ranging from 512 bytes to many thousands of bytes.

The chunks of data that are transferred each time are typically a multiple of 512 bytes, because that is the size of most disk sectors. If the acquisition tool encounters an error while reading data from the suspect drive, many of the tools will write zeros to the destination.

Data Acquisition Layers

The general theory of non-volatile data acquisition is to save every byte that we think may contain evidence. Data can be interpreted at different layers; for example, the disk, volume, file and application layers. At each layer of abstraction, data are lost. Therefore, the rule of thumb is to acquire data at the lowest layer that we think there will be evidence. For most cases, an investigator will acquire every sector of a disk, which is what we cover in this chapter. Note that when we save only the contents of each sector, we lose data that data recovery specialists may need.

To show why we typically acquire at the disk level, we will consider some scenarios. Suppose that we acquired a disk at the volume level and we made a copy of every sector in each partition. This would allow us to recover deleted files in each partition, but we would not be able to analyze the sectors that are not allocated to partitions. A disk that has DOS partitions may not use sectors 1 to 62 and they could contain hidden data. If we acquired at the volume level, the hidden data would be lost.

Suppose that we used a backup utiliıy and copied only allocated files. In this case, we would not be able to recover deleted files, we might not have access to all the temporal data and we would not be able to find data that has been hidden inside partition or file system data structures. Sometimes a backup is the only available data and the investigator needs to make the most of it. A scenario where a backup would be critical is in a corporate environment where a server is not responding because its disks were wiped with 0s and then rebooted. The last backups of the system might provide clues about who had access to the system and whether an attacker had compromised it.

For some systems, our rule of thumb about acquiring at the level where we think there will be evidence means that we need to copy only files. Consider an intrusion investigation where there is an *Intrusion Detection System* (IDS) that contains log entries corresponding to the attack. If we do not think that the IDS was compromised, the only evidence on the system is at the file level and we can simply copy the necessary logs and take the appropriate preservation steps. If we think that the IDS was compromised, we should acquire it at the disk level so that we can analyze all the data.

Acquisition Tool Testing

Acquisition is a crucial part of the investigation process and the *National Institute of Standards and Technology* (NIST) in US has conducted tests on common acquisition tools. The *Computer Forensic Tool Testing* (CFTT) project at NIST developed requirements and test cases for disk-imaging tools. The results and specifications can be found on their Web site (http://www.cftt.nist.gov/disk_imaging.htm).

Reading the Source Data

Using the general acquisition theory that was previously described, there are two major parts of the process. First, we need to read data from a source and then we need to write it to the destination. Since we are focussing on the analysis of volume and file system data, we are going to cover the process of acquiring at the disk level (because that is where the volume data structures are located). Let us examine the issues associated with reading a disk and the next major section examines the issues associated with writing to a destination. Let us assume that a typical IA32 system (such as x86/i386) is being used for the acquisition and we will discuss how to access the data, handle errors and reduce the risk of writing data to the suspect drive.

Direct versus BIOS Access

There are two methods in which the data on a disk can be accessed. In one method, the operating system or acquisition software accesses the hard disk directly which requires that the software know the hardware details. In the second method, the operating system or acquisition software accesses the hard disk through the *Basic Input/Output System* (BIOS), which should know all the hardware details.

At a casual glance, there do not seem to be many differences between these methods and using the BIOS seems easier because it takes care of the hardware details. Unfortunately, it is not that straightforward when it comes to doing an investigation.

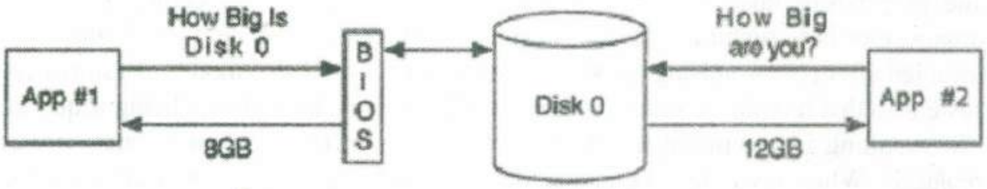


Fig. 1: Two applications are trying to determine the size of a disk. The BIOS is not properly configured and says that the 12GB disk is only 8GB

When the BIOS is used, there is a risk that it may return incorrect information about the disk. If the BIOS thinks that a disk is 8GB, but the disk is really 12GB, the INT13h functions will give you access to only the first 8GB. Therefore, if you are doing an acquisition of the disk, you will not copy the final 4GB. We can see this in Fig. 1, where two applications are trying to identify the size of a disk using different methods.

This scenario can happen in a couple of different ways. One case is when the BIOS is configured for a specific hard disk geometry that is different from the one installed. In another case, an acquisition tool uses a legacy method of requesting the size of the disk.

There are two ways that an application can ask the BIOS for a disk size. One is through the original INT13h function that suffers from the 8GB limit and returns the size using the disk's geometry in CHS format. The second method is to use an extended INT13h function that returns the size in LBA format. The CFTT group at NIST had a 2GB disk and a computer where two different sizes were returned from the INT13h and the extended INT13h functions.

The extended INT13h result was correct, but the legacy INT13h result was too small. Occasionally, an e-mail is sent to one to the digital forensic e-mail lists from someone who acquired a disk using two different tools and got different sized images. The reason is usually because one of the tools used the BIOS and the other did not. Make sure that you know how your acquisition tools access the disk and if the tool uses the BIOS, make sure it reports the full disk before you acquire the disk. The BIOS adds one more location where an error can be introduced into the final image and it should be avoided if better alternatives exist.

Dead vs Live Acquisition

An investigator has the choice of performing a dead or a live acquisition of data. A *dead acquisition* occurs when the data from a suspect system is being copied without the assistance of the suspect operating system. Historically, the term dead refers to the state of only the operating system, so a dead acquisition can use the hardware from the suspect system as long as it is booted from a trusted CD or floppy. A live acquisition is one where the suspect operating system is still running and being used to copy data.

The risk of conducting a *live acquisition* is that the attacker has modified the operating system or other software to provide false data during the acquisition. To provide an analogy to the physical world, imagine the police arriving at a crime scene where there are several people and it is unknown whether any were involved in the crime. A little while later, the police are looking for a certain object and they ask one of these unknown people to go into one of the rooms and look for the object. The person comes back to the officer and says that he could not find the object, but should the officer trust him? Maybe this person was involved in the crime and the object was in the room, but he destroyed it when he was sent in to look for it.

Attackers frequently install tools called rootkits into systems that they compromise and they return false information to a user. The rootkits hide certain files in a directory or hide running processes. Typically, the attackers hide the files that they installed after compromising the system. An attacker could also modify the operating system so that it replaces data in certain sectors of the disk while it is being acquired. The resulting image might not have any evidence of the incident because it was replaced. When possible, live acquisition should be avoided so that all evidence can be reliably collected.

It is common for an investigator to boot a suspect system using a trusted DOS floppy or Linux CD that has been configured to not mount drives or modify any data. Technically, it is possible for the suspect to have modified their hardware so that it returns false data even with a trusted operating system, but that is much less likely than the operating system being tampered with.

Error Handling

When an acquisition tool is reading data from a disk, it needs to be capable of handling errors. The errors could be caused by a physical problem where the entire drive no longer works or the errors could be in a limited number of sectors. If only a limited number of sectors is damaged, a normal acquisition can occur provided that the acquisition tool properly handles the errors.

The generally accepted behaviour for dealing with a bad sector is to log its address and write 0s for the data that could not be read. Writing 0s keeps the other data in its correct location. If the sector were ignored instead of writing 0s, the resulting copy would be too small and most analysis tools would not work. Fig. 2 shows a series of values that are being acquired.

Three of the values have errors and cannot be read, so 0s are written to the copy.

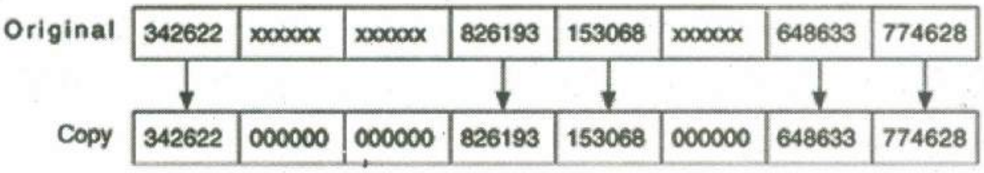


Fig. 2: The original has three errors in it that have been replaced by 0s

Host Protected Area

When acquiring data from an ATA disk, you should pay attention to the *Host Protected Area* (HPA) of the disk because it could contain hidden data. Unless an acquisition tool looks for an HPA, it will not be acquired. A tool can detect an HPA by comparing the output of two ATA commands. The READ_NATIVE_MAX_ADDRESS command gives the total number of sectors on the disk and the IDENTIFY_DEVICE returns the total number of sectors that a user can access. If an HPA exists, these two values will be different. If you do not have access to a tool that will execute the necessary ATA commands, you may have to compare the number of sectors that are copied during an acquisition with the number of sectors that is documented on the label of the disk. Many of the current acquisition tools on the market will detect an HPA and there are also specialized tools such as BXDR (http://www.sandersonforensics.co.uk/ BXDR.htm), diskstat in The Sleuth Kit, DRIVEID by MyKey Technology (http://www.mykeytech.com) and hpa by (http://www.dmares.com/ maresware/gk.htm#HPA). If you encounter a disk with an HPA and you want to gain access to the hidden data, you will need to change the disk configuration. An HPA is removed by setting the maximum user addressable sector to be the maximum sector on the disk. This can be done using the volatility bit such that the configuration change will be lost when the hard disk is powered off.

The process of removing an HPA involves changing the disk configuration. There is an extremely rare possibility that the disk controller or acquisition tool has not properly implemented HPA changes and data could be lost. Therefore, you might consider imaging the disk with the HPA before you remove it. If the removal process causes any damage, you still have the original image to analyze. We will see an example of a disk with an HPA in the dd case study later in this unit. If you need to remove an HPA, it should be documented in your notes.

Device Configuration Overlay

When acquiring data from a newer ATA disk, you should look for a *Device Configuration Overlay* (DCO) which could cause the disk to look smaller than it really is. A DCO is similar to an HPA and they can both exist at the same time.

A DCO is detected by comparing the output of two ATA commands. The READ_NATIVE_MAX_ADDRESS command returns the maximum sector of the disk that normal ATA commands have access to and the DEVICE_CONFIGURATION_IDENTIFY command returns the actual physical number of sectors. If these are different, a DCO exists and needs to be removed if all data are going to be acquired.

To remove a DCO, the disk configuration must be changed using the DEVICE_CONFIGURATION_SET or DEVICE_CONFIGURATION_RESET commands.

Both of these changes are permanent and will not be revoked at the next reset as is possible with HPA. Currently, there are few tools that detect and remove DCO. The Image MASSter Solo 2 from ICS (http://www.icsforensic.com) will copy the sectors hidden by a DCO.

As with HPA, it is safest to make a copy of the drive with the DCO in place and then remove it and make a second copy. When you remove a DCO, be sure to document the process. Also test whether your hardware write blockers allow the DCO to be removed.

Hardware Write Blockers

One of the investigation guidelines is to modify the original data as little as possible. There are many acquisition techniques that do not modify any of the original data but mistakes can happen. Further, there are also some acquisition techniques that can modify the original data and we may want to prevent that.

A hardware write protector is a device that sits in the connection between a computer and a storage device. It monitors the commands that are being issued and prevents the computer from writing data to the storage device. Write blockers support many storage interfaces such as ATA, SCSI, Firewire (IEEE 1394), USB or Serial ATA. These devices are especially important when using an operating system that could mount the original disk, such as Microsoft Windows.

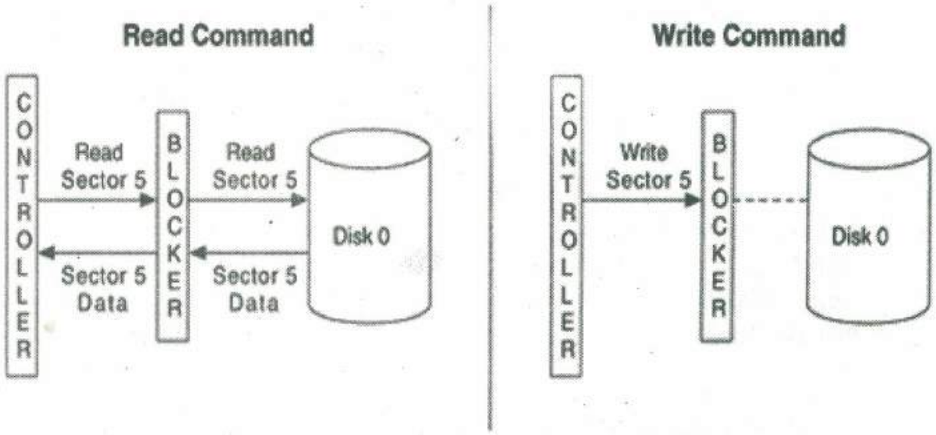


Fig. 3: The read request for sector 5 is passed through the write blocker, but the write command for the same sector is blocked before it reaches the disk

A disk should not perform any actions until its command register is written to. So in theory, the most basic type of ATA hardware write blocker is a device that prevents the controller from writing any values to the command register that could cause data to be written to or erased from the disk. However, such a device might allow the controller to write data into other registers. This is analogous to being able to load a gun but not being able to pull the trigger. We can see in Fig. 3 that read commands are passed to the disk but write commands are not.

The NoWrite device by MyKey Technologies has a more advanced design and works as a state-based proxy between the controller and hard disk. It does not send any data or command to the hard disk until it knows that it is a safe command. Therefore, the command arguments are not written to the registers until the NoWrite device knows what command they are for. This makes the data transfers slower but it is easier to show that no dangerous commands were written. Using the previous gun analogy, this process checks each bullet and allows only blanks to be loaded.

To remove an HPA or DCO, commands are sent to the disk. These commands modify the device and should be stopped by hardware write blockers. The NoWrite device makes an exception and allows the SET_MAX command to be executed if the volatile bit is set such that the change is not permanent. All other SET_MAX and DEVICE_CONFIGURATION commands are blocked. Other write blockers may choose to allow all these commands to pass and others may block them all. At the time of this writing, there is little documentation on which commands are being blocked, so you should check with your vendor and conduct your own tests.

Like all investigation tools, testing of hardware write blockers is important and the CFTT group at NIST has published a specification for hardware write blockers (http://www.cftt.nist.gov/hardware_write_block.htm). The specification classifies the ATA commands as non-modifying, modifying and configuration. The specification states that modifying commands must be blocked and optionally return success or failure.

Software Write Blockers

In addition to hardware write blockers, there are also software write blockers. At one point most digital forensic tools were DOS-based and used the INT13h method to access a disk. Software write blockers were frequently used to prevent the disk from being modified during the acquisition and examination. We will now examine how they work and what their limitations are.

The software write blockers work by modifying the interrupt table, which is used to locate the code for a given BIOS service. The interrupt table has an entry for every service that the BIOS provides and each entry contains the address where the service code can be found. For example, the entry for INT13h will point to the code that will write or read data to or from the disk.

A software write blocker modifies the interrupt table so that the table entry for interrupt 0x13 contains the address of the write blocker code instead of the BIOS code. When the operating system calls INT13h, the write blocker code is executed and examines which function is being requested. Fig. 4 shows an example where the software write block has been installed and blocks a write command. A write blocker allows a non-write function to execute by passing the request directly to the original INT13h BIOS code.

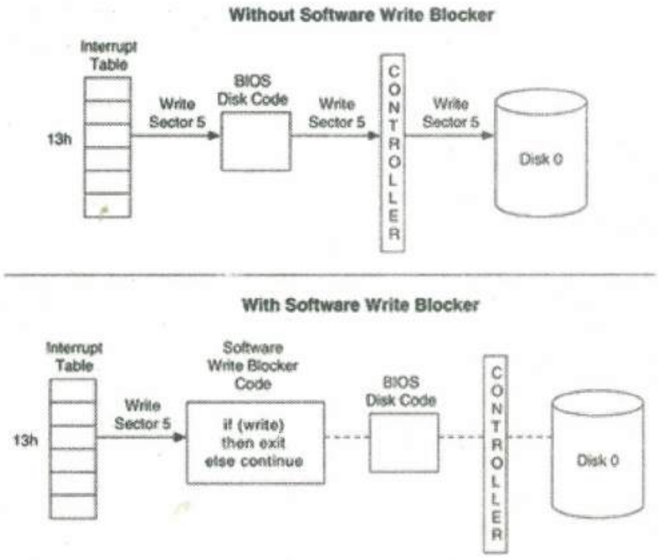


Fig. 4: A BIOS interrupt table without a write block installed and with a software write block installed that prevents writes from being executed

Software write blockers are not as effective as hardware blockers because software can still bypass the BIOS and write data directly do the controller and the BIOS can still write data to the disk because it has direct access to the controller. In general, if you want to control access to a device, you should place the controls as close to the device as possible. The hardware write blockers are as close to the hard disk as possible on the ribbon cable.

The CFTT group at NIST has developed requirements and has tested software write block devices. The details can be found on their Web site (*http://www.cftt.nist.gov/software_write_block.htm*).

Writing the Output Data

After we read the data from the source disk, we need to write them somewhere. We will now discuss where to save data and the various formats in which data can be saved.

- **Destination Location**

When we save the data, we can write them either directly to a disk or to a file.

Before there was specialized analysis software, an investigator either booted the suspect system or mounted the disks in her analysis system. She acquired the drive by copying the data directly to another disk. In other words, sector 0 of the source disk was identical to sector 0 of the destination disk. The resulting disk was frequently called a duplicate copy or a cloned copy. This method can cause problems when the destination disk is bigger than the source disk because it can be difficult to tell exactly where the copy ends. When acquiring directly to disk, it is recommended that the disk be wiped with zeros before acquisition so that unrelated data, possibly from a previous investigation are not confused with data from the suspect system. A second problem with acquiring to disk is that some operating systems, such as Microsoft Windows, will try to mount any disk and the copy could be mounted by the acquisition system and have its data changed. You also can run into difficulties if the original and destination disks have different geometries because some of the data structures rely on the geometry to describe locations.

Currently, the most common output location is to save the data to a file on a hard disk or CDROM. With a file, it is easy to know the boundaries of the data and operating systems will not try to mount it automatically. The file is frequently called an *image* or a duplicate image.

Many tools will allow you to break an image file into smaller pieces so that they fit onto CDs or DVDs. Some investigators will wipe the disks that store image files so that they can more easily testify that there could not have been any contamination from a previous case.

- **Image File Format**

If we save the data to a file, we have a choice of in what format the image will be. A raw *image* contains only the data from the source device and it is easy to compare the image with the source data. An *embedded image* contains data from the source device and additional descriptive data about the acquisition such as hash values, dates and times. Some tools will create a raw image and save the additional descriptive data to a separate file. Recall that hash values, such as CRC, MD5 and SHA-1 are used to show the integrity of data. Examples of image formats can be seen in Fig. 5.

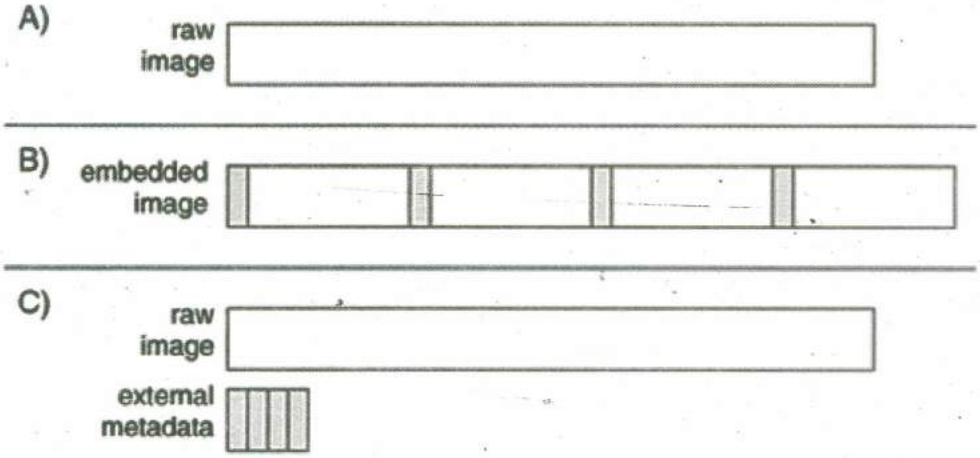


Fig. 5: Examples of (A) a raw image, (B) an embedded image with meta data interleaved in the raw data and (C) an image with the data stored in a raw format and the meta data stored in a second file

In current implementations of acquisition tools, many of the embedded image formats are proprietary such as those from Guidance Software's EnCase and NTI's SafeBack and some are documented such as the format used by Technology Pathway's ProDiscover. Most analysis tools import a raw image; therefore, it is the most flexible format. The SMART tool from ASR Data and the dcfldd/dccidd tools acquire data in a raw format and have an external file with additional data.

Compressing the Image File

When we write the data to a file, we may have the option to compress the file so that it takes up less storage space. Compression works by storing repetitive data more efficiently. For example, if the data have 10,000 consecutive 1s, a compressed format may be able to describe that in a few hundred bits instead of 10,000 bits. If the data are random, there will be little repetition and compression will not be as effective. If you compress data that have already been compressed, the result will not be much smaller. For example, JPEG images have compression in them and their size does not change if they are compressed.

When an image is compressed, any analysis tool you use it with must support the compression type. This is similar to using an image format in which data are embedded. Most general types of compression require you to decompress the entire file before it can be used.

Examples of this include the Winzip tools for Windows and the gzip tools in Unix. Special compression algorithms will allow you to uncompress a small part of the compressed file and those are the ones that should be used by acquisition tools so that you do not have to uncompress the entire image.

The benefit of compression is that you can acquire a storage device to a smaller image file, although the actual amount of data saved depends on the acquired data. The negatives of compression are as follows:

- You might be limited by the number of analysis tools that support the format.
- Acquisition might take longer because the software must perform the compression.
- Analysis might be slower because the analysis tool must decompress the image when it reads data from it.

Network-based Acquisition

The basic acquisition theory also allows you to create an image file on a remote computer by using a network. In this case, data are read from the source disk, transmitted to the destination host via a network and written to a file. This method of acquisition is convenient if you cannot get access to the suspect disk or do not have the correct adaptors or interface for the suspect disk. Many current tools support network-based acquisition of dead and live systems.

Some offer encryption to provide confidentiality on the network. Compression can be useful for the transmission to reduce the amount of data sent over a slow network.

Integrity Hashes

One of the core concepts of an investigation is to calculate hash values for evidence so that we can later verify the integrity of the data. Some acquisition tools will calculate a hash while the data are being copied and others require a separate tool. In many cases, the hashes are stored in either an embedded image or an external file with a raw image. Having the hashes embedded in the image does not provide any additional security or integrity.

It is important to note what the hashes actually do for you. Any hash that is stored with the image will not ensure that someone has not modified the data. After all, if someone modifies the image, they can also recalculate the hashes, even if they are embedded in the format. A program could be easily written to do this. To prove the integrity of an image file using a digital hash, you will need to use a cryptographic signature and a trusted time source. This requires a lot of overhead; therefore, a much easier method is to write the hash value down in your notebook. Then someone will have to modify the image, recalculate the hash and rewrite your notebook.

While hashes are important to later prove the integrity of an image, they can also be used to show the accuracy of an acquisition process and that the acquisition process did not modify the original disk. By calculating the hash of the disk before it is acquired and comparing that value with the hash of a raw image, you can show that the raw image contains the same data that were on the original disk. Ideally, the original hash should be calculated with a tool that is independent of the acquisition tools so that any errors are not applied to both the control case and the actual image.

Note that the previous hashing process reads only the data that are available to the tool. If hardware or software problems prevent you from accessing all bytes in a disk, the hash of the disk can equal the hash of the image file even though the image file does not represent all data on the disk. For example, if the tool can read only the first 8GB of a 12GB disk, the tool will compute the hash of the first 8GB of the disk, copy the first 8GB of data and then compute the hash of the 8GB image file.

Another consideration for hashes is how often they are calculated. Hashes are most commonly used to identify when a value in a chunk of data has been changed. If the hash shows that a value in the chunk has been changed, the chunk must not be used. Calculating hashes of smaller chunks can minimize the impact of an integrity failure. If any chunk of data fails an integrity test, then it will not be used but the rest of the image will.

A Case Study Using dd

To illustrate the acquisition process, I will describe how we can do an acquisition with the dd tool. dd is one of the most simple and flexible acquisition tools but it is command line-based and can be more complex to learn than other tools because each feature and option can be specified. dd comes with many of the UNIX versions and is available for Windows. For this example we will focus on running it in Linux.

At its core, dd copies a chunk of data from one file and writes it to another. It does not care what type of data it is copying and does not know about file systems or disks, only files. dd reads data from the input source in block-sized chunks and the default block size is 512 bytes. It reads data from an input source, which is specified with the if= flag. If the if= flag is not given, it takes the standard input as the input source, which is typically the keyboard.dd writes the data to an output file, which is specified with the of= flag. If that is not given, the data are written to standard output, which is usually the display. As an example, to copy the contents of file1.dat, which is 1024 bytes, to file2.dat in 512-byte blocks, we use

```
# dd if=file1.dat of=file2.dat bs=512
2+0 records in
2+0 records out
```

The final two lines show that two complete blocks were read from file1.dat and two complete blocks were written to file2.dat. If a full block was not used during the last read and write, the final two lines would have ended with '+1' instead of '+0.' For example, if file1.dat were 1500 bytes instead of 1024 bytes, the following would have been seen:

```
# dd if=file1.dat of=file2.dat bs=512
2+1 records in
2+1 records out
```

Note that the resulting file will be the full 1500 bytes. dd will try to write in block-sized chunks, but if there is not enough data, it will only copy what it has.

Input Sources

In Linux, there is a device for each storage device and partition and it can be used as the input file. For example, the master ATA disk on the first channel is/dev/hda and we can use that device name with the if= flag to tell dd to copy data from the disk to a file.

Microsoft Windows does not have an actual device file for the hard disks but you can use the \\.\ syntax to reference a disk, \\.\PhysicalDrive0, for example.

The default block size is 512 bytes, but we can specify anything we want using the bs= flag.

We can copy 1 byte at a time or we can copy 1GB at a time. Any value will work, but some values will give you better performance than others. Most disks read a minimum of 512 bytes at a time and can easily read more at the same time. Using a value that is too small is wasteful because the disk will need to be frequently read and time will be wasted in the copying process. If you choose a value that is too large, you will waste time filling up the buffer in dd before the copy is performed. I have found that values in the 2KB to 8KB range work well.

Linux accesses the hard disk directly and does not use the BIOS, so we do not risk getting incorrect data from the BIOS about the size of the disk. That also means that there are not software write blockers for Linux, but you can use a hardware device if you want.

HPA

As previously stated, dd knows about only files and therefore does not know anything about ATA HPAs. There are several methods of detecting an ATA HPA in

Linux and we will cover those here.

The scenario for this example is a 57GB disk with 120,103,200 sectors. We have placed the string "here i am" in sector 15,000, as seen here:

```
# dd if=/dev/hdb bs=512 skip=15000 count=1 | xxd
1+0 records in
1+0 records out
0000000: 6865 7265 2069 2061 6d0a 0000 0000 0000 here i am.......
```

Next, We created an HPA in the final 120,091,200 sectors. In other words, there are only 12,000 sectors that the OS or an application can access. We can see this because we can no longer see the string in sector 15,000:

```
# dd if=/dev/hdb bs=512 skip=15000 count=1 | xxd
0+0 records in
0+0 records out
```

No records were copied because it could not read the data. There are several ways of detecting an HPA in Linux. Newer versions of Linux display a message in the dmesg log.

Note that this log has a limited size and entries will be overwritten if there is an application that is writing a lot of warning or error messages. Its output for our disk is as follows:

```
# dmesg | less
[REMOVED]
hdb: Host Protected Area detected.
current capacity is 12000 sectors (6 MB)
native capacity is 120103200 sectors (61492 MB)
```

Not all versions of Linux will display this message, though. Another method of detecting an HPA is using the hdparm tool that comes with Linux. It displays details about a hard disk and we need to use the -I flag to obtain the total number of sectors. We will compare this value with the value written on the disk or from the vendor's Web site. This output will also tell us if the drive supports HPA, which older disks do not.

```
# hdparm -I /dev/hdb
[REMOVED]
CHS current addressable sectors: 11088
LBA user addressable sectors: 12000
LBA48 user addressable sectors: 12000
[REMOVED]
Commands/features:
Enabled Supported:
* Host Protected Area feature set
```

In this case, the label of my drive says that it has 120,103,200 sectors; therefore, many sectors are not addressable. Lastly, you can use the diskstat tool from 'The Sleuth Kit'. It displays the maximum native address and the maximum user address.

```
# diskstat /dev/hdb
Maximum Disk Sector: 120103199
Maximum User Sector: 11999
** HPA Detected (Sectors 12000 - 120103199) **
```

To access the data, we need to reset the maximum address. One tool that allows us to do this is setmax (http://www.win.tue.nl/~aeb/linux/setmax.c). We will run this tool and set the maximum number of sectors in the drive, which is 120,103,200 in this example. This tool modifies the configuration of your drive and extreme care must be taken (which means you should also take good notes while you doing it). Also note that this tool sets the maximum address as non-volatile, so the change is permanent. If you are going to use a tool like this, test it on other drives before you use it on a disk that may contain evidence.

```
# setmax --max 120103200 /dev/hdb
```

After resetting the maximum address, you can use dd to acquire the full disk. Record the location of the HPA so that you can return the disk to its original state and so that you will know where it started when you analyze the data.

Output Destinations

The output from dd can be either a new file or another storage device. For example, the two following examples are performed in a Linux environment. The first copies the master ATA disk on the primary channel to a file and the second example copies the master ATA disk on the primary channel to the slave ATA disk on the second channel.

```
# dd if=/dev/hda of=/mnt/hda.dd bs=2k
# dd if=/dev/hda of=/dev/hdd bs=2k
```

If you do not specify the output file, the data will be written to the display. This can be useful to calculate the MD5 hash, to extract the ASCII strings or to send the data to a remote system using a network. For example, to hash a disk, we could use the md5sum command that comes with Linux:

```
# dd if=/dev/hda bs=2k | md5sum
```

We can also send data to a server using the netcat (http://www.atstake.com/research/tools/) or cryptcat (http: //sf.net/projects/cryptcat) tools. With netcat, a trusted evidence server at IP address 10.0.0.1 would run the following to open a network port on port 7000 and save incoming data to a file:

```
# nc -l -p 7000 > disk.dd
```

The system with the source disk in it would be booted from a trusted Linux CD and dd would be executed with the data piped to netcat, which would send data to the server at 10.0.0.1 at port 7000. The connection would close after three seconds of no activity:

```
# dd if=/dev/hda bs=2k | nc -w 3 10.0.0.1 7000
```

Error Handling

If dd encounters an error while reading the input file, the default action is to stop copying data. If you specify the conv=noerror flag, dd will report the error and not stop.

Unfortunately, this method skips the blocks with bad data and the image will be the wrong size and the data will be at the wrong addresses.

To maintain the addresses in the image, the sync flag should be given. The sync flag forces dd to write data in block-sized chunks and if there is not enough data for a full block, it pads the data with 0s. Therefore, when an error is encountered, the invalid data will be replaced with 0s. The downside of always using these flag options is that the resulting image will always be a multiple of the block size which may not be the actual size of the original storage device. For example, if we choose a block size of 4,096 bytes, but the size of my (really small) disk is 6,144 bytes, the resulting image file will be 8,192 bytes instead of 6,144 bytes.

An example of using the error handling options is

```
# dd if=/dev/hda of=hda.dd bs=2k conv=noerror,sync
```

Cryptographic Hashes

Normally, when you want a cryptographic hash of a file and you are using dd, you must use another utility, such as md5sum. The cryptographic hash of an image is calculated to later prove an image's integrity.

The same basic flags that we saw for dd also apply to tools used for cryptographic hashes. The hashwindow= flag allows you to specify how frequently a hash should be calculated. If the value is 0, only one hash is calculated of the entire file. If a non-zero byte size is given, a hash is calculated at each point in the file and a final hash is calculated. The hashes can be saved to an output file using the hashlog= flag. dcfldd computes only the MD5 hash, but dccidd has the hash= flag that allows you to specify which hashes should be calculated. By default, the MD5 and SHA-1 are calculated in parallel, but you can specify 'md5,' 'sha1,' or 'sha256.'

For example, if you wanted to image a Linux hard disk and calculate hashes for every 1MB you would use the following:

```
# dcfldd if=/dev/hda of=/mnt/hda.dd bs=2k hashwindow=1M hashlog=/mnt/hda.hashes
```

The hashlog has the following format:

```
0 - 1048576: 970653da48f047f3511196c8a230f64c
1048576 - 2097152: b6d81b360a5672d80c27430f39153e2c
...
103809024 - 104857600: b6d81b360a5672d80c27430f39153e2c
104857600 - 105906176: 94a171ec3908687fd1f456087576715b
Total: 28d34393f36958f8fc822ae3980f37c3
```

Each line starts with the range of bytes that the hash applies to and ends with the hash value.

The last value is the hash for the entire image. The log file for dccidd is slightly different because it includes the SHA-1 hash and the range field is padded with 0s. Here is the output when the hashwindow was set to 512 bytes (the SHA-1 and MD5 hashes are typically on the same line):

```
000000 - 000511: 5dbd121cad07429ed176f7fac6a133d6
09cae0d9f2a387bb3436a15aa514b16f9378efbf
000512 - 001023: 91cf74d0ee95d4b60197e4c0ca710be4
0f71d8729ad39ae094e235ab31a9855b2a5a5900
001024 - 001535: 8a0a10f43b2bcd9e1385628f7e3a8693
641b9b828e41cd391f93b5f3bfaf2d1d7b393da0
[REMOVED]
```

The Windows version dd also has built-in MD5 features. With this tool, supplying the -md5sum flag calculates the MD5 hash for the file. It can also save the hash to a file using the -md5out flag.

Digital forensics has existed for as long as computers have stored data that could be used as evidence. For many years, digital forensics was performed primarily by government agencies, but has become common in the commercial sector over the past several years.

Digital forensics has three major phases:

- **Electronic Acquisition**
- **Data Analysis**
- **Information Presentation**

The ***Electronic Acquisition Phase*** saves the state of a digital system so that it can be later analyzed. This is analogous to taking photographs, fingerprints, blood samples or tire patterns from a crime scene. As in the physical world, it is unknown which data will be used as digital evidence so the goal of this phase is to save all digital values. At a minimum, the allocated and unallocated areas of a hard disk are copied, which is commonly called an image.

The specific challenges and requirements associated with complex forensic examinations and in the acquisition, preservation and analysis of digital information need to be faced with lot of care. The tools used in the acquisition phase to copy data from the suspect storage device to a trusted device do not modify the suspect device and copy all data.

Electronic Acquisition is in short,

- Evidence collection and preservation;
- Analysis of data modification, access and creation; and
- Intelligent and robust techniques result in faster searches and recovery of information.

The ***Data Analysis Phase*** uses the acquired data and examines it to identify pieces of evidence. There are three major categories of evidence we are looking for:

- **Inculpatory Evidence:** That which supports a given theory
- **Exculpatory Evidence:** That which contradicts a given theory
- **Evidence of tampering:** That which cannot be associated to any theory but shows that the system was tampered with to avoid identification.

This phase includes examining file and directory contents and recovering deleted content. Our patent pending data analysis technique enables us to search for relevant information, develop insights and analyze the results very quickly. Our technology can perform analysis on digital content from multiple sources in various formats, structured or unstructured. Our techniques allow legal experts to spend more time developing their case instead of searching for information.

Regardless of the investigation setting (corporate or government), the steps involved during acquisition and analysis phases are similar because they are dominated by technical issues, rather than legal processes.

Data Analysis and Recovery are therefore,

- Secure data recovery and analysis;
- Organize data by categories without requiring prior information about dataset;
- Recover data in a fraction of the time when compared to traditional keyword searches; and
- Discover hidden patterns, relationships and trends.

While acquiring evidence one often runs some very interesting situations with our clients. Often it is observed that former employees break back in to system to cause havoc to conducting covert data acquisition in the middle of the night of current employees suspected of wrongdoing. Often organizations are left to balance the need to get to information and gathering that information in a manner that doesn't trample all over the effectiveness of the data.

As an example, say key individuals are being laid off from the organization and they have information on their laptop that you need. One approach would be just to have technical support team come in and copy off the data via Windows copy or use Ghost to make a copy of hard drive. These two options will get the data copied, but at what cost?

- Will you have access to deleted data?
- What if the data collected reveals criminal behavior or behavior that warrants litigation – do you have the data collected in a manner that can be used in court?
- Have you taken the steps to be able to show a clear picture of what occurred on the computer?

The ***Information Presentation Phase*** though is based entirely on policy and law, which are different for each setting. In this phase we present the conclusions and corresponding evidence from the investigation in our patent pending proprietary framework.

So what do you do to protect yourself and the company? Here's a list of the some of the top Best Practices to digital forensics:

1) Document everything.

2) Never mishandle data.

3) Never work on the original data.

4) Never trust the custodian's software/hardware.

5) Maintain chain-of-custody throughout the process.

6) Only use courtroom admissible and licensed tools.

7) Be sure to be fully trained in the use of digital forensic tools.

8) Don't forget other devices such as PDAs, Blackberries, iPhones etc.

9) Use write-blocking hardware when doing physical acquisitions.

The hard disk is where most of the evidence is found in current investigations, which will likely be the case for many years to come, at least until all hard disks are encrypted.

Acquisitions are very important in the investigation process because if they are not performed correctly, data may not exist for the investigation. This section has outlined the general theory of acquisitions and given a case studying using *dd*. dd is a fairly simple tool, but it is command line and can cause confusion because it has many options.

Check Your Progress 2

Note: a) Space is given below for writing your answers.

b) Compare your answers with the one given at the end of this Unit.

1) Define freezing the scene.

....

....

....

....

2) What is honeypotting?

....

....

....

....

2.4 LET US SUM UP

This Unit has attempted to present the concept of data acquisition and information gathering along with different scientific methods followed in investigation. It also describes various methods of collection and its steps and discussed standard procedures for dealing with digital evidence, as well as specific evidence location and examination techniques such as recovering supposedly deleted files, finding steganographic data, locating "forgotten" data and decrypting encrypted data. Procedures for documenting digital evidence are also outlined and examination some of the legal issues involved in evidence collection and handling will also be done.

2.5 CHECK YOUR PROGRESS: THE KEY

Check Your Progress 1

General Procedure

There is a general four-step procedure to follow for collecting and analyzing evidence.

- **Identification of Evidence**

  One must be able to distinguish between evidence and junk data. For this purpose, one should know what the data is, where it is located and how it is stored. Once this is done, one will be able to work out the best way to retrieve and store any evidence.

- **Preservation of Evidence**

  The evidence find must be preserved as close as possible to its original state. Any changes made during this phase must be documented and justified.

- **Analysis of Evidence**

  The stored evidence must then be analyzed to extract the relevant information and recreate the chain of events. Analysis requires in-depth knowledge of what you are looking for and how to get it. Always be sure that the person or people who are analyzing the evidence are fully qualified to do so.

- **Presentation of Evidence**

  Communicating the meaning of your evidence is vitally important-otherwise you can't do anything with it. The manner of presentation is important and it must be understandable by a layman to be effective. It should remain technically correct and credible. A good presenter can help in this respect.

Check Your Progress 2

1) **Freezing the Scene**

   It involves taking a snapshot of the system in its compromised state. The necessary authorities should be notified (the police and your incident response and legal teams) but you shouldn't go out and tell the world just yet. You should then start to collect whatever data is important onto removable non-volatile media in a standard format. Make sure the programs and utilities used to collect the data are also collected onto the same media as the data. All data collected should have a cryptographic message digest created and those digests should be compared to the originals for verification.

2) **Honeypotting**

   It is the process of creating a replica system and luring the attacker into it for further monitoring. A related method (sandboxing) involves limiting what the attacker can do while still on the compromised system, so he can be monitored without (much) further damage. The placement of misleading information and the attacker's response to it is a good method for determining the attacker's motives. You must make sure that any data on the system related to the attacker's detection and actions is either removed or encrypted; otherwise they can cover their tracks by destroying it. Honeypotting and sandboxing are extremely resource intensive, so they may be infeasible to perform.

2.6 SUGGESTED READINGS

- MyKey Technology, Inc. "Technical White Paper: No Write Design Notes." 2003.https://mykeytech.com/nowritepaper1.html.
- Skoudis, Ed and Lenny Zeltser. *Malware: Fighting Malicous Code*. Upper Saddle River: Prentice Hall, 2004.
- Technology Pathways, Inc. "ProDiscover Image File Forma." 2003. https://wwwtechpathways.com/uploads/ProDiscoverImageFileFormatv4.pdf.
- U.S. Department of Justice. "Test Results for Disc Imaging Tools: SafeBack 2.18." *NCJ* 200032, June 2003. https://www.ncjrs.org/pdffiles1/nij/20032.pdf.

UNIT 3 FORENSIC EXAMINATION OF SYSTEMS

Structure

3.0 INTRODUCTION

The field of computer forensics involves identifying, extracting, documenting and preserving information that is stored or transmitted in electronic or magnetic form (i.e. digital evidence). Like fingerprints, digital evidence can be visible (such as files stored on disk that can be accessed via the normal directory structure using standard file management tools such as Windows Explorer) or it can be latent (not readily visible or accessible, requiring some sort of processing – via special software or techniques-to locate and identify it). An important aspect of computer forensics involves finding and evaluating this "hidden data" for its evidentiary value.

Computer forensics standards have been developed that apply to the collection and preservation of digital evidence, which differs in nature from most other types of evidence and thus requires different methods of handling. Following procedures that are proper, accepted and in some cases, prescribed by law in dealing with evidence is vital to the successful prosecution of a cybercrime case. The proper handling of these procedures comes into play at two different points in a trial:

- If the evidence is not collected and handled according to the proper standards, the judge may deem the evidence inadmissible when it is presented (usually based on the opposing attorney's motion to suppress) and the jury members will never get a chance to evaluate it or consider it in making their decision.
- If the evidence is admitted, the opposing attorney will attack its credibility during questioning of the witnesses who testify regarding it. Such an attack can create doubt in jury members' minds that will cause them to disregard the evidence in making their decision-and perhaps even taint the credibility of the entire case.

The entire investigation will be of little value if the evidence that shows the defendant's guilt is not allowed into the trial or if the jury gives it no weight.

Thus proper handling of evidence is one of the most important issues facing all criminal investigators and because of the intangible nature of digital evidence, cyber crime investigators in particular. Because this is such an important topic-not only for investigators but for prosecutors, judges and justice system professionals involved in cybercrime cases-many organizations and publications are devoted solely to issues concerning digital evidence. The International Organization of Computer Evidence (IOCE) was established in 1995 to provide a forum for law enforcement agencies across the world to exchange information about computer forensics issues.

3.1 OBJECTIVES

After studying this unit, you should be able to:

- understand various search techniques;
- explain different keyword research tools;
- understand and explain about data recovery;
- elucidate various tools used in data recovery; and
- elucidate various tools used in forensic examintion of systems.

3.2 SEARCH TECHNIQUES

This group of techniques searches collects information to answer the question whether objects of given type, such as hacking tools or pictures of certain kind are present in the collected information. According to the level of search automation, this unit classifies techniques into manual browsing and automated searches. Automated searches include keyword search, regular expression search, approximate matching search, custom searches and search of modifications.

3.2.1 Manual Browsing

Manual browsing means that the forensic analyst browses collected information and singles out objects of desired type. The only tool used in manual browsing is a viewer of some sort. It takes a data object, such network packet, decodes the object and presents the result in a human-comprehensible form.

Manual browsing is slow. Most investigations collect large quantities of digital information, which makes manual browsing of the entire collected information unacceptably time consuming.

3.2.2 Keyword Search

Keyword search is automatic search of digital information for data objects containing specified key words. It is the earliest and the most widespread technique for speeding up manual browsing. The output of keyword search is the list of found data objects (or locations thereof). Keywords are rarely sufficient to specify the desired type of data objects precisely. As a result, the output of keyword search can contain false positives, objects that do not belong to the desired type even though they contain specified keywords. To remove false positives, the forensic scientist has to manually browse the data objects found by the keyword search.

Another problem of keyword search is false negatives. They are objects of desired type that are missed by the search. False negatives occur if the search utility cannot properly interpret the data objects being searched. It may be caused by encryption, compression or inability of the search utility to interpret novel data format.

A strategy for choosing key words and phrases prescribes (1) to choose words and phrases highly specific to the objects of the desired type such as specific names, addresses, bank account numbers etc. and (2) to specify all possible variations of these words.

3.2.3 Regular Expression Search

Regular expression search is an extension of keyword search. Regular expressions provide a more flexible language for describing objects of interest than keywords. Apart from formulating keyword searches, regular expressions can be used to specify searches for Internet e-mail addresses, etc. Forensic utility EnCase performs regular expression searches.

Regular expression searches suffer from false positives and false negatives just like keyword searches because not all types of data can be adequately defined using regular expressions.

3.2.4 Approximate Matching Search

Approximate matching search is a development of regular expression search. It uses matching algorithm that permits character mismatches when searching for keyword or pattern. The user must specify the degree of mismatches allowed.

Approximate matching can detect misspelled words but mismatches also increase the number of false positives. One of the utilities used for approximate search is "agrep".

3.2.5 Custom Search

The expressiveness of regular expressions is limited. Searches for objects satisfying more complex criteria are programmed using a general purpose programming language. For example, the FILTER I tool from new Technologies Inc. uses heuristic procedure to find full names of persons in the collected information. Most custom searches, including FILTER I tool suffer from false positives and false negatives.

3.2.6 Search Modifications

Search of modification is automated search for data objects that have been modified since specified moment in the past. Modification of data objects that are not usually modified such as operating system utilities can be detected by comparing their current hash with their expected hash. A library of expected hashes must be build prior to the search.

Modification of a file can also be inferred from modification of its timestamp. Although plausible in many cases, this inference is circumstantial. Investigator assumes that a file is always modified simultaneously with its timestamp and since the timestamp is modified, he infers that the file was modified too. This is a form of event reconstruction.

3.2.7 Reconstruction of Events

Search techniques are commonly used for finding incriminating information. However, the mere fact of presence of objects does not prove that the owner of the computer is responsible for putting the objects in it. Apart from the owner, the objects can be generated automatically by the system. Or they can be planted by an intruder or virus program. Or they can be left by the previous owner of the computer. To determine who is responsible, the investigator must reconstruct events in the past that caused presence of the objects.

Reconstruction of events inside a computer requires understanding of computer functionality. Many techniques emerged for reconstructing events in specific operating systems. This dissertation classifies these techniques according to the primary object of analysis. Two major classes are identified as log file analysis and file system analysis.

3.3 LOG FILE ANALYSIS

A log file is a purposefully generated record of past events in a computer system. It is organised as a sequence of entries. A log file entry usually consists of a timestamp, an identifier of the process that generated the entry and some description of the reason for generating an entry.

It is common to have multiple log files on a single computer system. Different log files are usually created by the operating system for different types of events. In addition, many applications maintain their own log files.

Log file entries are generated by the system processes when something important (from the process's point of view) happens. For example, a TCP wrapper process generates one log file entry when a TCP connection is established and another log file entry when the TCP connection is released.

The knowledge of circumstances, in which processes generate log file entries permits forensic scientist to infer from presence or absence of log file entries that certain events happened. For example, from presence of two log file entries generated by TCP wrapper for some TCP connection X, forensic scientist can conclude that:

- TCP connection X happened;
- X was established at the time of the first entry; and
- X was released at the time of the second entry.

This reasoning suffers from implicit assumptions. It is assumed that the log file entries were generated by the TCP wrapper, which functioned according to the expectations of the forensic scientist; that the entries have not been tampered with; and that the timestamps on the entries reflect real time of the moments when the entries were generated. It is not always possible to ascertain these assumptions, which results in several possible explanations for appearance of the log file entries. For example, if possibility of tampering cannot be excluded, then forgery of the log file entries could be a possible explanation for their existence. To combat uncertainty caused by multiple explanations, forensic analyst seeks corroborating evidence, which can reduce number of possible explanations or give stronger support to one explanation than another.

3.3.1 Determining Temporal Order with Timestamps

Timestamps on log file entries are commonly used to determine temporal order of entries from different log files. The process is complicated by two time related problems, even if the possibility of tampering is excluded.

First problem may arise if the log file entries are recorded on different computers with different system clocks. Apart from individual clock imprecision, there may be an unknown skew between clocks used to produce each of the timestamps. If the skew is unknown, it is possible that the entry with the smaller timestamp could have been generated after the entry with the bigger timestamp.

Second problem may arise if resolution of the clocks is too coarse. As a result, the entries may have identical timestamps, in which case it is also not possible to determine whether one entry was generated before the other.

3.3.2 File System Analysis

Log files are not the only source of evidence that can be used for event reconstruction. Other data objects can also be used. This subsection describes how structural information stored by the file system can be used for event reconstruction.

In most operating systems, a data storage device is represented at the lowest logical level by a sequence of equally sized storage blocks that can be read and written independently. Most file systems divide all blocks into two groups. One group is used for storing user data and the other group is used for storing structural information.

Structural information includes structure of directory tree, file names, locations of data blocks allocated for individual files, locations of unallocated blocks etc. Operating system manipulates structural information in a certain well-defined way that can be exploited for event reconstruction.

3.3.3 Detection of Deleted Files

Information about individual files is stored in standardised file entries whose organisation differs from file system to file system. In Unix file systems, the information about a file is stored in a combination of i-node and directory entries pointing to that i-node. In Windows NT file system (NTFS), information about a file is stored in an entry of the Master File Table.

When a disk or a disk partition is first formatted, all such file entries are set to initial "unallocated" value. When a file entry is allocated for a file, it becomes active. Its fields are filled with proper information about the file. In most file systems, however, the file entry is not restored to the "unallocated" value when the file is deleted. As a result, presence of a file entry whose value is different from the initial "unallocated" value, indicates that that file entry once represented a file, which was subsequently deleted.

3.3.4 File Attributes Analysis

Every file in a file system either active or deleted has a set of attributes such as name, access permissions, timestamps and location of disc blocks allocated to the file. File attributes change when applications manipulate files via operating system calls.

File attributes can be analysed in much the same way as log file entries.

Timestamps are a particularly important source of information for event reconstruction.

In most file systems a file has at least one timestamp. In NTFS, for example, every active (i.e. non-deleted) file has three timestamps which are collectively known as MAC-times.

- Time of last Modification (M)
- Time of last Access (A)
- Time of Creation (C)

Imagine that there is a log file that records every file operation in the computer. In this imaginary log file, each of the MAC-times would correspond to the last entry for the corresponding operation (modification, access or creation) on the file entry in which the timestamp is located. To visualise this similarity between MAC-times and the log file, the "mactimes tool" from the coroner's toolkit sorts individual MAC-times of files both active and deleted and presents them in a list which resembles a log file.

Signatures of different activities can be identified in MAC-times like in ordinary log files. Given below are several such signatures which have been published.

3.3.5 Restoration of a Directory from a Backup

The fact that a directory was restored from a backup can be detected by inequality of timestamps on the directory itself and on its sub-directory '.' or '..'. When the directory is first created, both the directory timestamp and the timestamp on its sub-directories '.' and '..' are equal. When the directory is restored from a backup, the directory itself is assigned the old timestamp but its subdirectories '.' and '..' are timestamped with the time of backup restoration.

3.3.6 Exploit Compilation, Running and Deletion

The signature of compiling, running and deleting an exploit program is explored. It is concluded that when someone compiles, runs and deletes an exploit program, we expect to find traces of the deleted program source file of the deleted executable file as well as traces of compiler temporary files.

3.3.7 Moving a File

When a file is being moved in Microsoft FAT file systems, the old file entry is deleted and a new file entry is used in the new location. The new file entry maintains same block allocation information as the old entry. Thus, the discovery of a deleted file entry whose allocation information is identical to some active file, supports possibility that the file was moved.

3.3.8 Reconstruction of Deleted Files

In most file systems file deletion does not erase the information stored in the file. Instead, the file entry and the data blocks used by the file are marked as unallocated, so that they can be reused later for another file. Thus, unless the data blocks and the deleted file entry have been re-allocated to another file, the deleted file can usually be recovered by restoring its file entry and data blocks to active status.

Even if the file entry and some of the data blocks have been re-allocated, it may still be possible to reconstruct parts of the file. The lazarus tool, for example, uses several heuristics to find and piece together blocks that (could have) once belonged to a file. Lazarus uses the following heuristics about file systems and common file formats.

- In most file systems, a file begins at the beginning of a disk block;
- Most file systems write file into contiguous blocks, if possible;
- Most file formats have a distinguishing pattern of bytes near the beginning of the file;
- For most file formats, same type of data is stored in all blocks of a file; and
- Lazarus analyses disc blocks sequentially. For each block, lazarus tries to determine (1) the type of data stored in the block by calculating heuristic characteristics of the data in the block; and (2) whether the block is a first block in a file using well known file signatures. Once the block is determined as a "first block", all subsequent blocks with the same type of information are appended to it until new "first block" is found.

This process can be viewed as a very crude and approximate reconstruction based on some knowledge of the file system and application programs. Each reconstructed file can be seen as a statement that that file was once created by an application program, which was able to write such a file.

Since lazarus makes very bold assumptions about the file system, its reconstruction is highly unreliable. Despite that fact, states that lazarus works well for small files that fit entirely in one disk block. The effectiveness of tools such as lazarus can probably be improved by using more sophisticated techniques for determining the type of information contained in a disk block.

Check Your Progress 1

Note: a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of this Unit.

What is log file analysis?

....

....

....

....

3.4 KEYWORD SEARCH

3.4.1 Preparation – Creating Your Master List

The first step in the process is the groundwork. Always allocate a certain amount of time up front to plan and prepare the list of initial keywords to be used as a basis for conducting keyword research. We need to have an inventory of words or phrases to get started, so why not put some thought and effort into generating a solid list to work from. From one's perspective, the better the plan, the better the results. So let's get to it.

3.4.2 Preliminary Evaluation and Client Input

Conduct a brief audit on the existing content to determine what, if anything can be used to help get you started. I'll take a quick look at the site and start a list of keywords found in page titles, metadata, headlines and body copy as well as those appearing in site navigation. In addition to providing a place to start, it also gives me the opportunity to become more familiar with the client's content and how it has been structured. The next task is to meet with a variety of stakeholders from throughout the organization with the intention of soliciting the keywords they would typically use to describe the content being targeted on the site. This helps to uncover internal language use and identify marketing jargon or industry speak.

3.4.3 Competitive Analysis

While surveying stakeholders normally a list of who they believe their competitors are online are sought from the complainants. This information also becomes an important part of the overall plan and is a good way to see if their direct competition is actively pursuing their own keyword strategies. Next you start plugging some of the initial terms into Google (or another search engine) to see what you get. This is somewhat of an iterative process but has the potential to provide some valuable insight into the industry. Items of interest found on competitor sites along with those from the search results are documented and added to the list.

Search Tip

Placing a tilde (~) immediately before a keyword when conducting a web search using Google will return the search term in addition to synonyms for that keyword.

3.4.4 Recursive Term Expansion

Now that we have spoken with the client and gathered data from the competition and the search index, we have a fairly complete list to start our research with. There is, however, one last thing we need to do to make it slightly more comprehensive. We need to expand our root words to include things like prefixes, suffixes and plurals (a process also applied to our synonyms and related concepts). Any new or related terms added to the list are also put through this same process until the results have been exhausted, hence the name recursive term expansion. Finally, our list of keywords along with the synonyms and related concepts are then integrated with what we will call descriptors which assist with further defining the words and phrases to better describe the content we're trying to target. This includes things like consultants or consulting if the company is offering a service or a list of brand names if it is for a product. Let us now walk through a simple example to illustrate the whole process.

Example: "Digital Forensics"

Take for instance, a company that employs techniques in forensic sciences to solve computer crimes. This company intends to create content on their site targeted at a specific keyword. Let's look at the phrase "digital forensics" and follow the method outlined above to build a master list of keywords. A brief competitive analysis and term expansion leads to the following:

Root Terms: Digital, Forensics Related Concepts [Digital] – electronic, computer

Related Concepts [Forensics] – evidence, examination, examinations, investigation, investigations, science, sciences

Related Concepts [Digital Forensics] – firewall forensics, database forensics, mobile forensics, data remanence

Term Descriptors [Corporate] – business, businesses, engineer, engineers, company, companies, service, services, expert, experts, specialist, specialists, professional, professionals

Term Descriptors [Actions] – analyze, analyzing, collect, collecting, collection, identify, identifying, identification.

After putting it all together we get a total of more than 700 keywords in our master list! That's a far cry from the single two-word phrase we began with. Examples from the master list include:

- analyzing digital evidence;
- digital evidence experts;
- digital forensics companies;
- digital investigations;
- electronic evidence specialists; and
- electronic forensics.

This may sound like a lot of work but it really isn't. To complete the process, the final copy of the master list is sent to the client for review before the keyword research is started to ensure what's been added is relevant to the organization.

The purpose of this step was to give us the ability to cast as wide a net as possible in an effort to uncover as much of the language being used by our potential customers when searching for our content, products and/or services online. Doing so not only gives us the opportunity to wisely target the correct keywords, but also lets us craft our content in such a way as to tap into as much into the long tail as possible. To illustrate, we'll use the following Top Content report from Google Analytics. This particular page, although targeted toward a specific set of keywords, generated traffic from an amazing 5,766 unique keyword combinations! This alone demonstrates the power of the long tail in driving significant amounts of traffic to any website.

Keep in mind that no one wants to generate traffic just for traffic's sake; all want these visitors to do something while on the site, whether it's to buy product, fill out a form or contact your organization. Web analytics aside, now that we've done all the groundwork and assembled our master list of terms, we're ready to tackle the research part of our keyword research.

3.4.5 Keyword Research Tools

There are a number of different tools available on the market to conduct your keyword research with. Below is a list of some of the more popular ones but if you look around there are certainly others out there you might want to use.

- WordTracker;
- Trellian's Keyword Discovery Tool;
- Google AdWords Keyword Tool;
- Wordze; and
- Microsoft AdCenter Labs Keyword Tools.

We will use WordTracker (this is free tool available on web) to begin with then, if needed, the Google AdWords Keyword Tool to help verify the validity of some of the results. One should not typically put much faith in the actual numbers returned by the tools themselves. What we are really looking for are trends or general guidelines that help in choosing high value keywords for the site. For instance, we might find out that keyword X is searched for five times more often than keyword Y and therefore keyword X is a better candidate for inclusion in our content.

After logging in WordTracker account, we navigate to the keyword universe section of the site and begin to copy from the master list we have created and paste into the tool. WordTracker allows you to input up to 100 keywords at a time so with our list we will need to repeat the process at least 8 times. As we work our way down the list, sets of words and phrases that match or are related to the terms we've entered are returned. Each time through, we save the results to my keyword basket.

After we've made our way through the entire list, there's one last thing to do. We take all the keywords we've saved and perform a competition search on them. The outcome of the competition search is a metric called the Keyword Effectiveness Index or KEI. This metric is intended to identify keywords that are highly searched while at the same time having low competition, thus potentially providing an improved opportunity of making it to the top of the rankings. Typically however, most if not all of these keywords require further research and analysis, preferably in another tool. After collecting search frequencies and competitive search numbers, the next step is to perform some analysis.

3.4.6 Keyword Analysis: Interpreting the Results

The analysis part of the process is where the real insight comes from. This step provides the ability to determine, based on searcher behavior which additional words should be used in close proximity to the target keywords as well as those that should be leveraged in an effort to tap into the long tail.

Doing keyword research for as long as we have the need for the custom development of some specialized software has been a necessity. Software like this provides the ability to import, tag, slice and dice very large data sets in a multitude of ways. On an average 7.5 times more people search for the phrase "computer forensics" than those that search for "digital forensics".

Putting it into Action

Now that we have analyzed the data, we are ready to choose and implement the keywords into the copy on my page. Based on a quick analysis of the research, here's what we've come up with as a template for this page's metadata and key headlines:

Title Tag: *Computer Forensics: Experts in Digital Forensic Investigations*

Meta Description: *Computer forensic investigators specializing in the discovery of electronic evidence. <Company Name> leverages expertise in digital forensics for electronic discovery and data investigations*

Meta Keywords: *computer, forensics, forensic, expert, experts, digital, investigate, investigation, investigation, investigations, investigating, investigator, investigators, electronic, data, evidence, database, discovery, crime, crimes, analysis, examiner, company, service, services, <company name>*

Page Header 1: *Using Computer Forensics for Solving Digital Investigations*

Page Header 2: *Forensic Investigators providing Expertise to Clients around the Globe*

As you can see, based on the research rather than going with our original plan to target the phrase "digital forensics", we've decided to go after "computer forensics" instead. Further, as part of the long tail, we might also see visits from a variety of combinations taken from the above including "digital forensic(s)", "digital investigation(s)", "computer investigation(s)", "expert forensic(s)" and many others.

Without performing keyword research in the way that we did, we would never have been able to determine which keywords to target and how best to implement them within the content on my page. Further, we would've surely missed out on the opportunity to capture searchers using these keywords in this niche.

3.5 DATA RECOVERY

3.5.1 Salvaging Deleted Data

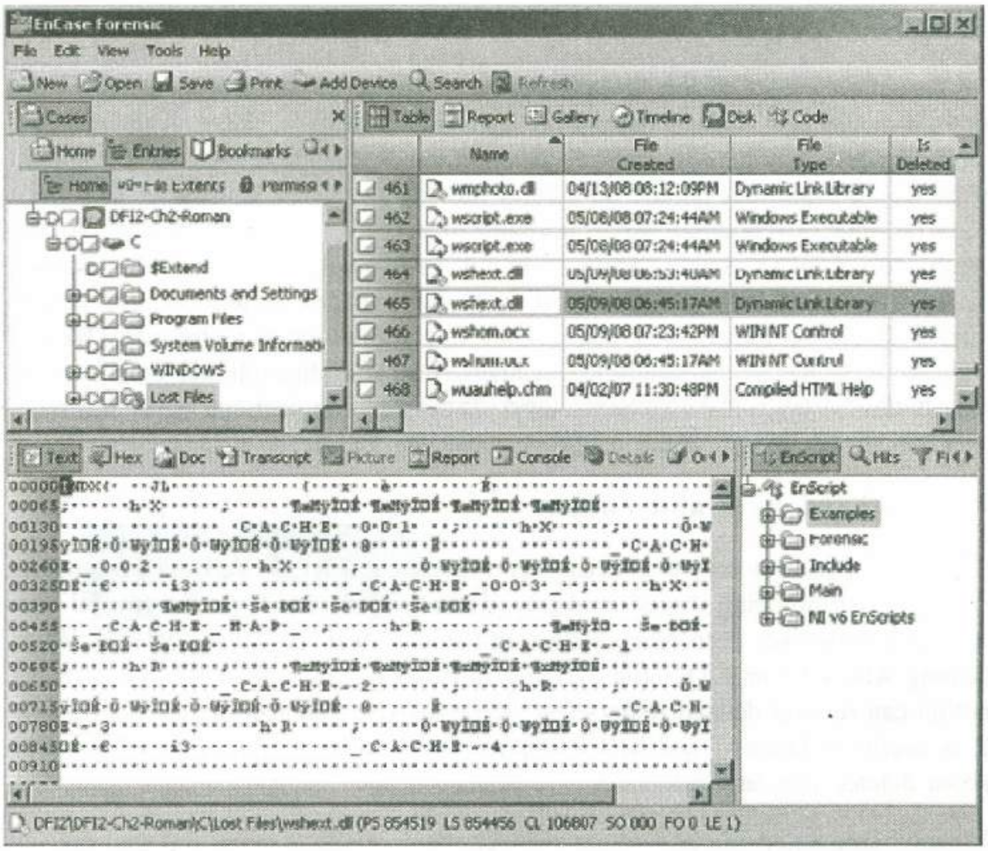
An important aspect of the forensic analysis process is to salvage all data from storage media and convert unreadable data into a readable form. Although data can be hidden on a drive in many ways and it is not feasible to look for all of them in all cases, examiners should be able to identify the major sources of data or at least be able to recognize large amounts of missing data. For instance, if the combined size of all visible partitions on a drive is much smaller than the capacity of the drive, this may be an indication that a partition is not being detected. Similarly, if a large amount of data cannot be classified or there are many files of a known type that are unusually large, this may be an indication that some form of data hiding or encryption is being used. The following sections cover the main areas on storage media where useful data may be found.

3.5.2 Deleted Files and Folders

Criminals often take steps to conceal their crimes and deleted data can often contain the most incriminating digital evidence. Therefore, one of the most fruitful data salvaging processes is to recover files and folders that have been deleted. When dealing with FAT or NTFS file systems, most tools can recover deleted files but not all can recover deleted folders that are no longer referenced by the file system. It is useful to know if and how different tools recover and present information about deleted files and folders.

Notably, automated file and folder recovery tools make assumptions that are not always correct. For instance, when recovering deleted files, many tools take the starting cluster and file size from a folder entry and assign the next free clusters to the file sequentially. The underlying assumption in this process breaks down when the starting cluster of one deleted file is followed by free clusters that belonged to a different deleted file. So, automated tools can generate correct or incorrect results depending on the assumptions they make and the particular situation. Furthermore, if two deleted directory entries point to the same cluster, it can take some effort to determine which filename and associated date-time stamps referenced that cluster most recently. Some automated file recovery tools distinguish between directory entries that have been deleted versus those that have been deleted and overwritten. However, care must be taken not to assume that files with newer date-time stamps accessed the associated clusters more recently. There are sufficient nuances to file datetime stamps that an apparently newer file could be created before the apparently old one.

Fig. 1 provides an example of an apparently recoverable deleted file. A closer inspection of the data that is displayed for this file reveals that it is not the actual/ original contents of the file. These problems are compounded when more aggressive salvaging techniques are used. Some forensic analysis tools have a powerful feature that scours a disk for deleted folders on FAT and NTFS systems. For instance, X–Ways provides a "Thorough" salvaging capability and EnCase has a "Recover Folders" feature. The resulting salvaged file system information may enable forensic analysts to recovered data that was associated with deleted files that are no longer referenced by the current file system. However, different implementations of this salvaging process can lead to inconsistent results between different tools or even different versions of the same tool. Many existing forensic tools combine salvaged file system details with the existing file system and this mixing of two separate file system states often leads to more conflicts of the type described in the previous section. The conflicts arise when a new file has reused the clusters that were previously allocated to a salvaged file. To guard against mistakes and misinterpretations, it is critically important that forensic analysts conceptually separate these two system states, thinking of the salvaged file system information as separate and overlaid on top of the existing file systems.



	Name	File Created	File Type	Is Deleted
461	wmphoto.dll	04/13/08 08:12:09PM	Dynamic Link Library	yes
462	wscript.exe	05/08/08 07:24:44AM	Windows Executable	yes
463	wscript.exe	05/08/08 07:24:44AM	Windows Executable	yes
464	wshext.dll	05/09/08 06:53:40AM	Dynamic Link Library	yes
465	wshext.dll	05/09/08 06:45:17AM	Dynamic Link Library	yes
466	wshom.ocx	05/09/08 07:23:42PM	WIN NT Control	yes
467	wshom.ocx	05/09/08 06:45:17AM	WIN NT Control	yes
468	wuauhelp.chm	04/02/07 11:30:48PM	Compiled HTML Help	yes

DF12\DF12-Ch2-Roman\C\Lost Files\wshext.dll (PS 654519 LS 654456 CL 106807 SO 000 FO 0 LE 1)

Fig. 1

For the simplest conflicts, both the current and salvaged file system have file system metadata pointing to same area on disk, generally providing forensic analysts with sufficient information to resolve the conflict. More difficult conflicts arise when the space allocated to a salvaged file has been overwritten by newer data but file system details about the newer file are not recoverable. In this situation, the diligent forensic analyst will determine that the file content is not consistent with the salvaged file system details. This inconsistency may be as simple as a salvaged Microsoft Word document pointing to an area of the disk that contains a JPEG graphics file. However, some inconsistencies are more difficult to detect and require a more thorough comparison between the recovered data and the salvaged file system information. In more complicated situations, multiple salvaged filenames point to the same area of the disk, creating additional levels of complexity when trying to reconstruct activities on a computer system. Therefore, forensic analysts must take additional steps to determine whether the data that appears to have been allocated to a salvaged file in fact was recovered properly and accurately.

When a few deleted files or folders are critical to a case, forensic analysts should examine them closely for inconsistencies and seek corroborating evidence from other areas of the computer system or connected networks. The same caveats apply to deleted items in physical memory of computers and mobile devices. Forensic tools are emerging that capture the full contents of memory from various devices and attempt to reconstruct the data structures containing files, processes, call logs, SMS messages and other information.

Given the variety of data structures, mistakes may be made when parsing physical memory dumps and attempting to recover deleted items. Documentation relating to this recovery process, such as an inventory of the recovered files and a description of how they were salvaged should be maintained to enable others to assess what was recovered. Given the variations between tools and the potential for error, it is advisable to compare the results from one tool using another. Such a comparison can be made by comparing the inventories of undeleted files from both tools for any differences.

3.5.3 File Carving

On storage media, the space that is available to store new data is called unallocated space. This area on a disk is important from an investigative standpoint because it often contains significant amounts of data from deleted files.

File carving tools like Foremost, Scalpel, DataLifter and PhotoRec scour unallocated space for characteristics of certain file types in an effort to salvage deleted files. This salvaging process generally produces a large percentage of incomplete and corrupt files because file carving tools rarely know the size of the original files and the deleted files may be fragmented or partially overwritten. If the approximate size of the original files is known, forensic analysts can adjust a parameter in file carving tools to set the maximum size of the salvaged files to increase the number of successfully salvaged files. When specific files are of interest like deleted NT Event Logs, file carving tools can be customized with the associated file signature to salvage these file types.

Some forensic utilities break large files such as unallocated and swap space into smaller pieces to facilitate processing such as file carving and indexing. If an examiner exports and processes these segments of unallocated space individually with standalone file carving utilities, it is possible that, depending on where the boundaries are portions of salvaged items may be missing.

Keep in mind that most file carving methods work on the assumption that a file is stored in contiguous clusters. The advantage of performing file carving on extracted unallocated space rather than on a full forensic image is that the data on disk may not have been arranged in consecutive clusters but may become consecutive when extracted into a single file. Also keep in mind that files salvaged using this technique do not have file names or date-time stamps associated with them so the examiner needs to assign them names in a systematic way.

Researchers are developing more sophisticated techniques to salvage fragmented deleted files in response to DFRWS Forensic Challenges (www.dfrws.org). Currently, these advanced salvaging methods work for only certain file types like PDF and ZIP files.

3.5.4 Handling Special Files

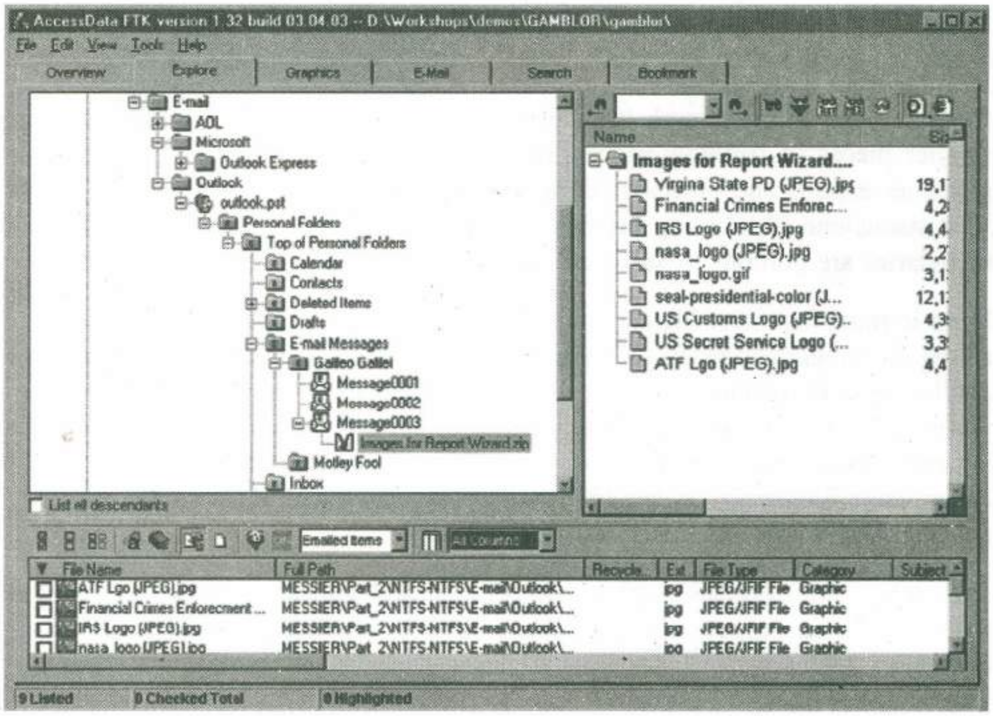
There are certain files that will not be immediately accessible to keyword searching. In addition to encrypted and password protected files, certain file formats store text in binary or proprietary formats. These "special" files include compressed archives, encoded attachments in e-mail messages and encrypted and password protected files.

Another very common example is Adobe PDF documents that have been "secured" to prevent extraction of text and TIFF fax documents. Keyword searches on these of course will be useless. If there is a large quantity of such unsearchable files that are relevant to a case, more than can be reviewed manually, the typical approach is to use optical character recognition (OCR) to convert them to text, thus rendering them searchable. Attention must be paid to ensure that mistakes are not introduced by the OCR process.

A criminal investigation may centre around kickbacks on contracts and forensic analysts may focus their attention on Microsoft Office documents and other files that are keyword searchable. However, the smoking gun evidence may actually be in TIFF e-mail attachments, JPG items in "My Scans" or received fax items. An effective identification and pre-processing procedure process will identify such items whereas just performing keyword would fail.

MIME encoding adds an additional layer of encoding to e-mail attachments. An alternate approach to searching for e-mail messages is to use a tool that understands the specific file format and makes it accessible for keyword searching.

FTK can interpret and index an Outlook PST file as shown in Fig. 2, giving forensic analysts efficient access to e-mail messages and many attachments. EnCase can also interpret some of these proprietary formats using the View File Structure feature. An added advantage of using this type of specialized tool is that they support analysis of metadata within e-mail. For instance, a search can be restricted to a date range of interest. When using such specialized tools for processing digital evidence, it is important to understand their inner workings in order to avoid mistakes and misinterpretations. For instance, exporting items found in e-mail using certain tools may not preserve the parent-child relationship between messages and attachments, which can be important in certain situations.



```
E-mail
  AOL
  Microsoft
    Outlook Express
  Outlook
    outlook.pst
      Personal Folders
        Top of Personal Folders
          Calendar
          Contacts
          Deleted Items
          Drafts
          E-mail Messages
            Galileo Galilei
              Message0001
              Message0002
              Message0003
                Images for Report Wizard.zip
            Motley Fool
          Inbox
```

Name	Si
Images for Report Wizard....	
Virgina State PD (JPEG).jp	19,1
Financial Crimes Enforec...	4,2
IRS Logo (JPEG).jpg	4,4
nasa_logo (JPEG).jpg	2,2
nasa_logo.gif	3,1
seal-presidential-color (J...	12,1
US Customs Logo (JPEG)..	4,3
US Secret Service Logo (...	3,3
ATF Lgo (JPEG).jpg	4,4

File Name	Full Path	Recycle...	Ext	File Type	Category	Subject
□ ATF Lgo (JPEG).jpg	MESSIER\Part_2\NTFS-NTFS\E-mail\Outlook\...		jpg	JPEG/JFIF File	Graphic	
□ Financial Crimes Enforecment ...	MESSIER\Part_2\NTFS-NTFS\E-mail\Outlook\...		jpg	JPEG/JFIF File	Graphic	
□ IRS Logo (JPEG).jpg	MESSIER\Part_2\NTFS-NTFS\E-mail\Outlook\...		jpg	JPEG/JFIF File	Graphic	
□ nasa_logo (JPEG).jpg	MESSIER\Part_2\NTFS-NTFS\E-mail\Outlook\...		jpg	JPEG/JFIF File	Graphic	

Fig. 2

Another approach to viewing proprietary formats, such as Lotus Notes, is to restore them to a disk and view them via the native client application. In some cases it is possible to recover messages that have been deleted but have not been purged from e-mail containers.

When special files are corrupt, it may be possible to repair the damage using specially designed utilities. For example, EasyRecovery Professional from Ontrack can repair a variety of file types from Windows systems including Outlook files and Zip archives. On UNIX systems, there are tools for repairing a more limited set of files such as tarfix and fixcpio.

3.5.5 Extracting Embedded Metadata

The purpose of this step is to harvest additional metadata relating to the files of interest to support further analysis. Embedded metadata can answer a variety of questions regarding a document, including its provenance and authenticity. Embedded metadata can also help generate important leads, pointing to other sources of digital evidence on the system or Internet. Photographs taken by digital cameras can contain details such as the make and model of the camera, the date and time the picture was taken (according to the camera's clock) and with some models the GPS coordinates of the camera when the photograph was taken. The data in such photographs found on a computer or on the Internet can be compared with those of a digital camera seized in the defendant's home to determine if they are consistent helping to establish a link.

A review of all installed and uninstalled applications may provide significant information. For example, installed "covert" communications (steg), privacy software (Tor), specialized applications to destroy data (e.g. BCWipe or Evidence Eliminator) or transfer mechanisms (e.g. Winsock FTP, peer-to-peer) may dictate a much more detailed analysis in an attempt to document user activity associated with such utilities. Installed applications may also provide insight as the user's knowledge level. An example would be existence and use of hexadecimal editors, "patch" files and low-level programming utilities such as Microsoft Assembler along with user generated source code. Specialized user generated material such as source code may require review by an expert.

In addition to review of applications, an examiner typically performs an antivirus and malware scan. Knowing which, if any, virus programs and/or malware or remote access tools are present may be an important aspect of the case. For example, forensic analysis of malware may be necessary when a defendant uses the "Trojan defence," claiming that all incriminating material on his computer is attributable to a remote intruder who compromised the system using a remote administration tool (a.k.a. Trojan horse program). Assessing the capabilities of malware found on the system may reveal that it could not have been used to place the incriminating files on the computer and analyzing activities on the computer around the time in question may support the conclusion that the defendant was using the computer when the incriminating files were placed on the system.

Once elements of a user's activity are processed, the examiner can identify significant elements for reporting or further analysis. An example would be the user executed Google searches on "deletion utilities" followed by "prevent undeleted" with subsequent download and execution of such a utility to delete several company confidential documents the day of her previously unannounced resignation but only after she connected a USB thumb drive, reviewed the confidential documents, performed a File->Save As to save them to the thumb drive.

3.5.6 Using Data from Data Files

A data file (or simply called a file) is a collection of information logically grouped into a single entity and referenced by a unique name, such as a filename. A file can be of many data types, including a document, an image, a video or an application. Successful forensic processing of computer media depends on the ability to collect, examine and analyze the files that reside on the media.

This section provides an overview of the most common media types and file systems, methods for naming, storing organizing and accessing files. It then discusses how files should be collected and how the integrity of the files should be preserved. The section also discusses various technical issues related to file recovery such as recovering data from deleted files. The last portion of the section describes the examination and analysis of files, providing guidance on tools and techniques that can assist analysts.

3.5.7 File Storage Media

The widespread use of computers and other digital devices has resulted in a significant increase in the number of different media types that are used to store files. In addition to traditional media types such as hard drives and floppy disks, files are often stored on consumer devices such as PDAs and cell phones, as well as on newer media types, such as flash memory cards, which were made popular by digital cameras.

3.5.8 File Systems

Before media can be used to store files, the media must usually be partitioned and formatted into logical volumes. Partitioning is the act of logically dividing a media into portions that function as physically separate units. A logical volume is a partition or a collection of partitions acting as a single entity that has been formatted with a file system. Some media types, such as floppy disks, can contain at most one partition (and consequently, one logical volume). The format of the logical volumes is determined by the selected file system.

A file system defines the way that files are named, stored organized and accessed on logical volumes. Many different file systems exist, each providing unique features and data structures. However, all file systems share some common traits. First, they use the concepts of directories and files to organize and store data. Directories are organizational structures that are used to group files together. In addition to files, directories may contain other directories called subdirectories. Second, files ystems use some data structure to point to the location of files on media. In addition, they store each data file written to media in one or more file allocation units. These are referred to as clusters by some files ystems (e.g. File Allocation Table [FAT], NT File System [NTFS]) and as blocks by other file systems (e.g. UNIX and Linux). A file allocation unit is simply a group of sectors, which are the smallest units that can be accessed on media.

Some commonly used file systems are as follows:

- **FAT12**. FAT12 is used only on floppy disks and FAT volumes smaller than 16 MB. FAT12 uses a 12-bit file allocation table entry to address an entry in the file system.
- **FAT16**. MS-DOS, Windows 95/98/NT/2000/XP, Windows Server 2003 and some UNIX OSs support FAT16 natively. FAT16 is also commonly used for multimedia devices such as digital cameras and audio players. FAT16 uses a 16-bit file allocation table entry to address an entry in the file system. FAT16 volumes are limited to a maximum size of 2 GB in MS-DOS and Windows 95/98. Windows NT and newer OSs increase the maximum volume size for FAT16 to 4 GB.
- **FAT32.** Windows 95 Original Equipment Manufacturer (OEM) Service Release 2 (OSR2), Windows 98/2000/XP and Windows Server 2003 support FAT32 natively, as do some multimedia devices. FAT32 uses a 32-bit file allocation table entry to address an entry in the file system. The maximum FAT32 volume size is 2 terabytes (TB).
- **NTFS**. Windows NT/2000/XP and Windows Server 2003 support NTFS natively. NTFS is a recoverable file system, which means that it can automatically restore the consistency of the file system when errors occur. In addition, NTFS supports data compression and encryption and allows user and group-level access permissions to be defined for data files and directories. The maximum NTFS volume size is 2 TB.

- **High-Performance File System (HPFS)**. HPFS is supported natively by OS/2 and can be read by Windows NT 3.1, 3.5 and 3.51. HPFS builds on the directory organization of FAT by providing automatic sorting of directories. In addition, HPFS reduces the amount of lost disk space by utilizing smaller units of allocation. The maximum HPFS volume size is 64 GB.

- **Second Extended File System (ext2fs)**.ext2fs is supported natively by Linux. It supports standard UNIX file types and file system checks to ensure file system consistency. The maximum ext2fs volume size is 4 TB.

- **Third Extended File System (ext3fs)**. ext3fs is supported natively by Linux. It is based on the ext2fs file system and provides journaling capabilities that allow consistency checks of the file system to be performed quickly on large amounts of data. The maximum ext3fs volume size is 4 TB.

- **ReiserFS**. ReiserFS is supported by Linux and is the default file system for several common versions of Linux. It offers journaling capabilities and is significantly faster than the ext2fs and ext3fs file systems. The maximum volume size is 16 TB.

- **Hierarchical File System (HFS)**. HFS is supported natively by Mac OS. HFS is mainly used in older versions of Mac OS but is still supported in newer versions. The maximum HFS volume size under Mac OS 6 and 7 is 2 GB. The maximum HFS volume size in Mac OS 7.5 is 4 GB. Mac OS 7.5.2 and newer Mac OSs increase the maximum HFS volume size to 2 TB.

- **HFS Plus**. HFS Plus is supported natively by Mac OS 8.1 and later and is a journaling file system under Mac OS X. It is the successor to HFS and provides numerous enhancements, such as long file name support and Unicode file name support for international filenames. The maximum HFS Plus volume size is 2 TB.

- **UNIX File System (UFS)**. UFS is supported natively by several types of UNIX OSs, including Solaris, FreeBSD, OpenBSD and Mac OS X. However, most OSs have added proprietary features, so the details of UFS differ among implementations.

- **Compact Disk File System (CDFS)**. As the name indicates, the CDFS file system is used for CDs.

- **International Organization for Standardization (ISO) 9660 and Joliet**. The ISO 9660 file system is commonly used on CD-ROMs. Another popular CD-ROM file system, Joliet, is a variant of ISO 9660. ISO 9660 supports file name lengths of up to 32 characters, whereas Joliet supports up to 64 characters. Joliet also supports Unicode characters within file names.

- **Universal Disk Format (UDF)**. UDF is the file system used for DVDs and is also used for some CDs.

3.5.9 Other Data on Media

File systems are designed to store files on media. However, file systems may also hold data from deleted files or earlier versions of existing files. This data can provide important information. The following items describe how this data can still exist on various media:

- **Deleted Files**. When a file is deleted, it is typically not erased from the media; instead, the information in the directory's data structure that points to the location of the file is marked as deleted. This means that the file is still stored on the media but is no longer enumerated by the OS. The operating system considers this to be free space and can overwrite any portion of or the entire deleted file at any time.

- **Slack Space**. As noted previously, file systems use file allocation units to store files. Even if a file requires less space than the file allocation unit size, an entire file allocation unit is still reserved for the file. For example, if the file allocation unit size is 32 kilobytes (KB) and a file is only 7 KB, the entire 32 KB is still allocated to the file, but only 7 KB is used, resulting in 25 KB of unused space. This unused space is referred to as file slack space and it may hold residual data such as portions of deleted files.

- **Free Space**. Free space is the area on media that is not allocated to any partition; it includes unallocated clusters or blocks. This often includes space on the media where files (and even entire volumes) may have resided at one point but have since been deleted. The free space may still contain pieces of data.

- Another way in which data might be hidden is through Alternate Data Streams (ADS) within NTFS volumes. NTFS has long supported multiple data streams for files and directories. Each file in an NTFS volume consists of an unnamed stream that is used to store the file's primary data and optionally one or more named streams (i.e., file.txt:Stream1, file.txt:Stream2) that can be used to store auxiliary information, such as file properties and picture thumbnail data. For instance, if a user right-clicks on a file in Windows Explorer, views the files properties and then modifies the information displayed in the summary tab, the OS stores the summary information for the file in a named stream.

- All data streams within a file share the file's attributes (e.g. timestamps, security attributes). Although named streams do affect the storage quota of a file, they are largely concealed from users because standard Windows file utilities, such as Explorer, only report the size of a file's unnamed stream. As a result, a user cannot readily determine whether a file contains ADS using the standard Windows file utilities. This allows hidden data to be contained within any NTFS file system. Moving files with ADS to non-NTFS file systems effectively strips ADS from the file, so the ADS can be lost if analysts are not aware of their presence. Software and processes are available to identify ADS.

3.5.10 Collecting Files

During data collection, the analyst should make multiple copies of the relevant files or file systems typically a master copy and a working copy. The analyst can then use the working copy without affecting the original files or the master copy. It is often important to collect not only the files, but also significant timestamps for the files, such as when the files were last modified or accessed.

3.5.11 Copying Files from Media

Files can be copied from media using two different techniques:

- **Logical Backup**. A logical backup copies the directories and files of a logical volume. It does not capture other data that may be present on the media, such as deleted files or residual data stored in slack space.

- **Bit Stream Imaging**. Also known as disk imaging, bit stream imaging generates a bit-for-bit copy of the original media, including free space and slack space. Bit stream images require more storage space and take longer to perform than logical backups.

If evidence may be needed for prosecution or disciplinary actions, the analyst should get a bit stream image of the original media, label the original media and store it securely as evidence. All subsequent analysis should be performed using the copied media to ensure that the original media is not modified and that a copy of the original media can always be recreated if necessary. All steps that were taken to create the image copy should be documented. Doing so should allow any analyst to produce an exact duplicate of the original media using the same procedures. In addition, proper documentation can be used to demonstrate that evidence was not mishandled during the collection process. Besides the steps that were taken to record the image, the analyst should document supplementary information such as the hard drive model and serial number, media storage capacity and information about the imaging software or hardware that was used (e.g. name, version number, licensing information). All of these actions support the maintenance of the chain of custody.

When a bit stream image is executed, either a disk-to-disk or a disk-to-file copy can be performed. A disk-to-disk copy, as its name suggests, copies the contents of the media directly to another media. A disk-to-file copy copies the contents of the media to a single logical data file. A disk-to-disk copy is useful since the copied media can be connected directly to a computer and its contents readily viewed. However, a disk-to-disk copy requires a second media similar to the original media. A disk-to-file copy allows the data file image to be moved and backed up easily. However, to view the logical contents of an image file, the analyst has to restore the image to media or open or read it from an application capable of displaying the logical contents of bit stream images. The details of this are OS and forensics tool-dependent.

Numerous hardware and software tools can perform bit stream imaging and logical backups. Hardware tools are generally portable, provide bit-by-bit images, connect directly to the drive or computer to be imaged and have built-in hash functions. Hardware tools can acquire data from drives that use common types of controllers, such as Integrated Drive Electronics (IDE) and Small Computer System Interface (SCSI). Software solutions generally consist of a startup diskette, CD or installed programs that run on a workstation to which the media to be imaged is attached. Some software solutions create logical copies of files or partitions and may ignore free or unallocated drive space, whereas others create a bit-by-bit image copy of the media.

In addition to their primary function, some disk imaging tools can also perform forensic recordkeeping, such as automated audit trails and chain of custody. The use of such tools can support consistency in the examination process and the accuracy and reproducibility of results. An increasing number of disk imaging tools are becoming available. In response to this proliferation and the lack of a standard for testing them, NIST's Computer Forensics Tool Testing (CFTT) project has developed rigorous testing procedures for validating the tools' results. Currently, only a few disk imaging tools have undergone CFTT testing.

Generally, tools that perform bit stream imaging should not be used to acquire bit-by-bit copies of an entire physical device from a live system. A system currently in use because the files and memory on such a system are changing constantly and therefore cannot be validated. However, a bit-by-bit copy of the logical areas of a live system can be completed and validated. When logical backups are being performed, it is still preferable not to copy files from a live system; changes might be made to files during the backup and files that are held open by a process might not be easy to copy. Accordingly, analysts should decide whether copying files from a live system is feasible based on which files need to be obtained, how accurate and complete the copying needs to be and how important the live system is. For example, it is not necessary to take down a critical server used by hundreds of people just to collect files from a single user's home directory. For logical backups of live systems, analysts can use standard system backup software. However, performing a backup could affect the performance of the system and consume significant amounts of network bandwidth, depending on whether the backup is performed locally or remotely.

Organizations should have policy, guidelines and procedures that indicate the circumstances under which bit stream images and logical backups (including those from live systems) may be performed for forensic purposes and which personnel may perform them. It is typically most effective to establish policy, guidelines and procedures based on categories of systems (i.e. low, moderate or high impact) and the nature of the event of interest; some organizations also choose to create separate policy statements, guidelines and procedures for particularly important systems. The policy, guidelines or procedures should identify the individuals or groups with authority to make decisions regarding backups and images; these people should be capable of weighing the risks and making sound decisions. The policy, guidelines or procedures should also identify which individuals or groups have the authority to perform the backup or imaging for each type of system. Access to some systems might be restricted because of the sensitivity of the operations or data in the system.

3.5.12 Data File Integrity

During backups and imaging, the integrity of the original media should be maintained. To ensure that the backup or imaging process does not alter data on the original media, analysts can use a write-blocker while backing up or imaging the media. A write-blocker is a hardware or software-based tool that prevents a computer from writing to computer storage media connected to it. Hardware write-blockers are physically connected to the computer and the storage media being processed to prevent any writes to that media. Software write-blockers are installed on the analyst's forensic system and currently are available only for MS-DOS and Windows systems. (Some OSs [e.g. Mac OS X, Linux] may not require software write-blockers because they can be set to boot with secondary devices not mounted. However, attaching a hardware write-blocking device will ensure that integrity is maintained.) MS-DOS based software write-blockers work by trapping Interrupt 13 and extended Interrupt 13 disk writes. Windows-based software write-blockers use filters to sort interrupts sent to devices to prevent any writes to storage media.

In general, when using a hardware write-blocker, the media or device used to read the media should be connected directly to the write-blocker and the write-blocker should be connected to the computer or device used to perform the backup or imaging. When using a software write-blocker, the software should be loaded onto a computer before the media or device used to read the media is connected to the computer. Write-blockers may also allow write-blocking to be toggled on or off for a particular device. It is important when write-blocking is used, that it be toggled on for all connected devices.37 Write-blockers also should be tested routinely to ensure that they support newer devices. For example, a new device might make use of reserved or previously unused functions or placeholders to implement device-specific functions that might ultimately write to the device and alter its contents.

After a backup or imaging is performed, it is important to verify that the copied data is an exact duplicate of the original data. Computing the message digest of the copied data can be used to verify and ensure data integrity. A message digest is a hash that uniquely identifies data and has the property that changing a single bit in the data will cause a completely different message digest to be generated. There are many algorithms for computing the message digest of data, but the two most commonly used are MD5 and Secure Hash Algorithm 1 (SHA-1). These algorithms take as input data of arbitrary length and produce as output 128-bit message digests.

- When a bit stream image is performed, the message digest of the original media should be computed and recorded before the image is performed. After the imaging, the message digest of the copied media should be computed and compared with the original message digest to verify that data integrity has been preserved. The message digest of the original media should then be computed again to verify that the imaging process did not alter the original media and all results should be documented. The process should be used for logical backups, except that message digests should be computed and compared for each data file. For both bit stream images and logical backups, the message digests created to ensure data integrity should be stored on read-only or write-once media or printed and then secured in a proper location.

3.5.13 File Modification, Access and Creation Times

It is often important to know when a file was created, used or manipulated and most OSs keep track of certain timestamps related to files. The most commonly used timestamps are the modification, access and creation (MAC) times as follows:

- **Modification Time**. This is the last time a file was changed in any way, including when a file is written to and when it is changed by another program.
- **Access Time**. This is the last time any access was performed on a file (e.g. viewed, opened, printed).
- **Creation Time**. This is generally the time and date the file was created; however, when a file is copied to a system, the creation time will become the time the file was copied to the new system. The modification time will remain intact.

Different types of file system may store different types of times. For example, Windows systems retain the last modified time, the last access time and the creation time of files. UNIX systems retain the last modification, last inode change and last access times; however, some UNIX systems (including versions of BSD and SunOS) do not update the last access time of executable files when they are run. Some UNIX systems record the time when the metadata for a file was most recently altered. Metadata is data about data; for file systems, metadata is data that provides information about a file's contents.

If an analyst needs to establish an accurate timeline of events, then the file times should be preserved. Accordingly, analysts should be aware that not all methods for collecting data files can preserve file times. Bit stream images can preserve file times because a bit-for-bit copy is generated; performing a logical backup using some tools may cause file creation times to be altered when the data file is copied. For this reason, whenever file times are essential, bit stream imaging should be used to collect data.

Analysts should also be aware that file times may not always be accurate. Among the reasons for such inaccuracies are the following:

- The computer's clock does not have the correct time. For example, the clock may not have been synchronized regularly with an authoritative time source.
- The time may not be recorded with the expected level of detail, such omitting the seconds or minutes.
- An attacker may have altered the recorded file times.

Several technical issues may arise in collecting data files. The primary issue is the collection of deleted files and remnants of files existing in free and slack space on media. Individuals can use a variety of techniques to hinder the collection of such data. For example, there are many utilities available that perform wiping – the overwriting of media (or portions of media, such as particular files) with random or constant values (e.g. all 0.s). Such utilities vary in services and reliability, but most are effective in preventing easy collection of files, especially if several wipes are performed. Individuals can also use physical means to prevent data collection, such as demagnetizing a hard drive (also known as degaussing) or physically damaging or destroying media. Both physical and software-based techniques can make it very difficult or even impossible, to recover all of the data using software. Recovery attempts in these cases necessitate the use of highly specialized forensic experts with advanced facilities, hardware and techniques, but the cost and effort involved in making use of such means are prohibitive for general use. In some cases, the data is simply not recoverable.

Another common issue is the collection of hidden data. Many OSs permit users to tag certain files, directories or even partitions as hidden, which means that by default they are not displayed in directory listings. Some applications and OSs hide configuration files to reduce the chance that users will accidentally modify or delete them. Also, on some OSs, directories that have been deleted may be marked as hidden. Hidden data may contain a wealth of information; for example, a hidden partition could contain a separate OS and many data files. Users may create hidden partitions by altering the partition table to disrupt disk management and prevent applications from seeing that the data area exists. Hidden data can also be found within ADSs on NTFS volumes, in the end-of-file slack space and free space on a medium and in the Host Protected Area (HPA) on some hard drives, which is a region of a drive intended to be used by vendors only. Many collection tools can recognize some or all of these methods of hiding data and recover the associated data.

Yet another issue that may arise is collection of data from RAID arrays that use striping (e.g. RAID-0, RAID-5). In this configuration, a striped volume consists of equal-sized partitions that reside on separate disk drives. When data is written to the volume, it is evenly distributed across the partitions to improve disk performance. This can cause problems because all partitions of a striped volume must be present for the examination of its contents, but in this case the partitions reside on separate physical disk drives. To examine a striped volume, each disk drive in the RAID array needs to be imaged and the RAID configuration has to be recreated on the examination system. The examination system needs to be booted using a forensic boot disk that can recognize and use the RAID array and that prevents writes to the array. Some imaging tools can acquire striped volumes and preserve unused data areas of the volume, such as free space and slack space.

3.5.14 Examining Data Files

After a logical backup or bit stream imaging has been performed, the backup or image may have to be restored to another media before the data can be examined. This is dependent on the forensic tools that will be used to perform the analysis. Some tools can analyze data directly from an image file, whereas others require that the backup or image be restored to a medium first. Regardless of whether an image file or a restored image is used in the examination, the data should be accessed only as read-only to ensure that the data being examined is not modified and that it will provide consistent results on successive runs. Write-blockers can be used during this process to prevent writes from occurring to the restored image. After restoring the backup (if needed), the analyst begins to examine the collected data and performs an assessment of the relevant files and data by locating all files, including deleted files, remnants of files in slack and free space and hidden files. Next, the analyst may need to extract the data from some or all of the files which may be complicated by such measures as encryption and password protection.

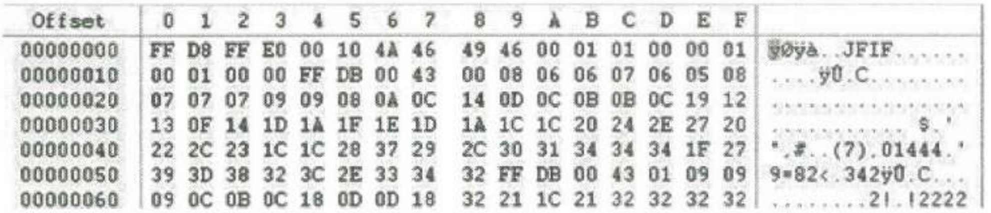
3.5.15 Locating the Files

The first step in the examination is to locate the files. A disk image can capture many gigabytes of slack space and free space, which could contain thousands of files and file fragments. Manually extracting data from unused space can be a time-consuming and difficult process because it requires knowledge of the underlying file system format. Fortunately, several tools are available that can automate the process of extracting data from unused space and saving it to data files, as well as recovering deleted files and files within a recycling bin. Analysts can also display the contents of slack space with hex editors or special slack recovery tools.

3.5.16 Extracting the Data

The rest of the examination process involves extracting data from some or all of the files. To make sense of the contents of a file, an analyst needs to know what type of data the file contains. The intended purpose of file extensions is to denote the nature of the file's contents; for example, a jpg extension indicates a graphic file and an mp3 extension indicates a music file. However, users can assign any file extension to any type of file, such as naming a text file mysong.mp3 or omitting a file extension. In addition, some file extensions might be hidden or unsupported on other OSs. Therefore, analysts should not assume that file extensions are accurate.

Analysts can more accurately identify the type of data stored in many files by looking at their file headers. A file header contains identifying information about a file and possibly metadata that provides information about the file's contents. The file header contains a file signature that identifies the type of data that particular file contains. If a file has a file header of FF D8, that indicates that this is a JPEG file. A file header could be located in a file separate from the actual file data. Another effective technique for identifying the type of data in a file is a simple histogram showing the distribution of ASCII values as a percentage of total characters in a file. For example, a spike in the "space", "a" and "e" lines generally indicates a text file, while consistency across the histogram indicates a compressed file. Other patterns are indicative of files that are encrypted or that were modified through steganography.



```
Offset    0  1  2  3  4  5  6  7   8  9  A  B  C  D  E  F
00000000  FF D8 FF E0 00 10 4A 46  49 46 00 01 01 00 00 01  ÿØÿà..JFIF......
00000010  00 01 00 00 FF DB 00 43  00 08 06 06 07 06 05 08  ....ÿÛ.C........
00000020  07 07 07 09 09 08 0A 0C  14 0D 0C 0B 0B 0C 19 12  ................
00000030  13 0F 14 1D 1A 1F 1E 1D  1A 1C 1C 20 24 2E 27 20  ............ $.'
00000040  22 2C 23 1C 1C 28 37 29  2C 30 31 34 34 34 1F 27  ",#..(7),01444.'
00000050  39 3D 38 32 3C 2E 33 34  32 FF DB 00 43 01 09 09  9=82<.342ÿÛ.C...
00000060  09 0C 0B 0C 18 0D 0D 18  32 21 1C 21 32 32 32 32  ........2!.!2222
```

Analysts may also need to access non-stegged files that are protected by passwords. Passwords are often stored on the same system as the files they protect but in an encoded or encrypted format. Various utilities are available that can crack passwords placed on individual files, as well as OS passwords. Most cracking utilities can attempt to guess passwords, as well as performing brute force attempts that try every possible password. The time needed for a brute force attack on an encoded or encrypted password can vary greatly, depending on the type of encryption used and the sophistication of the password itself. Another approach is to bypass a password. For example, an analyst could boot a system and disable its screensaver password or bypass a Basic Input/Output System (BIOS) password by pulling the BIOS jumper from the systems motherboard or using a manufacturer's backdoor password. Of course, bypassing a password might mean rebooting the system which might be undesirable. Another possibility is to attempt to capture the password through network or host-based controls (e.g. packet sniffer, keystroke logger) with proper management and legal approval. If a boot-up password has been set on a hard drive, it might be possible to guess it (i.e. a default password from a vendor) or to circumvent it with specialized hardware and software.

3.5 ENCRYPTION AND STEGANOGRAPHY

Encryption can present a significant challenge for digital forensic practitioners, particularly full disk encryption. Even when full disk encryption is not used or can be circumvented, additional effort is required to salvage data from password protected or encrypted files. When dealing with individually protected files, it is sometimes possible to use a hexadecimal editor like WinHex to simply remove the password within a file. There are also specialized tools that can bypass or recover passwords of various files.

Currently, the most powerful and versatile tools for salvaging password protected and encrypted data are PRTK and DNA from AccessData. The Password Recovery Toolkit can recover passwords from many file types and is useful for dealing with encrypted data. Also, it is possible for a DNA network to try every key in less time by combining the power of several computers. Distributed Network Attack (DNA) can brute-force 40-bit encryption of certain file types including Adobe Acrobat and Microsoft Word and Excel. Using a cluster of approximately 100 off-the-shelf desktop computers and the necessary software, it is possible to try every possible 40-bit key in five days. Rainbow tables can be used to accelerate the password guessing process. Some vendors also have hardware decryption platforms based on implementation of field programmable gate arrays that can increase the speed of brute force attacks.

When strong encryption is used such as BestCrypt, PGP or Windows Encrypting File System, a brute-force approach to guessing the encryption key is generally infeasible. In such cases, it may be possible to locate unencrypted versions of data in unallocated space, swap files and other areas of the system. For instance, printer spool files on Windows and UNIX systems can contain data from files that have been deleted or encrypted. Alternatively, it may be possible to obtain an alternate decryption key. For instance some encryption programs advise users to create a recovery disk in case they forget their password. When EFS is used, Windows automatically assigns an encryption recovery agent that can decrypt messages when the original encryption key is unavailable. In Windows 2000, the built-in administrator account is the default recovery agent (an organization can override the default by assigning a domain-wide recovery agent provided the system is part of the organization's Windows 2000 domain).

Notably, prior to Windows XP, EFS private keys were weakly protected and it was possible to gain access to encrypted data by replacing the associated NT logon password with a known value using a tool like ntpasswd and logginginto a bootable/virtualized clone of the system with the new password.

When investigating a child exploitation case, it is advisable to be on the lookout for other forms of data concealment such as steganography. Forensic analysts can make educated guesses to identify files containing hidden data-the presence of steganography software and uncharacteristically large files should motivate examiners to treat these as special files that require additional processing. In such cases, it may be possible to salvage the hidden data by opening the files using the steganography software and providing a password that was obtained during the investigation. More sophisticated techniques are available for detecting hidden data. Even if encryption or steganography cannot be bypassed, documenting which files are concealing data can help an investigator determine the intent of the defendant.

Encryption often presents challenges for analysts. Users might encrypt individual files, folders, volumes or partitions so that others cannot access their contents without a decryption key or passphrase. The encryption might be performed by the OS or a third-party program. Although it is relatively easy to identify an encrypted file, it is usually not so easy to decrypt it. The analyst might be able to identify the encryption method by examining the file header, identifying encryption programs installed on the system or finding encryption keys (which are often stored on other media). Once the encryption method is known, the analyst can better determine the feasibility of decrypting the file. In many cases, it is not possible to decrypt files because the encryption method is strong and the authentication (e.g. passphrase) used to perform decryption is unavailable.

Although an analyst can detect the presence of encrypted data rather easily, the use of steganography is more difficult to detect. Steganography, also known as steg, is the embedding of data within other data. Digital watermarks and the hiding of words and information within images are examples of steganography. Some techniques an analyst can use to locate stegged data include looking for multiple versions of the same image, identifying the presence of grayscale images, searching metadata and registries, using histograms and using hash sets to search for known steganography software. Once certain that stegged data exists, analysts might be able to extract the embedded data by determining what software created the data and then finding the stego key or by using brute force and cryptographic attacks to determine a password. However, such efforts are often unsuccessful and can be extremely time-consuming, particularly if the analyst does not find the presence of known steganography software on the media being reviewed. In addition, some software programs can analyze files and estimate the probability that the files were altered with steganography.

Check Your Progress 2

Note: a) Space is given below for writing your answers.

b) Compare your answers with the one given at the end of this Unit.

1) How is search important step in investigation?

....

....

....

....

....

2) What is salvaging of data?

....

....

....

....

....

3) What is file carving?

....

....

....

....

....

4) Compare file recovery and file carving?

....

....

....

....

....

3.7 LET US SUM UP

This unit covers various techniques of search which are group of techniques searches collects information to answer the question whether objects of given type, such as hacking tools or pictures of certain kind, are present in the collected information. According to the level of search automation, this unit classifies techniques into manual browsing and automated searches. Automated searches include keyword search, regular expression search, approximate matching search, custom searches and search of modifications. Different scientific methods followed in data search are also explained. It also describes about data recovery and various tools used in data recovery.

3.8 CHECK YOUR PROGRESS : THE KEY

Check Your Progress 1

Log file

A log file is a purposefully generated record of past events in a computer system. It is organised as a sequence of entries. A log file entry usually consists of a timestamp, an identifier of the process that generated the entry and some description of the reason for generating an entry.

It is common to have multiple log files on a single computer system. Different log files are usually created by the operating system for different types of events. In addition, many applications maintain their own log files.

Log file entries are generated by the system processes when something important (from the process's point of view) happens. For example, a TCP wrapper process generates one log file entry when a TCP connection is established and another log file entry when the TCP connection is released.

Check Your Progress 2

1) Most searching for evidence is done in a file system and inside files. A common search technique is to search for files based on their names or patterns in their names. Another common search technique is to search for files based on a keyword in their content. We can also search for files based on their temporal data, such as the last accessed or written time. Hash databases can be used to find files that are known to be bad or good. Another common method of searching is to search for files based on signatures in their content. This allows us to find all files of a given type even if someone has changed their name. When analyzing network data, we may search for all packets from a specific source address or all packets going to a specific port. We also may want to find packets that have a certain keyword in them.

2) **Salvaging of data**

The latest computer forensic technologies enable salvaging data from damaged and destroyed electronic devices. Both defence and prosecution teams benefit from this advanced technology. Analyzing data from computers provides e-mail information, including sent, received and deleted messages. Active and deleted computer files undergo scrutiny. SIM cards removed from cell phones can yield lists of dialled numbers, contact numbers, texts that include sent, drafted and deleted messages, locations where cell phones were last used and overseas network providers.

Data recovery is the process of salvaging data from damaged, failed, corrupted or inaccessible secondary storage media when it cannot be accessed normally. Often the data are being salvaged from storage media such as hard disk drives, storage tapes, CDs, DVDs, RAID and other electronics.

Recovery may be required due to physical damage to the storage device or logical damage to the file system that prevents it from being mounted by the host operating system.

The most common "data recovery" scenario involves an operating system (OS) failure (typically on a single-disk, single-partition, single-OS system), in which case the goal is simply to copy all wanted files to another disk. This can be easily accomplished with a Live CD, most of which provide a means to mount the system drive and backup disks or removable media and to move the files from the system disk to the backup media with a file manager or optical disc authoring software. Such cases can often be mitigated by disk partitioning and consistently storing valuable data files (or copies of them) on a different partition from the replaceable OS system files.

There could be disk-level failure, such as a compromised file system or disk partition or a hard disk failure. In any of these cases, the data cannot be easily read. Depending on the situation, solutions involve repairing the file system, partition table or master boot record or hard disk recovery techniques ranging from software-based recovery of corrupted data to hardware replacement on a physically damaged disk. If hard disk recovery is necessary, the disk itself has typically failed permanently and the focus is rather on a one-time recovery, salvaging whatever data can be read.

Sometimes, files have been "deleted" from a storage medium. Typically, deleted files are not erased immediately; instead, references to them in the directory structure are removed and the space they occupy is made available for later overwriting. In the meantime, the original file may be restored. Although there is some confusion over the term, "data recovery" may also be used in the context of forensic applications or espionage.

3) **File carving**

File Carving is the process of extracting a collection of data from a larger data set. Data carving techniques frequently occur during a digital investigation when the unallocated file system space is analyzed to extract files. The files are "carved" from the unallocated space using file type-specific header and footer values.

File system structures are not used during the process. File carving is a powerful tool for recovering files and fragments of files when directory entries are corrupt or missing. Carving is also especially useful in criminal cases, where the use of carving techniques can recover evidence. In certain cases related to child pornography, Law Enforcement agents were abl to recover more images from the suspect's hard-disks by using carving techniques.

Memory carving is a useful technique for analyzing physical and virtual memory dumps when the memory structures are unknown or have been overwritten. An example of memory dump carving is the recovery of files from a mobile phone.

4) **Compare file recovery and file carving**

When data is lost on a medium, people want to recover it. There is a big difference between file recovery techniques and carving. File recovery techniques make use of the file system information that remains after deletion of a file. By using this information, many files can be recovered. A disadvantage is that the file system information needs to be correct. If not, the files can't be recovered. If a system is formatted, the file recovery techniques will not work either. Carving works with the raw data and doesn't use the file system structure during its process. A File system is a structure for storing and organizing computer files and the data they contain. Examples of often used file systems are: FAT16, FAT32, NTFS, EXT etc. Although carving doesn't care about which file system is used to store the files, it could be very helpful to understand how the specific file system works.

Carving makes use of the internal structure of a file. A file is a block of stored information like an image in a jpeg file. A computer is using extensions in file names to identify what these files contain.

3.9 SUGGESTED READINGS

- www.aafs.org
- www.cops.org
- www.forensic-computing.com
- www.ijde.org
- www.shk-dplc.com

UNIT 4 FORENSIC EXAMINATION OF NETWORK DEVICES

Structure

4.0 INTRODUCTION

Network forensics is the capture, recording and analysis of network events in order to discover the source of security attacks or other problem incidents. (The term, attributed to firewall expert Marcus Ranum, is borrowed from the legal and criminology fields where forensics pertains to the investigation of crimes).

Digital investigation surrounding network forensics starts with obtaining the event data in the first place. Many times intrusions occur and the events get deleted by the perpetrator on the system that was compromised. If these events have not been stored or sent to another location then they are usually gone forever. Another obstacle to actually obtaining network or system events is that appliances and applications that provide this type of capability are usually extremely expensive and difficult to implement, in essence becoming cost prohibitive in regards to Return on Investment (ROI). To compound the pressures organizations face in regards to implementing proper network forensics and log management techniques, several organization have implemented policies that require organizations keep all network event data. In such situations there is no other choice but to implement expensive archiving equipment and analysis software to monitor and archive network security events. Organizations can only hope they can prove at some future date that the network security events gathered have not been altered. Assuming, of course, they even have any events at all.

However, numerous types of disparate devices and event log formats exist making them difficult to monitor, manage or correlate for any action, typically requiring a combination of tools and consoles for an incident to materialize. Also, until recently there has not been an easy way to correlate IDS alerts with firewall logs, system logs or vulnerability scans. Being notorious for high false positive rates, a correlated IDS alert is much more meaningful.

There is general information overload with millions of appliance, application and system event logs being generated every day. The final result and primary problem with network forensics is the dynamic nature of network event data and the fact that it is rarely audited, inadequately archived and easily lost, deleted or copied.

The Investigations Triad

All network forensic investigations revolve around what is known as the Investigations Triad. To meet the goals of the Investigations Triad, as it pertains to network forensics, we will discuss three main topics:

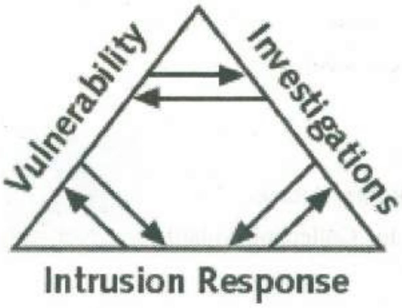


Fig. 1: Investigations Triad

Implementing real-time network forensic techniques is an effective method of initially identifying and responding to computer crimes and policy violations. With a proper security management tool an analyst can monitor, automate and investigate network forensic event data, as well as respond much quicker to IDS events by minimizing false positives. Correlating security events, investigating and acting according to policy and properly archiving network and system events over time, are critical elements of preparing an organization to be successful in current and future network forensic investigations. In tandem, vulnerability assessment and risk management are required elements of any investigation, to test and verify the integrity of computer systems, servers and enterprise networks. A tool to monitor network IDS and provide incident response functions, is desirable because it helps identify anomalies, such as covert channels and intruder attacks using automated tools and of course helps in correlating these anomalies on the network with system and firewall logs. Computer investigative functions are necessary to manage, protect and maintain the forensic integrity of network-based systems and devices.

4.1 OBJECTIVES

After studying this unit, you should be able to:

- understand network devices;

- explain different scientific methods followed in investigation;
- elucidate various tools used in forensic examination of network devices; and
- explain various areas of concern in the networks and network devices.

4.2 INTRUSION DETECTION SYSTEMS

Intrusion detection systems help computer systems prepare for and deal with attacks. They collect information from a variety of vantage points within computer systems and networks and analyze this information for symptoms of security problems.

Vulnerability assessment systems check systems and networks for system problems and configuration errors that represent security vulnerabilities. Both intrusion detection and vulnerability assessment technologies allow organizations to protect themselves from losses associated with network security problems. This section explains how intrusion detection and vulnerability assessment fits into the overall framework of security products and techniques used in computer forensics.

Protecting critical information systems and networks is a complex operation, with many tradeoffs and considerations. The effectiveness of any security solution strategy depends on selecting the right products with the right combination of features for the system environment one wishes to protect. This section also provides the information one needs to be a savvy consumer in the areas of intrusion detection and vulnerability assessment.

4.2.1 Definition of Intrusion Detection

Intrusion detection systems help computer systems prepare for and deal with attacks. They accomplish this goal by collecting information from a variety of system and network sources and then analyzing the information for symptoms of security problems. In some cases, intrusion detection systems allow the user to specify real-time responses to the violations. Intrusion detection systems perform a variety of functions:

- Monitoring and analysis of user and system activity;
- Auditing of system configurations and vulnerabilities;
- Assessing the integrity of critical system and data files;
- Recognition of activity patterns reflecting known attacks;
- Statistical analysis of abnormal activity patterns; and
- Operating system audit trail management, with recognition of user activity reflecting policy violations.

Some systems provide additional features, including:

- Automatic installation of vendor-provided software patches; and
- Installation and operation of decoy servers to record information about intruders.

The combination of these features allows system managers to more easily handle the monitoring, audit and assessment of their systems and networks. This ongoing assessment and audit activity is a necessary part of sound security management practice.

4.2.2 Vulnerability Assessment

Vulnerability assessment products (also known as scanners) perform rigorous examinations of systems in order to determine weaknesses that might allow security violations. These products use two strategies for performing these examinations. First, passive, host-based mechanisms inspect system configuration files for unwise settings, system password files for weak passwords and other system objects for security policy violations. These checks are followed, in most cases, by active, network-based assessment, which re-enacts common intrusion scripts, recording system responses to the scripts.

The results of vulnerability assessment tools represent a snapshot of system security at a point in time. Although these systems cannot reliably detect an attack in progress, they can determine that an attack is possible and furthermore, they can sometimes determine that an attack has occurred. Because they offer benefits that are similar to those provided by intrusion detection systems, they are included in the sphere of intrusion detection technologies and products.

Products Can Be Successfully Deployed in Operational Environments

The objective of intrusion detection and vulnerability assessment is to make complex, tedious and sometimes virtually impossible system security management functions possible for those who are not security experts. Products are therefore designed with user-friendly interfaces that assist system administrators in their installation, configuration and use. Most products include information about the problems they discover, including how to correct these problems and provide valuable guidance for those who need to improve their security skills. Many vendors provide consulting and integration services to assist customers in successfully using their products to achieve their security goals.

4.2.3 Network Security Management

Network security management is a process in which one establishes and maintains policies, procedures and practices required for protecting networked information system assets. Intrusion detection and vulnerability assessment products provide capabilities needed as part of sound network security management practice.

4.2.4 Trust and Intrusion Detection

Another area of discussion when considering the value of intrusion detection systems is the need to monitor the rest of the security infrastructure. Firewalls, identification and authentication products, access control products, virtual private networks, encryption products and virus scanners all perform functions essential to system security. Given their vital roles, however, they are also prime targets of attack by adversaries. On a less sinister note, they are also managed by mere mortals and therefore subject to human error. Be it due to mis-configuration, outright failure or attack, the failure of any of these components of the security infrastructure jeopardizes the security of the systems they protect.

By monitoring the event logs generated by these systems, as well as monitoring the system activities for signs of attack, intrusion detection systems provide an added measure of integrity to the rest of the security infrastructure. Vulnerability assessment products also allow system management to test new configurations of the security infrastructure for flaws and omissions that might lead to problems.

4.2.5 System Security Management: A Process View

Securing systems is not a point fix. It is an ongoing process targeting a dynamic environment in which new threats arise daily. Prevention covers those proactive measures taken by organizations to mitigate risks to their system security. Much of the classic, government-sponsored work in computer security addresses this area by focusing on the design and implementation of more secure operating systems and applications software. Prevention also includes security policy formation, encryption, strong identification and authentication and firewalls. .

Functions in the detection phase are primarily provided by intrusion detection systems, although virus scanners also fall into this category. Thus, detection involves monitoring the targeted system(s), analyzing the information gathered for problems and then, based on the system settings, responding to the problems, reporting the problems or both.

The results of the detection process drive the other two stages of managing security:

a) Investigating problems that are discovered and documenting the cause of the problem and

b) Either correcting the problem or devising a means of dealing with it should it occur again. A common vision for future intrusion detection systems is that of performing these last two stages automatically or else performing the functions internal to detection so well that the need for the last two stages is virtually eliminated.

The combination of the investigation and diagnosis/resolution phases is often called *incident response* or *incident handling*. Organizations should specify policies, procedures and practices to address this area as it does the rest of security.

4.2.6 Intrusion Detection Systems and Related Technologies

Every new market suffers from exaggeration and misconception. Some of the claims made in marketing materials are reasonable and others are misleading. Here is a primer on how to read intrusion detection marketing literature.

Realistic Benefits

First of all, intrusion detection systems (IDSs) can lend a greater degree of integrity to the rest of your security infrastructure. The reason for this is because they monitor the operation of firewalls, encrypting routers, key management servers and files critical to other security mechanisms, thus providing additional layers of protection to a secured system. Therefore, the strategy of a system attacker will often include attacking or otherwise nullifying security devices protecting the intended target. Intrusion detection systems can recognize these first hallmarks of attack and potentially respond to them, mitigating damage. In addition, when these devices fail, due to configuration, attack or user error, intrusion detection systems can recognize the problem and notify the right people.

Second, intrusion detection systems can also make sense of often obtuse system information sources, telling you what's really happening on your systems. Operating system audit trails and other system logs are a treasure trove of information about what's going on internal to your systems. They are also often incomprehensible, even to expert system administrators and security officers. Intrusion detection systems allow administrators and managers to tune, organize and comprehend what these information sources tell them, often revealing problems before loss occurs.

Third, intrusion detection systems can also trace user activity from the point of entry to the point of exit or impact. Intrusion detection systems offer improvements over perimeter protections such as firewalls. Expert attackers can often penetrate firewalls; therefore, the ability to correlate activity corresponding to a particular user is critical to improving security.

Fourth, intrusion detection systems can recognize and report alterations to data files. Putting trojan horses in critical system files is a standard attack technique.

Similarly, the alteration of critical information files to mask illegal activity, damage reputations or commit fraud is common. File integrity assessment tools utilize strong cryptographic checksums to render these files tamper-evident and in the case of a problem, quickly ascertain the extent of damage.

Fifth, intrusion detection systems can also spot errors of your system configuration that have security implications, sometimes correcting them if the user wishes. Vulnerability assessment products allow consistent auditing and diagnosis of system configuration settings that might cause security problems. These products offer extensive vendor support and turnkey design so that even novice security personnel can look for hundreds of problems by pushing a button. Some of these product offerings even offer automated fixes for the problems uncovered.

Sixth, intrusion detection systems can recognize when your system appears to be subject to a particular attack. Vulnerability assessment products also allow the user of a system to quickly determine what attacks should be of concern to that system.

Again, strong vendor support allows novice security personnel to re-enact scores of hacker attacks against their system, automatically recording the results of these attack attempts. These products also provide a valuable sanity check for those installing and setting up new security infrastructures. It is far better for a system manager to determine that his or her firewall is incorrectly configured immediately than to discover this after an attacker has successfully penetrated it.

Seventh, intrusion detection systems can relieve your system management staff of the task of monitoring the Internet, searching for the late hacker attacks. Many intrusion detection and assessment tools come with extensive attack signature databases against which they match information from your system. The firms developing these products have expert staffs that monitor the Internet and other sources for reports and other information about new hacker attack tools and techniques. They then use this information to develop new signatures that are provided to customers for download from Web sites, downloaded to customers via encrypted e-mail messages or both.

Eighth, intrusion detection systems can make the security management of your systems by non-expert staff possible. Some intrusion detection and assessment tools offer those with no security expertise the ability to manage security-relevant features of your systems from a user-friendly interface. These are window-based, point-and-click screens that step users through setup and configuration in a logical readily understood fashion.

Finally, intrusion detection systems can provide guidelines that assist you in the vital step of establishing a security policy for your computing assets. Many intrusion detection and assessment products are part of comprehensive security suites that include security policy building tools. These provide easy-to-understand guidance in building your security policy, prompting you for information and answers that allow you to articulate goals and guidelines for the use of your computer systems.

Unrealistic Expectations

First, intrusion detection systems are not silver bullets. Security is a complex area with myriad possibilities and difficulties. In networks, it is also a "weakest link" phenomenon (it only takes one vulnerability on one machine to allow an adversary to gain entry and potentially wreak havoc on the entire network). The time it takes for this to occur is minuscule. There are no magic solutions to network security problems and intrusion detection products are no exception to this rule. However, as part of a comprehensive security management they can play a vital role in protecting your systems.

Second, intrusion detection systems cannot compensate for weak identification and authentication mechanisms. Although leading-edge research in intrusion detection asserts that sophisticated statistical analysis of user behavior can assist in identification of a particular person by observing their system activity, this fact is far from demonstrated. Therefore, you must still rely on other means of identification and authentication of users. This is best accomplished by strong authentication mechanisms (including token-based or biometric schemes and one-time passwords). A security infrastructure that includes strong identification and authentication and intrusion detection is stronger than one containing only one or the other.

Third, intrusion detection systems cannot conduct investigations of attacks without human intervention. In very secure environments, incidents happen. In order to minimize the occurrence of incidents (and the possibility of resulting damage) one must perform *incident handling*. One must investigate the attacks, determine, where possible, the responsible party and then diagnose and correct the vulnerability that allowed the problem to occur, reporting the attack and particulars to authorities where required. In some cases, especially those involving a dedicated attacker, finding the attacker and pursuing criminal charges against the attacker is the only way to make the attacks cease. However, the intrusion-detection system is not capable of identifying the person at the other end of the connection without human intervention.

The best that it can do is identify the IP address of the system that served as the attacker's point of entry-the rest is up to a human incident handler.

Fourth, intrusion detection systems cannot intuit the contents of your organizational security policy. Intrusion-detection expert systems increase in value when they are allowed to function as both hacker/burglar alarms and policy-compliance engines.

These functions cannot only spot the high-school hacker executing the "teardrop" attack against your file server, but also spot the programmer accessing the payroll system after hours. However, this policy compliance checking can exist only if there is a security policy to serve as a template for constructing detection signatures.

Fifth, intrusion detection systems cannot compensate for weaknesses in transmission control protocol (TCP)/IP and many other network protocols do not perform strong authentication of host source and destination addresses. This means that the source address that is reflected in the packets carrying an attack does not necessarily cor's system; it is very difficult to prove the identity of an attacker in a court of law-for example, in civil or criminal legal processes.

Sixth, intrusion detection systems cannot compensate for problems in the quality or integrity of information the system provides. In other words, "garbage in garbage out" still applies. System information sources are mined from a variety of points within the system. Despite the best efforts on the part of system vendors, many of these sources are software-based; as such, the data are subject to alteration by attackers. Many hacker tools (for example "cloak" and "zap") explicitly target system logs, selectively erasing records corresponding to the time of the attack and covering the intruders' tracks. This argues for the value of integrated, sometimes redundant, information sources; each additional source increases the possibility of obtaining information not corrupted by an attacker.

Seventh, intrusion detection systems cannot analyze all of the traffic on a busy network. Network-based intrusion detection is capable of monitoring traffic on a network, but only to a point. Given the vantage point of network-based intrusion detection sources that rely on network adapters set to promiscuous mode, not all packets are visible to the systems. Also, as traffic levels rise, the associated processing load required to keep up becomes prohibitive and the analysis engine either falls behind or fails. In fact, vendors themselves characterized the maximum bandwidth at which they had demonstrated their products to operate without loss with 100% analysis coverage at 65 MB/sec.

Eighth, intrusion detection systems cannot always deal with problems involving packet-level attacks. There are weaknesses in packet-capture-based network intrusion detection systems. The heart of the vulnerabilities involves the difference between the intrusion detection systems' interpretation of the outcome of a network transaction (based on its reconstruction of the network session) and the destination node for that network session's actual handling of the transaction.

Therefore, a knowledgeable adversary can send series of fragmented and otherwise doctored packets that elude detection but launch attacks on the destination node. Worse yet, an adversary can use this sort of packet manipulation to accomplish a denial of service attack on the intrusion detection systems itself by overflowing memory allocated for incoming packet queues.

Finally, intrusion detection systems cannot deal with modern network hardware and features. Dealing with fragmented packets can also be problematic. This problem has serious ramifications when one considers modern high-speed asynchronous transfer mode (ATM) networks that use packet fragmentation as a means of optimizing bandwidth. Other problems associated with advances in network technologies include the effect of switched networks on packet-capture-based network intrusion detection systems. As the effect of switched networks is to establish a network segment for each host, the range of coverage for a network intrusion system is reduced to a single host. This problem can be mitigated in those switches offering monitoring ports or spanning capability; however, these features are not universal in current equipment.

The capabilities for intrusion detection are growing as new technologies enter the marketplace and existing organizations expand their product offerings to allow additional sensor inputs, improved analysis techniques and more extensive signature databases.

Thanks to so much interest in information warfare, of which intrusion detection is a vital defensive component, funding of research efforts has skyrocketed, with no end in sight. This increased activity will result in enhanced understanding of the intrusion detection process and new features in future products. Intrusion detection products have now been embedded as standard components of major governmental and financial networks.

As intrusion detection remains an active research area, look for future technologies to implement new techniques for managing data and detecting scenarios of interest. Also look for products that function at application level and that interoperate with network management platforms. Finally, look for product features that are integrated into a bevy of special purpose devices, ranging from bandwidth management products to "black box" plug-ins for targeted environments.

Check Your Progress 1

Notes: a) Space is given below each of the questions for writing your answer.

b) Compare your answer with those given at the end of this Unit.

What is vulnerability assesment?

....

....

....

4.3 FIREWALL SECURITY SYSTEMS

Today, when an organization connects its private network to the Internet, security has to be one of primary concerns. In the past, before the widespread interest in the Internet, most network administrators were concerned about attacks on their networks from within, perhaps from disgruntled workers. For most organizations now connecting to the Internet and big business and big money moving toward electronic commerce at warp speed, the motive for mischief from outside is growing rapidly and creating a major security risk to enterprise networks.

Reacting to this threat, an increasing number of network administrators are installing the latest firewall technology as a first line of defense in the form of a barrier against outside attacks. These firewall gateways provide a choke point at which security and auditing can be imposed. They allow access to resources on the Internet from within the organization while providing controlled access from the Internet to hosts inside the virtual private network (VPN).

The threat of attack on your network increases proportionally with the continued exponential growth of the Internet. If it is necessary for you to connect your network to the Internet, an appropriate security protocol should be decided on and implemented. This book illustrates many reasons why this is necessary, as well as many techniques to consider for your firewall solution. The bottom line is, do not connect your network to the Internet without some sort of protection. Also, do not put sensitive information in a place where it can be accessed over the Internet. The firewall you decide to use will prevent most of the attacks on your network; however, firewalls will not protect against dial-in modem attacks, virus attacks or attacks from within your company.

Nevertheless, a number of the security problems with the Internet can be remedied or made less serious through the use of existing and well-known techniques and controls for host security. A firewall can significantly improve the level of site security while at the same time permitting access to vital Internet services. This section provides an overview of firewall technology, including how they protect against vulnerabilities, what firewalls don't protect against and the components that make up a firewall.

4.3.1 Firewall

A firewall is a system or group of systems that enforces an access control policy between two networks. The actual means by which this is accomplished varies widely, but in principle, the firewall can be thought of as a pair of mechanisms: one that blocks traffic and one that permits traffic. Some firewalls place a greater emphasis on blocking traffic, while others emphasize permitting traffic. Probably the most important thing to recognize about a firewall is that it implements an access control policy.

If you don't have a good idea what kind of access you want to permit or deny or you simply permit someone or some product to configure a firewall based on what they or it think it should do, then they are making policy for your organization as a whole.

In other words, a firewall is a network security product that acts as a barrier between two or more network segments. The firewall is a system (which consists of one or more components) that provides an access control mechanism between your network and the network(s) on the other side(s) of the firewall. A firewall can also provide audit and alarm mechanisms that will allow you to keep a record of all access attempts to and from your network, as well as a real-time notification of things that you determine to be important.

Perhaps it is best to describe first what a firewall is not: a firewall is not simply a router, host system or collection of systems that provides security to a network.

Rather, a firewall is an approach to security; it helps implement a larger security policy that defines the services and access to be permitted and it is an implementation of that policy in terms of a network configuration, one or more host systems and routers and other security measures such as advanced authentication in place of static passwords. The main purpose of a firewall system is to control access to or from a protected network (a site). It implements a network access policy by forcing connections to pass through the firewall, where they can be examined and evaluated.

A firewall system can be a router, a personal computer, a host or a collection of hosts, set up specifically to shield a site or subnet from protocols and services that can be abused from hosts outside the subnet. A firewall system is usually located at a higher-level gateway, such as a site's connection to the Internet. However, firewall systems can be located at lower-level gateways to provide protection for some smaller collection of hosts or subnets.

Why do we need firewalls? What can a firewall do for you? Why would you want a firewall? What can a firewall not do for you? All of these burning questions are answered next for those inquiring security minds that want to know.

4.3.2 Reasons for Firewalls

The general reasoning behind firewall usage is that without a firewall, a subnet's systems are exposed to inherently insecure services such as Network File System (NFS) or Network Information Service (NIS) and to probes and attacks from hosts elsewhere on the network.

In a firewall-less environment, network security relies totally on host security and all hosts must, in a sense, cooperate to achieve a uniformly high level of security.

The larger the subnet, the less manageable it is to maintain all hosts at the same level of security. As mistakes and lapses in security become more common, break-ins occur not as the result of complex attacks, but because of simple errors in configuration and inadequate passwords.

4.3.3 Need for Firewalls

As technology has advanced to greatly expand the information technology systems capabilities of corporations, the threats to these systems have become numerous and complex. In today's world, corporations face a variety of information system attacks against their local area networks (LANs) and wide area networks (WANs).

Many of these attacks are directed through the Internet. These attacks come from three basic groups:

- Persons who see attacking a corporation's information system as a technological challenge;
- Persons with no identified political or social agenda who see attacking a corporation's information system as an opportunity for high-tech vandalism;
- Persons associated with a corporate competitor or political adversary who see;
- The corporation's information system as a legitimate strategic target;
- To combat this growing and complex threat to a corporation's LAN and Internet site, a series of protective countermeasures needs to be developed, continually updated and improved. Security services that are important to protecting a corporation's strategic information include:

- **Data Integrity**: Absolute verification that data has not been modified;
- **Confidentiality:** Privacy with encryption, scrambled text Authentication: Verification of originator on contract;
- **Non-Repudiation:** Undeniable proof-of-participation;
- **Availability:** Assurance of service demand.

The building and implementation of firewalls is an effective security countermeasure used to implement these security services. An external firewall is used to counter threats from the Internet. An internal firewall is primarily used to defend a corporation's LAN or WAN. The internal firewall is used to separate and protect corporate databases (for example, financial databases can be separated from personnel databases). In addition, internal firewalls can be used to separate different levels of information being sent over a corporate LAN or WAN (for example, corporate proprietary information dealing with research projects, financial data or personnel records).

Firewalls, however, are just one element in an array of possible information technology (IT) systems countermeasures. The most effective security countermeasure is a good corporate security strategy. The effectiveness of this strategy will have a direct bearing on the success of any firewall that a corporation builds or purchases.

For example, the two critical elements that form the basis of an effective corporate security strategy are: *least privilege* and *defence in depth.*

- **Least Privilege**

The principle of least privilege means that an object is given only the privileges it needs to perform its assigned tasks. Least privilege is an important principle in countering attacks and limiting damage.

- **Defence in Depth**

Don't depend on one security solution. Good security is usually found in layers. These layers should consist of a variety of security products and services. The solutions could be network security products (firewalls that could be both internal and external) and information systems security (INFOSEC) training (employee education through classes and threat and vulnerability briefings).

A firewall approach provides numerous advantages to sites by helping to increase overall host security. The following provides an overview of the primary benefits of using a firewall.

4.3.4 Benefits of Firewalls

A firewall provides a leveraged choke point for network security. It allows the corporation to focus on a critically vulnerable point: where the corporation's information system connects to the Internet. The firewall can control and prevent attacks from insecure network services. A firewall can effectively monitor all traffic passing through the system. In this manner, the firewall serves as an *auditor* for the system and can alert the corporation to anomalies in the system. The firewall can also log access and compile statistics that can be used to create a profile of the system.

Some firewalls, on the other hand, permit only e-mail traffic through them, thereby protecting the network against any attacks other than attacks against the e-mail service. Other firewalls provide less strict protections and block services that are known to be problems.

Generally, firewalls are configured to protect against unauthenticated interactive logins from the *outside* world. This, more than anything, helps prevent vandals from logging into machines on your network. More elaborate firewalls block traffic from the outside to the inside but permit users on the inside to communicate freely with the outside..

Firewalls are also important since they can provide a single *choke point* where security and audit can be imposed. Unlike in a situation where a computer system is being attacked by someone dialling in with a modem, the firewall can act as an effective *phone tap* and tracing tool. Firewalls provide an important logging and auditing function. Often, they provide summaries to the administrator about what kinds and amount of traffic passed through it, how many attempts there were to break into it, etc. The following are the primary benefits of using a firewall:

- Protection from vulnerable services
- Controlled access to site systems
- Concentrated security
- Enhanced privacy
- Logging and statistics on network use and misuse
- Policy enforcement Protection from Vulnerable Services

A firewall can greatly improve network security and reduce risks to hosts on the subnet by filtering inherently insecure services. As a result, the subnet network environment is exposed to fewer risks, since only selected protocols will be able to pass through the firewall.

For example, a firewall could prohibit certain vulnerable services such as NFS from entering or leaving a protected subnet. This provides the benefit of preventing the services from being exploited by outside attackers, but at the same time permits the use of these services with greatly reduced risk to exploitation. Services such as NIS or NFS that is particularly useful on a LAN basis can thus be enjoyed and used to reduce the host management burden.

Firewalls can also provide protection from routing-based attacks, such as source routing and attempts to redirect routing paths to compromised sites via Internet control message protocol (ICMP) redirects. A firewall could reject all source-routed packets and ICMP redirects and then inform administrators of the incidents.

4.3.5 Why Firewalls Aren't Enough?

A common question is how intrusion detection complements firewalls. One way of characterizing the difference is provided by classifying security violations by source- whether they come from outside the organization's network or from within. Firewalls act as a barrier between corporate (internal) networks and the outside world (Internet) and filter incoming traffic according to a security policy. This is a valuable function and would be sufficient protection were it not for these facts:

- Not all access to the Internet occurs through the firewall.
- Not all threat originates outside the firewall.
- Firewalls are subject to attack themselves

Not All Access to the Internet Occurs Through the Firewall

Users, for a variety of reasons ranging from naiveté to impatience, sometimes set up unauthorized modem connections between their systems connected to the internal network and outside Internet access providers or other avenues to the Internet.

The firewall cannot mitigate risk associated with connections it never sees.

Not All Threats Originate Outside the Firewall

A vast majority of loss from security incidents is traced to insiders. Again, the firewall only sees traffic at the boundaries between the internal network and the Internet.

If the traffic reflecting security breaches never flows past the firewall, it cannot see the problems.

As more organizations utilize strong encryption to secure files and public network connections, the focus of adversaries will shift to those places in the network in which the information of interest is not as likely to be protected: the internal network.

Intrusion detection systems are the only part of the infrastructure that is privy to the traffic on the internal network. Therefore, they will become even more important as security infrastructures evolve.

Firewalls Are Subject to Attack Themselves

Attacks and strategies for circumventing firewalls have been widely publicized since the first firewalls were fielded. A common attack strategy is to utilize *tunneling* to bypass firewall protections. Tunneling is the practice of encapsulating a message in one protocol (that might be blocked by firewall filters) inside a second message.

4.3.6 Controlled Access to Site Systems

A firewall also provides the ability to control access to site systems. For example, some hosts can be made reachable from outside networks, whereas others can be effectively sealed off from unwanted access. A site could prevent outside access to its hosts except for special cases such as e-mail servers or information servers.

This brings to the fore an access policy that firewalls are particularly adept at enforcing: do not provide access to hosts or services that do not require access. Put differently, why provide access to hosts and services that could be exploited by attackers when the access is not used or required? If, for example, a user requires little or no network access to his or her desktop workstation, then a firewall can enforce this policy.

ICMP is an extension to the IP defined by RFC 792. ICMP supports packets containing error, control and informational messages. The PING command, for example, uses ICMP to test an Internet connection.

4.3.7 Concentrated Security

A firewall can be less expensive for an organization, in that all or most modified software and additional security software could be located on the firewall systems as opposed to being distributed on many hosts. In particular, one-time password systems and other add-on authentication software could be located at the firewall as opposed to on each system that needed to be accessed from the Internet. Other solutions to network security such as Kerberos involve modifications at each host system. While Kerberos and other techniques should be considered for their advantages and may be more appropriate than firewalls in certain situations, firewalls tend to be simpler to implement in that only the firewall need run specialized software.

Kerberos is an authentication system developed at the Massachusetts Institute of Technology (MIT). Kerberos is designed to enable two parties to exchange private information across an otherwise open network. It works by assigning a unique key, called a ticket, to each user that logs on to the network. The ticket is then embedded in messages to identify the sender of the message.

4.4 ROUTERS

When we think about network security issues today, immediate thoughts of the many credit card and personal information disclosures that seem to be highlighted as a regular occurrence in the media come to mind. Routers are rarely if ever mentioned as having played any part in a high-profile cybercrime incident. Routers most often silently play a large part in the attack reconnaissance, providing the hacker with a detailed understanding of the layout and configuration of the intended victim's network.

The router for a given network essentially can provide the hacker with a complete road map of the intended victim's internal network that the hacker can later use to facilitate the attack. Historically, routers have long been the enablers of an attack but were not necessarily directly attacked themselves.

The only legacy exception, of course, is a denial of service (DoS) attack whereby the intent of the attack was to disable the victim's entire network.

If the router has been hacked, one should treat it as one would any other component involved in a digital forensic investigation. First, determine whether an incident response plan is available within your organization that covers, in detail, any incidents involving a router. If you are one of the fortunate few who have a detailed incident response plan which includes forensics for both volatile and non-volatile data within a router, this would be your road map of how to proceed.

4.4.1 Initial Steps

Although it may be tempting to simply dive in to the router to try to quickly determine what happened it is important to follow a few initial steps:

1) Interview the POC (Point of Contact) to gain an understanding of the incident and to gather information, such as router passwords, to facilitate logging in to the router to collect information.

2) Define your incident response plan; document everything and change nothing.

3) Begin evidence collection. Do not begin examining the router directly; examine only copies of the evidence and never the originals. Further, any evidence collection method used must be repeatable. If the issue ends up in court the opposing council must be able to follow the methods you used to arrive at the same result.

4) Analyze the evidence and build your case. Your report should be limited to specifics and should not draw conclusions.

Router security starts with understanding the difference between the perimeter router and an internal router. Internal routers are not directly exposed to the Internet and therefore they have fewer concerns. The two most common types of routers that are internal to your infrastructure are Inter-VLAN routers and Core WAN routers. Inter-VLAN routers are used to separate broadcast domains (virtual LANs) and move directed Internet Protocol (IP) traffic. Core WAN routers connect dedicated point-to-point connections between a central office location and remote offices. The hub and spoke topology used with Core WAN routers is disappearing as companies are using VPN connections to replace the point-to-point connectivity used here.

The perimeter router connects your company to the Internet. This router is exposed to billions of people worldwide, some of whom may not be nice.

4.4.2 Common Router Attacks

The three most common types of attack on the perimeter are:

- **Denial of service (DoS) attack** Using Internet Control Message Protocol (ICMP)
- **Telnet connection to the outside interface:** First using a packet capture program to obtain the enable secret password and then using the Telnet application
- **Simple Network Management Protocol (SNMP) attack** Enabled by default, on most routers can be exploited by people on the public Internet to gain information from the router

These three protocols should be blocked from the Internet to the perimeter router.

4.4.3 Procedure for Collecting of Volatile and Non-volatile Data

Although the incident that prompted examination of the router may not seem too illegal and a decision not to contact law enforcement may have been already made, the procedures used to perform router forensics should assume that the investigation could end up before a court of law.

Great care should be taken to use sound and repeatable practices throughout the examination to avoid having evidence omitted because of a lack of integrity of the evidence.

Create a bit-by-bit copy of the original evidence in such a manner that it does not change or alter the original evidence. It is critical that the examiner works with only the copy of the evidence and never the original to avoid any chance that the original evidence may be altered.

The copy of the evidence must be authenticated to prove that the copy is the same as the original. Hashing with Message Digest 5 (MD5) or SHA1 has become an industry accepted method for authenticating evidence. If the calculated MD5 or SHA1 hash of the original evidence matches the calculated hash of the copy of the evidence, it is reasonable to assume that the copy of the evidence is the same as the original evidence.

Examination of the copy of the evidence will differ based upon the circumstances of the incident. That is, in an incident involving the possible interception of a Voice over IP (VoIP) conversation more emphasis may be placed on possible alteration of routing tables while the investigation of a distributed denial of service attack (DDoS) attack against a router may require emphasis on the analysis of the connections to the router. As attack methodologies evolve, so must examination procedures; the potential for a rootkit operating within the flash memory of a router has become a reality.

You can connect to a router to collect its non-volatile data through the console or AUX port and over the network using Telnet, Secure Shell (SSH), Hypertext Transfer Protocol (HTTP) or Hypertext Transfer Protocol over Secure Sockets Layer (HTTPS). The direct connection approach of using the console port with an adapter cable and connecting to a laptop running Windows with the HyperTerminal application can be preferred. This approach requires physical access and simply may not be practical in all circumstances. A good alternative to a direct connection to the console port is to use SSH or HTTPS over the network, but the router must have already been configured to support that access methodology.

Serial Cable

Several routers, ship with the necessary management cable that connects the router to a PC serial port for configuration. If you are unable to locate the management cable, you can easily construct one. Connecting to the console port on a Cisco router with a serial cable to a serial port on a PC requires an adapter for the router's RJ-45 connection on the router and the PC's serial port. The adapter is referred to as an RJ-45 to DB-9 female adapter and is readily available from multiple suppliers on the Internet.

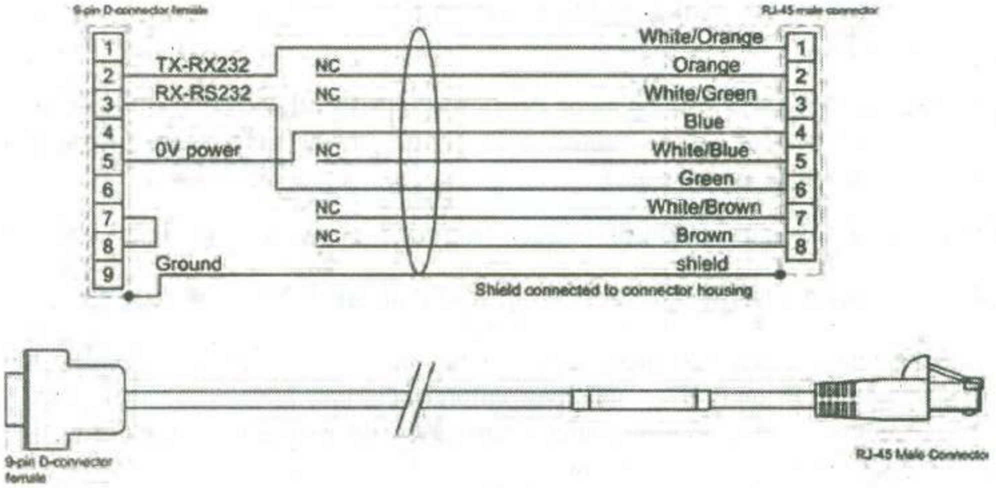


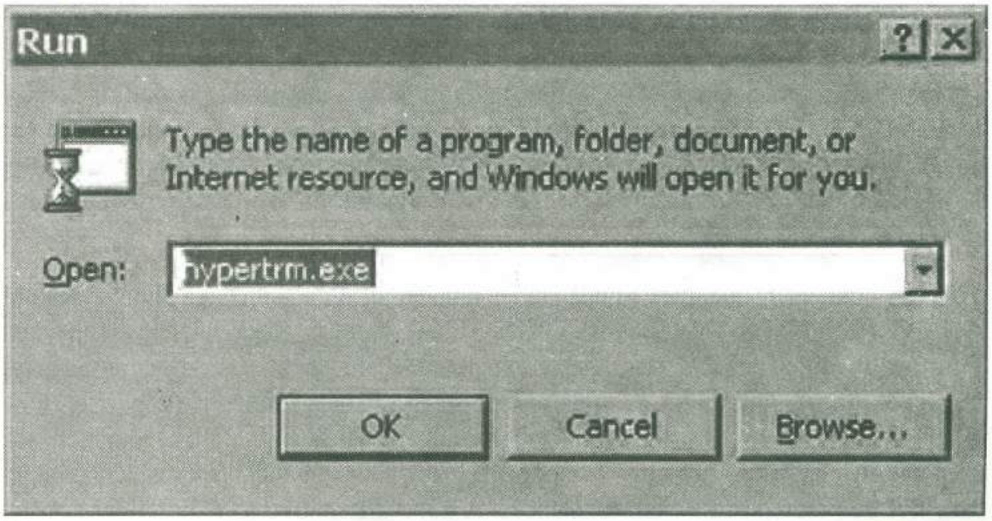
Fig. 1

USB Connection

As PC connectivity has evolved, it is not uncommon for a new laptop to be provided that does not have an available serial port. Hence, it may be necessary to use a converter to adapt a USB port to a serial port to connect the console port of the Cisco router to the laptop or PC. You can use a USB to DB-9 male adapter to connect to the Cisco management cable or a USB to RJ-45 adapter with a length of RJ-45 cable to connect to the console port of the router.

Hyperterminal

You can configure the Hyperterminal application that ships with Windows to be used as the interface to the router. Once you have connected the management cable to the router and the PC or laptop, you can start Hyperterminal from the Windows Start menu by selecting Run and entering hypertrm.exe;

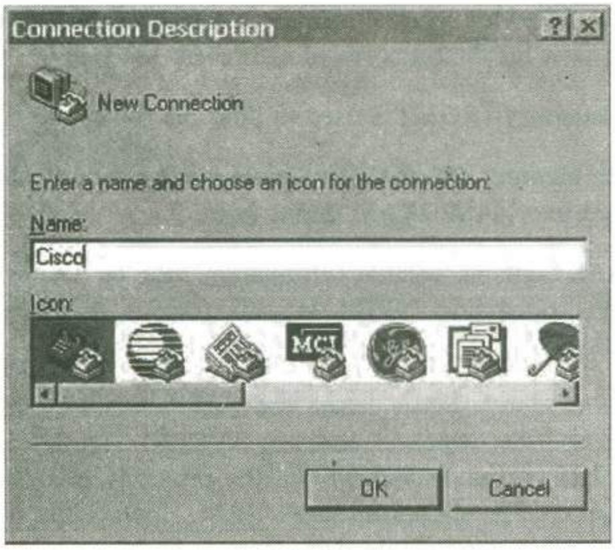

Run

Type the name of a program, folder, document, or Internet resource, and Windows will open it for you.

Open: hypertrm.exe

Fig. 2

Once Hyperterm is running, you begin by entering a connection description


Connection Description

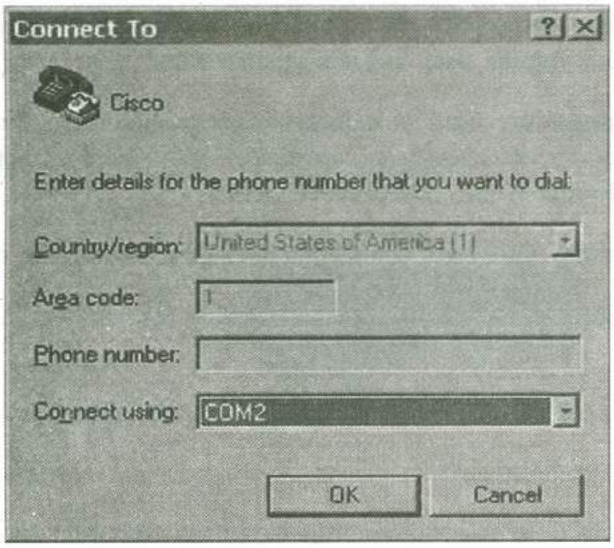
New Connection

Enter a name and choose an icon for the connection:

Name: Cisco

Fig. 3

Next, select the port to be used to communicate to the router.


Connect To

Cisco

Enter details for the phone number that you want to dial:

Country/region: United States of America (1)

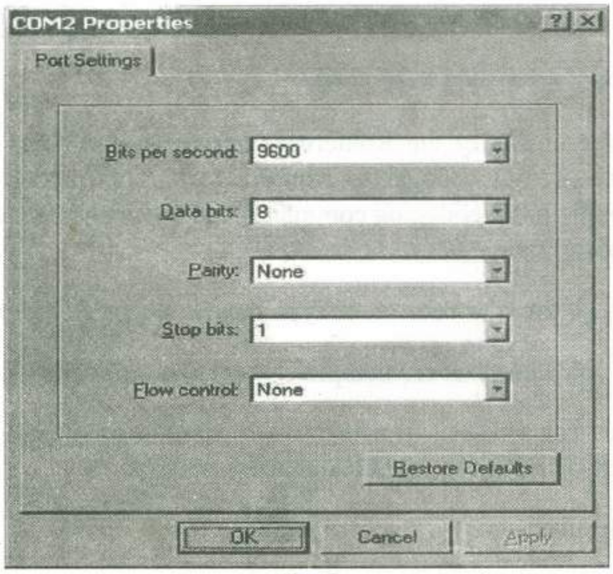
Area code: 1

Phone number:

Connect using: COM2

Fig. 4

Configure the correct communications properties for bits per second, data bits, parity, stop bits and flow control.


COM2 Properties

Port Settings

Bits per second: 9600

Data bits: 8

Parity: None

Stop bits: 1

Flow control: None

Fig. 5

Router Non-Volatile Data Collection Procedures

Routers utilize the following types of memory and this can aid in classifying volatile and non-volatile data for the purposes of a forensic analysis:

- **Read-only memory (ROM)** is used to boot the router into a minimal state.
- **Non-volatile memory (NVRAM)** contains a copy of the router configuration that is loaded into RAM/DRAM during boot.
- **RAM/DRAM** Contains volatile information such as a copy of the running configuration loaded from NVRAM, routing tables, statistics, local logs and the packet buffer.

For the purposes of clarity, let us define non-volatile data as that which is stored in flash memory and that which is stored in NVRAM. Hence, in an examination of non-volatile data on the router, one will be primarily interested in collecting copy of the stored configuration(s) from NVRAM. However, in the collection and analysis of the router's non-volatile data, there is some volatile data that should also be collected to aid in authenticating the non-volatile data that is collected from the router. That being said, it must also be noted that a collection of non-volatile router data represents only a part of a router forensic analysis and the only way to get the complete picture of an incident involving a router is to include the collection of both volatile data and non-volatile data.

Although any procedure used in a forensic investigation of router non-volatile data will depend upon the specific circumstances of the incident, here is an example of a non-volatile router data collection procedure:

1) Start Hyperterm and verify connection settings to connect to the Cisco router.
2) Enable logging by selecting Transfer/Capture Text and when the dialog box appears
    a) Enter a path to a file for storage of all HyperTerm session information.
    b) Log in to the router:
        - Press Enter to get the prompt to appear when connected to the console port.
        - The first prompt will look like Routername>; the greater than sign at the prompt tells you that you are in user mode.
        - In user mode, you can only view limited statistics of the router.
        - Enter privileged exec mode.
        - Type enable at the Routername> prompt and enter your password. The prompt changes to Routername#. This mode supports testing commands, debugging commands and commands to manage the router configuration files.
        - To go back to user mode type disable at the Routername# prompt.
        - If you want to leave completely, type logout at the user mode prompt.
        - You can also exit from the router while in privileged mode by typing exit or logout at the Routername# prompt.

Begin data collection:

- Use the router show commands to verify date and time.

- Use the router show commands to gather information such as the name of the CONFIG_FILE, the currently running system image filename, the system software release version, the configuration register setting and other data to facilitate copying of the data.
- Use the router *copy flash* commands to copy the respective named files:
  - To copy use TFTP, File Transfer Protocol (FTP) or Remote Procedure Call (RPC), FTP or RPC server.
  - To copy the contents of NVRAM such as the system image file, use TFTP, FTP or RPC to copy the router NVRAM image to the respective TFTP, FTP or RPC server.
- Use the router verify /md5 command to authenticate the copied software image.

Analysis of Gathered Non-Volatile Router Data from a Router

It can perhaps be assumed that if you are performing a non-volatile forensic analysis of a router you are doing so because volatile data is not available or reliable. An example is that the router was rebooted to allow a password reset to gain access to the router. In rebooting, the router erases all volatile data and the running configuration is also replaced by the startup configuration.

If you are performing a non-volatile examination in combination with a volatile examination, your findings in the non-volatile examination will complement your volatile examination.

The limited information in a non-volatile examination includes:

- Startup configuration
- Any files backed up to NVRAM or flash
- External log files

Log Files

Router log files are valuable non-volatile evidence and in an incident investigation you should handle them like any other evidence:

- Make a copy of the original log files. Sign and date the copy.
- Create an MD5 hash of the log file to later prove it was not modified.
- Never work with the original; work only with the copy.

Of particular interest in router log files are failed authentication attempts that may indicate a brute-force attack on the router's administrative passwords and denied connections, especially those from outside the network as they may indicate potential unauthorized attempts to access network resources.

Other areas of concern really depend on the type of incident. Here is a starting point to go by (this is by no means a complete list):

DDoS Attack

There are many types of DDoS attacks. Your first clue is higher than normal traffic on any given protocol. In today's world of botnets a DDoS typically uses legitimate looking traffic, but enormous amounts of it, sent from an army of bots within the botnet. Today, gigabytes of legitimate-looking traffic can be sourced from a botnet and easily take down any intended victim.

PI breach

- Increased protocol usage indicating large file transfers outside the network.
- Regular data transfers to an unknown outside host.
- Unusual increased usage of a protocol. Malicious hackers know you are looking at protocols such as Simple Mail Transport Protocol (SMTP).

Data Leakage

Protection (DLP) mechanisms for outbound PI and often encapsulate their stolen data on obscure protocols (even in DNS lookups).

Sniffing

Traffic forwarded to an unfamiliar network segment.

Collecting Volatile Data from Routers

Attacks against routers are becoming increasingly common due to their position in the network and their criticality for the continued operation of interconnected systems. The main reasons routers are attacked include the following:

- They provide a way to conduct denial of service (DoS) attacks against the network.
- They are a platform from which to compromise other systems.
- They can bypass firewalls, IDSs and other security devices through route changes.
- They can act as a sniffer on a network monitor.
- They can intercept and modify traffic.

When evidence is described as being volatile in nature, it means the evidence will be lost if certain events occur, such as a loss of power, timeouts and natural system purges. Furthermore, information contained in the active physical memory of a router will be lost when you power down the device. In addition, static memory sources (such as flash memory) may be overwritten if an orderly shutdown is allowed to occur. Indeed, much of the information contained within a router that is related to a forensic investigation is volatile in nature. This can include dynamic route updates, Address Resolution Protocol (ARP) information, dynamic name caching and even logs.

Pre-investigation Tasks

Before accessing the device, you need to perform a few preliminary tasks to ensure success. Many organizations will not have all of the documents mentioned in the following list, but they will generally have many of them. Starting this process will allow you to see what you have and what is missing.

1) Determine the scope. What are you planning to investigate?

2) Determine the risk. What information is the most crucial and what will be lost first? Detail what your requirements are.

3) Why are you conducting the investigation?

4) Collect the system and network design documentation. You can break this down into the following components:

   a) **System logical/infrastructure diagram** shows the system components in enough detail to support the Concept of Operations document.

b) **Concept of Operations document** details the purpose of each system (i.e. what each system does/provides).

i) How the system fulfills that purpose (How does it work?)

ii) In what ways the system depends on other components (What parts of the system rely on these other components? Why do they rely on them? How?)

5) List the mandatory requirements

6) List the risk-based requirements

a) This is a map of the prioritized countermeasures based on the risks identified in the risk assessment, with specific reference to countermeasures designed to counter the specific risks.

b) Evidence is required that illustrates why the countermeasures are considered effective.

7) List the critical configurations

a) These are the critical configurations that should be checked or changed on a regular basis, to ensure integrity of the system. This list may include:

i) Device configuration (rule sets, object definitions, filter lists)

ii) System passwords and access methods

iii) Logging and monitoring systems

iv) How these configurations/settings can be most efficiently checked on a regular basis

8) Document the configuration in detail

a) This document should cover the detailed configurations of each component on the system. For non-security-enforcing devices, it should cover at least the following information for each component:

i) Hostname

ii) Network address

iii) Function

iv) O/S version and patch level (e.g. IOS version)

v) Application configuration settings

vi) User accounts (including enable/privileged accounts)

vii) Integrity testing settings

viii) Interface details

9) Collect detailed network diagrams, which clearly indicate the following:

a) Host names of all components

b) Network addresses of all components

c) Function of all components

d Network addresses of all network segments

e) Netmasks of all network segments

   f) Any virtual local area networks (VLANs) and virtual private networks (VPNs)

10) Collect policy documents, which most likely include an access policy

   a) At a minimum, the access policy should include:

      i) Services allowed being externally accessible by anyone, externally accessible by customers and externally accessible by external support providers

      ii) Services available to all internally connected clients

   b) The access policy should also describe access allowed between internal networks, especially networks that have different requirements for different levels of security.

   c) It should detail services that are allowed between internal network segments, as well as the services to allow on an individual basis, the services available only from the system management segment and the services available only from the system console.

11) Collect procedures and plans

   a) Change implementation procedures

   b) Operational support procedures

   c) Contingency plans (something could go wrong during the test)

By following the preceding process, you should be able to collect information that will allow you to understand the following:

- What your organization needs to allow and the services it uses to conduct business
- What level of security is needed to validly conduct business, including that which is permitted, denied and logged
- From where and by whom the connections and services are needed
- When testing services and systems over the network, the end result is an increased understanding of what is running. Any interaction with a device will change the volatile evidence the device contains. Do not waste this. Use this to create an understanding of what happened and why. Most crucially, document each and every step you take.

Obtain the Router Password

Whenever possible it is essential that you obtain both the access and enable passwords for the router.

If an attacker has changed the privileged password or the organization has lost the privileged password for the router, you may have to reboot the router to gain access. This situation is far from ideal, as any volatile information stored on the device will be lost.

If access is available to the router, but you do not know the privileged password, there are still some actions that you can take. Depending on the router model, version and configuration, you may be able to run some commands in non-privileged mode. It is important to know which commands you can run without a privileged password and which ones will require further rights. Some of the commands you can run (by default) on a router include the following:

- show clock
- show version
- show users
- show configuration
- show ip route
- show access list
- show arp
- show cdp
- show frame-relay

If the privileged password for the router is unavailable, gather as much information as possible before rebooting the router. Depending on the router model and version, a wide range of information may be available to you.

4.5 SWITCHES

They are the backbone of the LAN and can support many different types of network interfaces, including Token Ring, Ethernet [10 mbs (megabits per second)], Fast Ethernet (10/100 mbs), Gigabit Ethernet (1,000 mbs) and Asynchronous Transfer Mode (ATM) Fiber. They can even support router and firewall cards to increase your ability to split a LAN into subnets or provide for higher security. The command structure uses a series of "Set" commands to configure and control the switch. Distribution switches and servers are commonly connected to switches.

There are other switches like CISCO's IOS-based switches that use commands similar to those used for routers. These switches have very little expansion and are considered distribution-level switches. They connect to the core switches and provide connectivity to workstations, printers and WAPs.

Menu-based switches are fading from view quickly, as they support only 10 MBPS Ethernet connections. Cisco no longer supports these switches, but because they are still in the field, network designers/administrators should be familiar with them as well.

Securing network switches starts with physical security. Locked cabinets located in locked closets are the first and most important step. If someone can connect to the console port of the switch, the LAN is severely breached.

4.5.1 Switch Concepts

The bottom line is that switches have many major advantages over hubs in terms of efficiency and security in communications. Much has been learned of using network switching technology and both IT administrators and purchasers have a deeper understanding of the costs and benefits of using switches in an enterprise. Before we dive into the technicalities, let's start with some terms that will help us along:

- **Collision** Occurs when two hosts attempt to access (or transmit) on a shared medium at the same time, resulting in a collision of their frames.
- **Broadcasts** Refers to both Open Systems Interconnection (OSI) Layer 2 (data link) broadcasts where frames are destined to all hosts on a sub network and OSI Layer 3 (Internet) broadcasts where packets are destined to all hosts on a network. Layer 2 broadcast frames have a destination Media Access Control (MAC) address of FF:FF:FF:FF:FF:FF and Layer 3 broadcast addresses have

a destination Internet Protocol (IP) address that is set for the broadcast of that particular network (the address varies, so don't always assume that an IP address ending with 255 is the broadcast address).

- **MAC address** Refers to the hardware, Ethernet or burned-in address of an Ethernet network interface. It is composed of a 48-bit address in a hexadecimal string of characters that designate the manufacturer ID and a unique serial number for the device.

- **Host** For the purposes of this discussion, a computer with a network card capable of communicating on an Ethernet network.

- **Bridges** The predecessors to switches and switching technology. Bridges have limitations that switches improve on.

- **Frame** A unit that is applied to the OSI model that defines the size and composition of a stream of network communication. In terms of the Ethernet specification, it is basically composed of a source MAC address, a destination MAC address, protocol information and a data payload consisting of data from the upper layers of the OSI model.

4.5.2 Advantages Over Hubs

Not long ago, switches were considered an extravagance and the mainstream network product to deploy onto a campus network was a hub. In fact, easy-to-remember formulas allowed anyone to determine in what circumstance hubs should be deployed in a network.

The price of switches has come down and they are easy to find on most any retail shelf. This helps organizations to motivate to take the plunge and purchase more switching hardware. In fact, when comparing cost to performance improvement switches cost an infrastructure less money and offer more performance if properly used. Not every system is pushing 100 million or 1 billion bits per second in and out of the switch, all of the time. Think of the bandwidth in terms of slices. For example, say that at one moment you are nearly saturating the network with a database query request that goes out of the switch's upstream port to somewhere out of the office. The next moment the system is quiet; this is where one worker is using the bandwidth to download a video from YouTube while the other is working normally without getting to know the load on the network.

Because switches are using switched architecture to keep these two communications separate from each other, the finite amount of bandwidth is appropriately used. If this occurred on a cheaper set of hubs, both you and your co-worker would have been saturating the network, preventing each other from transmitting any packets and possibly causing frame collisions.

The other reason switches are a better investment in terms of efficiency compared to cost is at the heart of the switching technology built inside switches. Without getting into electrical and computer engineering concepts, switches are effective at keeping conversations that occur between two ports separated from any other ports or pairs of ports, without sacrificing the speed of transmission/reception or bandwidth. So, suppose you want to download that Adam Sandler video from your co-worker. Both of you will make maximum use of the bandwidth between you as long as you are on the same switch.

But now say that two other co-workers are busy downloading PC games from a game-sharing Web site using Hypertext Transfer Protocol (HTTP). If they happened to be on the same switch both sets of network traffic communications would not interfere with each other and this raises the efficiency of the workplace, at least on a theoretical facilitauon aspect.

Now, it's pretty tough to find a hub these days, let alone one that has more than a handful of ports and that helps when it comes to computer security. A hub is really a multiport repeater. Given that you are now on a basic network hub, now the network traffic that hits the wire when you send your database query request actually goes to every port and every workstation that is connected to the hub, possibly causing frame collisions. (Remember back in the old days, Ethernet transmission collisions occurred when two workstations transmitted their bits onto the network at nearly the same time over a shared medium unbeknownst to each other. This series of bits overlapped each other, resulting in a collision. Then every workstation had to cease "talking" for a short but random period of time until everything settled back down on the network.) Switches manage to keep the medium shared in such a way that broadcast frames are transmitted to each port of the switch, but unicast frames are not, in most cases. A switch has to broadcast a unicast frame when it does not know which port a destination MAC address is connected to, so it has to broadcast it to every port and when its port-to-MAC address table (known as the content-addressable memory or CAM) is filled and cannot accept more entries, it is forced to revert to the behaviour of a hub. Otherwise, it keeps switched conversations apart from each host that is communicating on the switch.

Since the conversations are separated from each other, it also means that our workstation cannot eavesdrop on or "sniff " the unicast traffic using a network analyzer because of the separation in most cases. However, sometimes you can configure a computer network card to accept traffic destined to anyone else (called *promiscuous mode*) as well as being physically located on either one of the switch's truck ports or traffic spanning port that was left unsecured.

4.5.3 Volatile and Non-Volatile Data Collection Procedures

Now that we know many different ways to access a switch, the next step is to understand how to capture the information that is contained within so that you can use it for analysis. By capturing this information, you will be able to analyze the extent of the intrusion, the consequences of the intrusion and hopefully the individuals who caused the intrusion.

Hyperterminal

Hyperterminal is probably your key interface for performing console connections to the switch and for use in Telnet sessions. Hyperterminal has a great feature that allows you to create a log file and capture every keystroke along with the output that occurs on the screen. However, it captures every keystroke, including mistakes, so be careful as you are typing, as your skill will be on display. To set a switch so that there is no page break, use the following command:

```
Switch#Terminal Length 0
```

Telnet

If you are accessing a switch remotely through a network connection, you will need a Telnet program to make the connection. Any Telnet program will do, such as the PuTTY program mentioned previously. Because you are gathering information, it is necessary that the program have a logging feature. Whatever program you use, make sure you have enabled the logging feature prior to connecting to the switch so that you can capture all of the information from the moment you start until the moment you end.

Web-Based Interface

You can use the Web interface to collect information from a switch. Because this information is presented in your Web browser, you can copy and paste it into a word processing program for analysis, save the whole page as a Web site to your computer or take screenshots to capture the information. Because it's best to be able to search this information, screenshots may be the best way to capture this information.

Network-Based Backup of Config Files

There are several different ways to collect the configuration files from a Cisco switch, including using a network-based backup option. We will discuss two different methods, one of which allows authentication and security of the files and uses a reliable connection-oriented transport protocol (namely, File Transfer Protocol) and the other which provides no mechanism for security and utilizes a connectionless protocol for speed (namely, Trivial File Transfer Protocol). In the following sections, you will learn how to configure, upload and download the files using these methods.

TFTP

You can use Trivial File Transfer Protocol (TFTP) to gather information from your switch, by copying your startup config or running config to a TFTP server. In addition, you can use TFTP to upload files to the switch and download files from the switch. The two files that you will probably download to the TFTP server during your forensic analysis will be the startup and running config files. The following code snippets show the commands to use to copy the config files to a TFTP server at 10.0.0.7. The system will give you some suggested names for the destination filename, but I suggest that you alter them to include the date and time the files were created so that there is a clear audit trail for later purposes.

To copy the running config to a TFTP server, use the following command:

```
Switch#copy running-config tftp:

Address or name of remote host []? 10.0.0.7

Destination filename [Switch-running-config-11-07-08-06-10am]?

!!

2013 bytes copied in 1.387 secs (1451 bytes/sec)
```

To copy the startup-config to a TFTP server, use the following command:

```
Switch#copy startup-config tftp:

Address or name of remote host []? 10.0.0.7

Destination filename [Switch-startup-config-11-07-08-06-10am]?

!!

2013 bytes copied in 1.387 secs (1451 bytes/sec)
```

FTP

Alternatively, you can use File Transfer Protocol (FTP) to upload or download configuration files and IOS files to an FTP server. Two steps are necessary: First you set up the system to log into an FTP server; then you initiate the transfer. Whereas TFTP uses User Datagram Protocol (UDP) and therefore the transmission is unreliable, FTP uses TCP, so you can use it in even bad network situations where packets could be lost. The following code snippets show the commands to use to set up a switch for FTP login using a username of "fred" and a password of "cisco." The transfer is then done to an FTP server located at 10.0.0.7.

To set the FTP username and password, use the following command:

```
Switch#config terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)#ip ftp username fred
Switch(config)#ip ftp password cisco
Switch(config)#end
```

To copy the running config to the FTP server, use the following command:

```
Switch#copy running-config ftp:
Address or name of remote host []?
10.0.0.7
Destination filename [Switch-startup-config-11-07-08-06-10am]?
!!
2013 bytes copied in 1.387 secs (1451 bytes/sec)
```

To copy the startup config to the FTP server, use the following command:

```
Switch#copy startup-config ftp:
Address or name of remote host []? 10.0.0.7
Destination filename [Switch-startup-config-11-07-08-06-11am]?
!!
2013 es copied in 1.387 secs (1451 bytes/sec)
```

4.6 WIRELESS ACCESS POINTS

One of the largest areas of network growth in the business world today is wireless access. The idea is that if you can work from your laptop from anywhere in the building, you can be more creative. However, the more wireless capability that you add to your network, the more likely you will have a data exposure incident. Even the best wireless security can be breached by intent hackers. The two major wireless standards in use today are Wireless Encryption Protocol (WEP) and Wi-Fi Protected Access (WPA). Each standard can be broken, given enough time and access.

The hardest thing to protect with wireless is the ability for machines to connect from outside your building. Proper shielding, along with measuring the strength of the signal so that it does not extend beyond the walls of the building, are two ways to prevent unauthorized access; however, the best method is to put a physical firewall between the WAP and the wired network so that people have to VPN into the wired network. This will ensure stronger security access for people on the wireless network.

Check Your Progress 2

Notes: a) Space is given below each of the questions for writing your answers.

b) Compare your answers with those given at the end of this Unit.

1) What is intrusion detection system? How does it differ from firewall?

....

....

....

....

....

2) What are the advantages and disadvantages of hardware and software firewalls?

....

....

....

....

....

....

3) How are switches more superior to Hub in a network?

....

....

....

....

....

....

4.7 LET US SUM UP

In this unit we examined the processes involved when collecting volatile data from a router in a forensically sound manner. This starts prior to even connecting to the device with ensuring that you have all of the required equipment at hand and with planning what needs to be collected. The chapter covered how you can connect to the router, what you need to record and the commands that are associated with this process.

We then moved to capturing a core dump of the router image for offline analysis and to using router audit tools to evaluate the configuration. We introduced to the five phases of gathering data of value to your investigation. These processes, when introduced to the incident we are investigating will help preserve the necessary data and aid in stopping the incident and potentially catching the attacker.

Performing a forensic analysis on non-volatile data on a router can reveal valuable information, but you must understand that the non-volatile data may not represent the actual running configuration that was in operation at the time of the incident. If the router was rebooted since the incident occurred, the running configuration was overwritten by the start-up configuration.

It is preferred that an analysis of non-volatile data be used to complement a volatile data examination, but know that in many cases this will simply not be possible.

Logs maintained on a server for a router may be your best source of what actually happened, but many unfortunately choose not to go to the administrative burden, bandwidth consumption and storage costs associated with logging. All too often, logging is minimized and the period for which logs are held is too short and the actual steps leading up to an incident are no longer available.

Remember the importance of the documentation regarding the incident. We saw what data to gather from the non-volatile information for the incident report.

We covered many subjects related to network switching technology. We looked at why switches have performance and security advantages over hubs. We also explained switching modes and how they allow frames to go from port to port quickly and reliably, as well as the several levels of checking that we can perform. Hubs cannot do this.

Switches come with numerous status indicators and it's important to understand their meaning. In addition, you connect to a switch in much the same way you connect to a router; the difference is that switches have one less asynchronous interface, but plenty of additional Ethernet interfaces.

We must understand commands that involve VLAN information to effectively employ and detect switches.

4.8 CHECK YOUR PROGRESS: THE KEY

Check Your Progress 1

Vulnerability assessment

Vulnerability assessment products (also known as scanners) perform rigorous examinations of systems in order to determine weaknesses that might allow security violations. These products use two strategies for performing these examinations. First, *passive*, host-based mechanisms inspect system configuration files for unwise settings, system password files for weak passwords and other system objects for security policy violations. These checks are followed, in most cases, by *active*, network-based assessment, which re-enacts common intrusion scripts, recording system responses to the scripts.

The results of vulnerability assessment tools represent a snapshot of system security at a point in time. Although these systems *cannot* reliably detect an attack in progress, they *can* determine that an attack is possible and furthermore, they *can sometimes* determine that an attack has occurred. Because they offer benefits that are similar to those provided by intrusion detection systems, they are included in the sphere of intrusion detection technologies and products.

Check Your Progress 2

1) 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.

   There are several ways to categorize an IDS:

   - **Misuse detection vs. Anomaly detection:** in misuse detection, the IDS analyzes the information it gathers and compares it to large databases of attack signatures. Essentially, the IDS looks for a specific attack that has already been documented. Like a virus detection system, misuse detection software is only as good as the database of attack signatures that it uses to compare packets against. In anomaly detection, the system administrator defines the baseline or normal, state of the networks traffic load, breakdown, protocol and typical packet size. The anomaly detector monitors network segments to compare their state to the normal baseline and look for anomalies.

   - **Network-based vs. Host-based systems:** in a network-based system or NIDS, the individual packets flowing through a network are analyzed. The NIDS can detect malicious packets that are designed to be overlooked by a firewalls simplistic filtering rules. In a host-based system, the IDS examines at the activity on each individual computer or host.

- **passive system vs. reactive system:** in a passive system, the IDS detects a potential security breach, logs the information and signals an alert. In a reactive system, the IDS responds to the suspicious activity by logging off a user or by reprogramming the firewall to block network traffic from the suspected malicious source.

Though they both relate to network security, an IDS differs from a firewall in that a firewall looks out for intrusions in order to stop them from happening. The firewall limits the access between networks in order to prevent intrusion and does not signal an attack from inside the network. An IDS evaluates a suspected intrusion once it has taken place and signals an alarm. An IDS also watches for attacks that originate from within a system.

2) For any firewall the advantage is that a properly-configured firewall will help shield your computer from outside hacker attacks while the disadvantage is that firewalls can be difficult to configure correctly, especially for novices. Incorrectly configured firewalls may block users from performing certain actions on the Internet, until the firewall configured correctly.

Advantages and Disadvantages of Hardware-based Firewalls

Hardware-based firewalls have some advantages due to their independent nature. Because the firewall is external to the computer, no additional software needs to be configured or running for normal operation. The hardware-based firewall also protects many computers at once, since it intercepts all traffic from the Internet before it enters the internal private network.

Hardware-based firewalls also have some inherent limitations. If your computer is a laptop and travels to other locations, the hardware-based firewall in your home cannot protect you. Although there are small travel routers designed for use on a wired hotel network, a hardware-based firewall can't protect your computer if the Internet is accessed wirelessly.

If you have an always-on Internet connection, a hardware based firewall should be the first device on your network before you connect your computer to the link. Hardware-based firewalls for home and home-office use are relatively inexpensive and easy to install.

Advantages and Disadvantages of Software-based Firewalls

Software-based or "personal" firewalls are often the last line of defence between you and the Internet. Since a software-based firewall is physically part of your computer, this protection follows you everywhere. This is especially useful if a laptop is used at home, work and wireless hot-spots.

Personal firewalls have the advantage of identifying which applications on the computer are creating security risks. If a worm infects your system and attempts to open your computer to the world, a software-based firewall will identify this new application service. The personal firewall will prompt you to confirm the new application or to prevent its use. Your personal firewall may be your first warning that a malicious program is attempting to use the network.

Software-based firewalls aren't the ultimate security tool, however. A personal firewall can't prevent viruses from entering your system through legitimate sources such as a web browser or through e-mail. An anti-virus program with constantly updated virus signatures must always be included in an overall security strategy.

3) A hub is a layer 1 device that floods all its ports with any traffic passed through it, i.e. all workstations have to process all IP packets. Also most hubs only support a fixed speed/duplex configuration.

In contrast, a switch is a layer 2 device that only floods traffic to the ports that need it. This conserves bandwidth and means workstations will not have to process each and every packet passing through the switch. In addition, switches support any combination of speed/duplex.

That's the advantages of a switch. Disadvantages of using a switch are:

1) They are typically more expensive than a hub.

2) They may require some configuration depending on your needs. Hubs don't typically require any configuration (though most switches are plug-n-play and should work out the box.)

Student Satisfaction Survey

Student Satisfaction Survey of IGNOU Students

Enrollment No.	
Mobile No.	
Name	
Programme of Study	
Year of Enrolment	
Age Group	□ Below 30 □ 31-40 □ 41-50 □ 51 and above
Gender	□ Male □ Female
Regional Centre	
States	
Study Center Code	

Please indicate how much you are satisfied or dissatisfied with the following statements

Sl. No.	Questions	Very Satisfied	Satisfied	Average	Dissatisfied	Very Dissatisfied
1.	Concepts are clearly explained in the printed learning material	□	□	□	□	□
2.	The learning materials were received in time	□	□	□	□	□
3.	Supplementary study materials (like video/audio) available	□	□	□	□	□
4.	Academic counselors explain the concepts clearly	□	□	□	□	□
5.	The counseling sessions were interactive	□	□	□	□	□
6.	Changes in the counseling schedule were communicated to you on time	□	□	□	□	□
7.	Examination procedures were clearly given to you	□	□	□	□	□
8.	Personnel in the study centers are helpful	□	□	□	□	□
9.	Academic counseling sessions are well organized	□	□	□	□	□
10.	Studying the programme/course provide the knowledge of the subject	□	□	□	□	□
11.	Assignments are returned in time	□	□	□	□	□
12.	Feedbacks on the assignments helped in clarifying the concepts	□	□	□	□	□
13.	Project proposals are clearly marked and discussed	□	□	□	□	□
14.	Results and grade card of the examination were provided on time	□	□	□	□	□
15.	Overall, I am satisfied with the programme	□	□	□	□	□
16.	Guidance from the programme coordinator and teachers from the school	□	□	□	□	□

After filling this questionnaire send it to:
Programme Coordinator, School of Vocational Education and Training,
Room no. 19, Block no. 1, IGNOU, Maidangarhi, New Delhi- 110068

MPDD-IGNOU/P.O.1T/Feb,2012

ISBN-978-81-266-5923-4