IJCSN International Journal of Computer Science and Network, Volume 5, Issue 3, June 2016
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 1.02

Improved Technique for Simulation of Digital Forensic Architecture Framework in Cloud

[1] Smita Kamble, [2] Sulabha Patil, [3] Rajiv Dharaskar

[1] Computer Engineering, Department of M.TECH CSE
Tulsiramji Gaikwad-Patil College of Engineering and Technology Nagpur, India

[2, 3] Computer Engineering, Department of M.TECH CSE
Tulsiramji Gaikwad-Patil College of Engineering and Technology Nagpur, India

Abstract - Digital forensic is part of forensic science that implicitly covers crime that is related to computer technology. computer and network forensics have evolved to assure proper presentation of computer crime evidentiary data into court. The main purpose of forensic is to identify the origin while maintaining the chain of custody in order to enable the legal process to take its due course. Digital forensics is essential for the successful prosecution of digital criminals which involve diverse digital devices such as computer system devices, network devices, mobile devices and storage devices. If any computer related incident happens, fundamental questions to answer are when and where the incident occurred and, from which device, system and geographic location did the incident originate. In this paper a cyber crime, digital evidence investigation requires a special procedures and techniques in order to be used and be accepted in court of law. Generally, the goals of these special processes are to identify the origin of the incident reported as well as maintaining the chain of custody so that the legal process can take its option.While digital investigations have recently become more common, physical investigations have existed for thousands of years and the experience from them can be applied to the digital world.

Keywords - *SQL Injection, K-Means Algorithm, Aprioric Algorithm.*

1. Introduction

Since its inception, the field of digital forensic has not been significantly changed. It originates in solving pragmatic acquisition and chain of evidence problems related to investigations, performed by and large, by law enforcement personnel with little formal background in computing. The emergence of forensics comes from the incidence of criminal, illegal and inappropriate behaviors [3]. In general, the role of forensics can be classified in the following areas which are to facilitate investigations of criminal activities using forensic methodologies, [7] techniques and investigation frameworks. The areas are to preserve, gather, analyze and provide scientific and technical evidences for the criminal or civil courts of law; and to prepare proper documentations for law enforcement prosecution. In the digital forensics investigation practices, there are over hundreds of digital forensics investigation procedures developed all over the world [7]. Each organization tends to develop its own procedures and some focused on the technology aspects such as data acquisition or data analysis. The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations.

A good model of cybercrime investigations is important, because it provides an abstract reference framework, independent of any particular technology or organisational environment, for the discussion of techniques and technology for supporting the work of investigators.Such a model is useful not just for law enforcement [2]. It can also benefit IT managers, security practitioners, and auditors. These people are increasingly in the position of having to carry out investigations because of the increasing incidence not only of cybercrime, but of breaches of company policies and guidelines (e.g. the abuse of Internet connections in the workplace). Murder and rape suspects may, through a

search warrant, have their email and Internet activities analyzed to find evidence about their motives or hiding locations. Corporations investigate computers when an employee is suspected of unauthorized actions [9]. Fraud investigations collect transaction history evidence from servers. Murder and rape suspects may, through a search warrant, have their email and Internet activities analyzed to find evidence about their motives or hiding locations. Corporations investigate computers when an employee is suspected of unauthorized actions [3]. Fraud investigations collect transaction history evidence from servers. Furthermore, the model can be used in a proactive way to identify opportunities for the development and deployment of technology to support the work of investigators, and to provide a framework for the capture and analysis of requirements for investigative tools, particularly for advanced automated analytical tools [4]. At present, there is a lack of general models specifically directed at cybercrime investigations [12].

The available models concentrate on part of the investigative process (dealing with gathering, analysing and presenting evidence) but a fully general model must incorporate other aspects if it is to be comprehensive. The investigation of a computer or other digital device is also more similar to a physical crime scene investigation than a physical forensic analysis because of the amount of potential evidence [9]. A physical crime scene can be processed to identify many pieces of evidence. Blood on a wall is one piece of evidence and it can be analyzed to identify the owner of the blood, the type of object that struck the victim, the location of the victim, the location of the attacker, and the time of attack [2]. Similarly, a fingerprint is one piece of evidence that can be analyzed to show identity information and orientation information about how the person was facing.

2. Related Work

There are a myriad of existing digital forensic models some of which have been developed by organisations for their own use, or by law enforcement personnel for their own countries and even by other individuals based on their background, objective and even their employers' needs (Salemat et al 2008) and (Perumal 2009) [4]. These methodologies are in some part driven by the tools available to the investigator and focus on either the technical or legal aspects of the investigation [9]. However, there are other models that focus solely on the acquisition of the evidence ignoring all other phases that may be required by a "forensic" investigation [2]. These models all have positive and negative attributes most of which will be highlighted in this section.

2.1 Pollit Et Al. Methodology

One of the earlier models to be developed was the Computer Forensics Process by Pollitt (1995). This model is comprised of four stages and stresses the point that the digital forensics process should conform to the law while remaining committed to the scientific principles [7]. This model was however designed with the object of acquiring evidence from crimes committed in cyberspace and thus would need to be amended by the practitioner for use in other settings requiring such an acquisition.

2.2 Kruse & Heiser's Methodology

Kruse and Heiser (2001) was also one of the earlier models to be developed, though coming approximately six year after Pollitt's [12]. This model has three basic steps depicting the entire digital forensics process. The focus of this particular model is on the core aspects of digital evidence acquisition, acquiring, authenticating and analyzing the evidence. There is no mention of preparation, seeking authorization to acquire the evidence or identifying the evidence. Whereas these may have been assumed, as it seems with other models, to be discussed this is not enough especially where the legal issues are concerned [17].

2.3 H.C. Lee's Methodology

Also in 2001 H. C. Lee in his book 'Henry Lee's Crime Scene handbook' suggested a model that included an additional stage when compared to that of Kruse and Heiser. This model is more systematic and follows four very pertinent stages, which are recognition, identification, individualization and reconstruction [8]. This model is similar to the previous methodology proposed by Kruse and Heiser in that it assumes/ignores particular phases of the forensics process and does not include stages suggesting preservation or that of seeking authorization to access the evidence [11]. This model focuses mainly on the analysis of the evidence.

2.4 The DFRWS Methodology

The Digital Forensic research workshop (DFRWS) has also developed a model for the Digital Forensic process. This model is more extensive than the previous models highlighted [22]. It has seven stages and makes far fewer assumptions than the previous models covering integral stages not previously covered. However like a number of the other models, it ignores or assumes some of the legal aspects of the investigation and focuses more on the

technical aspects. It includes the stage "decision" which is somewhat out of the remit of the forensics process, which is concerned mainly with investigation and presentation of the findings.

2.5 Reith Et Al.

In 2002, Reith, Carr and Gunsch proposed a model that had a number of phases in which at least two phases overlap. This model is based on the one developed by DFRWS previously (DRFWS, 2002). The phases proposed include identify, prepare, approach strategy, preserve, collect, examine, analyse, present and return evidence [19]. This model, despite addressing some of the core areas of forensics, such as it does not include any suggestion of getting authorisation to preserve and /or collect the evidence, which is very important with regards to the legal aspects of any forensics process.

3. Proposed Methodology

3.1 SQL Injection

Causes-Simply stated, SQL injection vulnerabilities are caused by software applications that accept data from an untrusted source (internet users), fail to properly validate and sanitize the data, and subsequently use that data to dynamically construct an SQL query to the database backing that application [9]. For example, imagine a simple application that takes inputs of a username and password. It may ultimately process this input in an SQL statement of the form

string query = "SELECT * FROM users WHERE username = '" + username + "' ANDpassword = '" + password + "';

Since this query is constructed by concatenating an input string directly from the user, the query behaves correctly only if password does not contain a single-quote character. If the user enters

.

Impacts-Many of the high-profile intrusions in which SQL injection has been implicated have received attention because of the breach of confidentiality in the data stored in the compromised databases [17]. This loss of confidentiality and the resulting financial costs for recovery, downtime, regulatory penalties, and negative publicity represent the primary immediate consequences of a successful compromise. However, even sites hosting applications that do not use sensitive financial or customer information are at risk as the database's integrity can be compromised. Exploitation of SQL injection vulnerabilities may also allow an attacker to take advantage of persistent storage and dynamic page content generation to include malicious code in the compromised site. As a result, visitors to that site could be tricked into installing malicious code or redirected to a malicious site that exploits other vulnerabilities in their systems [23]. In many cases, exploitation of SQL injection vulnerabilities can also result in a total compromise of the database servers, allowing these systems to be used as intermediaries in attacks on third-party sites.

3.2 K-Means Algorithm

Simply speaking it is an algorithm to classify or to group your objects based on attributes/features into K number of group. K is positive integer number [2]. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data. The k-means method has been shown to be effective in producing good clustering results for many practical applications. However, a direct algorithm of k-means method requires time proportional to the product of number of patterns and number of clusters per iteration. This is computationally very expensive especially for large datasets [9].

K- means algorithm step wise execute is as follows

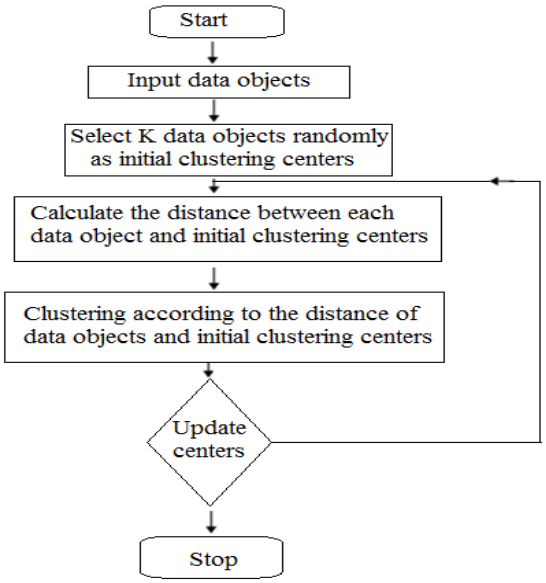

Fig 1. K-means algorithm

1. start
2. inpute data objects
3. select K data object randomly as initial clustering centers
4. calculate the distance between each data object and initial clustering centers
5. clustering according to the distance of data object and initial clustering centers
6. update center
   (repeated steps 4, 5 and 6)
7. stop

3.3 Aprioric Algorithm

Following the original definition by Agrawal the problem of association rule mining is defined as: Let **I = {i1, i2, ..., in}** be a set of *n* binary attributes called *items*. Let **D = {t1, t2, ..., tn}** be a set of transactions called the *database* [2]. Each transaction in **D** has a unique transaction ID and contains a subset of the items in **I**. A *rule* is defined as an implication of **the** form **X_Y** where **X, Y ⊆ I** and _ ∩ _ = □**.** The sets of items (for short *itemsets*) **X** and **Y** are called *antecedent* (left-hand-side or LHS) and *consequent* (right-hand-side or RHS) of the rule respectively. To illustrate the concepts, we use a small example from the supermarket domain. The set of items is ***I* = {milk,bread,butter,beer}** and a small database containing the items (1 codes presence and 0 absence of an item in a transaction) is shown in the table below [7]. An example rule for the supermarket could be {milk,bread}=>{butter} meaning that if milk and bread is bought, customers also buy butter.

General Process-Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all *frequent itemsets* in a database.
2. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

While the second step is straight forward, the first step needs more attention. Finding all frequent itemsets in a database is difficult since it involves searching all possible itemsets (item combinations). The set of possible itemsets is the power set over *I* and has size $2n − 1$ (excluding the empty set which is not a valid itemset) [16]. Although the size of the powerset grows exponentially in the number of items *n* in *I*, efficient search is possible using the ***downward-closure property*** of support (also called *anti-monotonicity*) which guarantees that for a frequent itemset, all its subsets are also frequent and thus for an infrequent itemset, all its supersets must also be infrequent [15]. Exploiting this property, efficient algorithms (e.g., Apriori and Eclat) can find all frequent itemsets.

Aprioric algorithm step wise execute is as follows

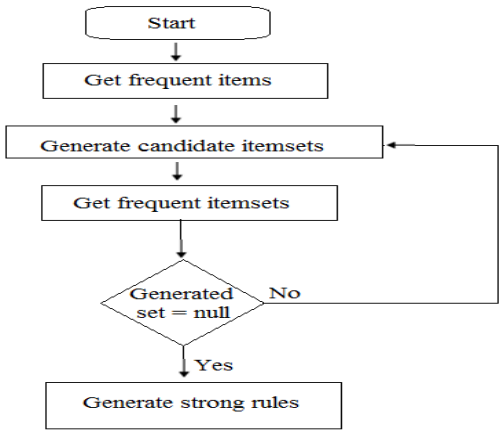

Fig. 2 aprioric algorithm

1.Start
2.Get frequent items
3.Generate candidate itemsets
4.Get frequent itemsets
5.Generated set= null
(repeated steps 3,4 and 5)
6.stop

4. Implementation of Propose Methodology

4.1. File Forensic

Most digital evidence is stored within the computer's file system, but understanding how file systems work is one of the most technically challenging concepts for a digital investigator because there exists little documentation. Now, security expert Brian Carrier has written the definitive reference for everyone who wants to understand and be able to testify about how file system analysis is performed. Carrier begins with an overview of investigation and computer foundations and then gives an authoritative, comprehensive, and illustrated overview of contemporary volume and file systems: crucial information for discovering hidden evidence, recovering deleted data, and validating your tools. Along the way, he describes data structures, analyzes example disk images, provides advanced investigation scenarios, and uses today's most valuable open source file system analysis tools - including tools he personally developed.

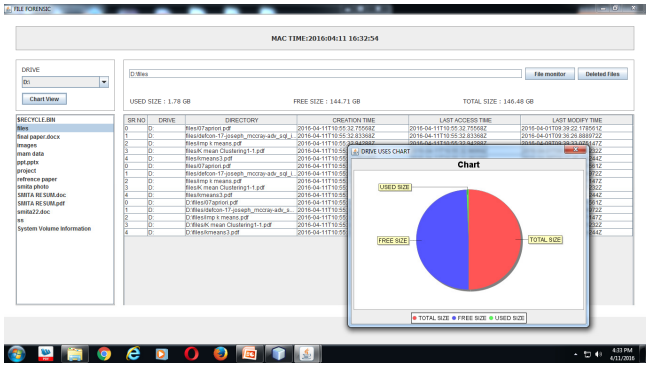

Fig. 3 file forensic

4.2 Network Forensic

Network forensics is a sub-branch of digital forensics relating to the monitoring and analysis of computer network traffic for the purposes of information gathering, legal evidence, or intrusion detection. Unlike other areas of digital forensics, network investigations deal with volatile and dynamic information. Network traffic is transmitted and then lost, so network forensics is often a pro-active investigation. Network forensics generally has two uses. The first, relating to security, involves monitoring a network for anomalous traffic and identifying intrusions. An attacker might be able to erase all log files on a compromised host; network-based evidence might therefore be the only evidence available for forensic analysis. The second form relates to law enforcement. In this case analysis of captured network traffic can include tasks such as reassembling transferred files, searching for keywords and parsing human communication such as emails or chat sessions.

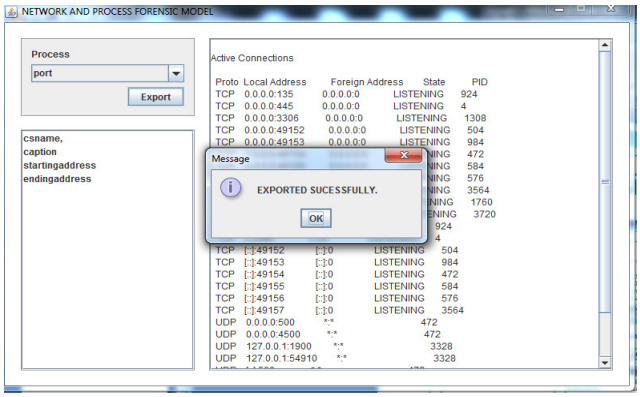

Fig 4. Network forensic

4.3. Data Mining

Data mining, *the extraction of hidden predictive information from large databases,* is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours , allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics.

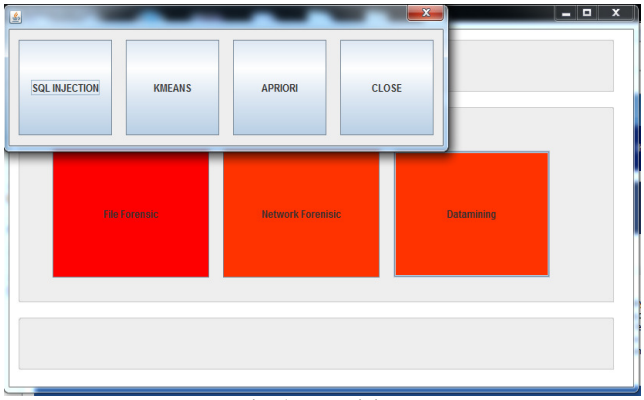

Fig. 5 Data Mining

A. SQL Injuction

Simply stated, SQL injection vulnerabilities are caused by software applications that accept data from an untrusted source (internet users), fail to properly validate and sanitize the data, and subsequently use that data to dynamically construct an SQL query to the database backing that application.

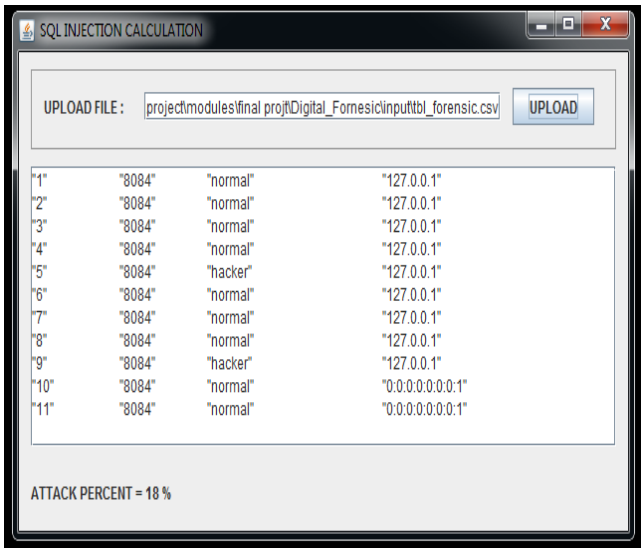

Fig. 6 SQL Injuction

B. k-Means

The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data. The k-means method has been shown to be effective in producing good clustering results for many practical applications.

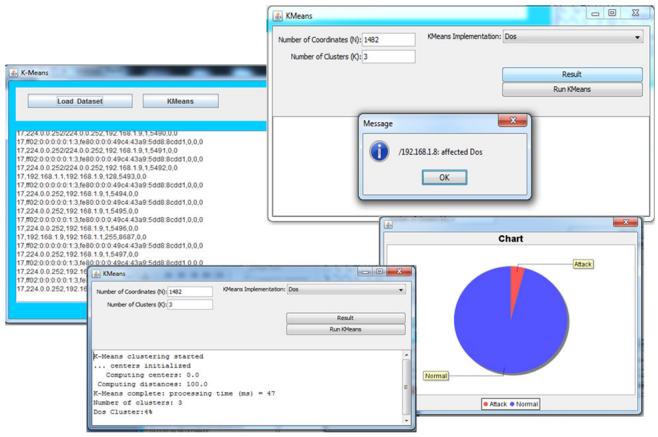

Fig. 7 k-Means

C. Apriori

Apriori is used to find the frequently occurred item in the data set. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation).

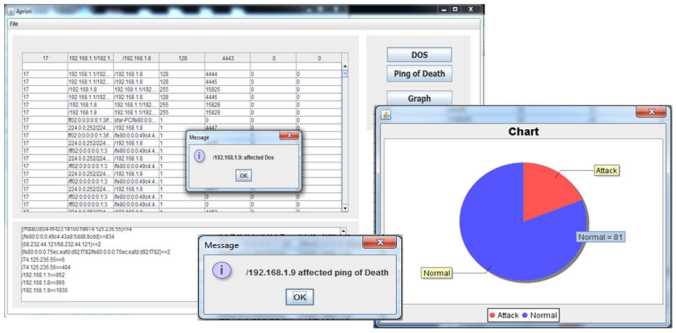

Fig 8 . Apriori

5. Conclusion

A new model of cybercrime investigations has been described. The inclusion of information flows in this model, as well as the investigative activities, makes it more comprehensive than previous models. It provides a basis for the development of techniques and especially tools to support the work of investigators. The viability and applicability of the model now needs to be tested in different organisational contexts and environments. A standardized methodology (way of working) will be of benefit to all involved in the world of digital forensics. The definition of a framework that includes all aspects and core fields that are involved in the digital forensics process will help to alleviate some of the issues that exist within the discipline at present. It has been identified that although several subject areas are impacting on the field there is no collaborative and integrated approach.

Digital forensics is a wide area and thus all professionals that are impacted must be able to communicate eliminating “area specific” jargon and assumptions that one field is more important than the other. Computer Scientists must accept that to be digital forensics practitioners they must become knowledgeable of the different laws that are related to the field. Legal experts must accept that digital forensics is more than just using a particular tool and become knowledgeable of the digital forensics field. Law enforcement officers must be cognisant of both of the above. Organisations must be made to realise though they may “forensics ready” (if there is such a term) and have various security personnel in place within the organisation it is not enough to use the Information technology department/technician to investigate a digital crime.

Appendix

File forensic

Most digital evidence is stored within the computer's file system, but understanding how file systems work is one of the most technically challenging concepts for a digital investigator because there exists little documentation. Overview of investigation and computer foundations and then gives an authoritative, comprehensive, and illustrated overview of contemporary volume and file systems.

Network forensic

Network forensics is a sub-branch of digital forensics relating to the monitoring and analysis of computer network traffic for the purposes of information gathering, legal evidence, or intrusion detection. Unlike other areas of digital forensics, network investigations deal with volatile and dynamic information. Network traffic is transmitted and then lost, so network forensics is often a pro-active investigation.

Datamining

Data mining, *the extraction of hidden predictive information from large databases,* is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools can answer business questions that traditionally were too time consuming to resolve.

SQL injuction

SQL injection vulnerabilities are caused by software applications that accept data from an untrusted source (internet users), fail to properly validate and sanitize the data, and subsequently use that data to dynamically construct an SQL query to the database backing that application.

K-means

The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data. The k-means method has been shown to be effective in producing good clustering results for many practical applications.

Apriori

Apriori is used to find the frequently occurred item in the data set. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation).

DOS

In computing, a denial-of-service (DoS) attack is an attempt to make a machine or network resource unavailable to its intended users, such as to temporarily or indefinitely interrupt or suspend services of a host connected to the Internet. A distributed denial-of-service (DDoS) is where the attack source is more than one, often thousands of, unique IP addresses. It is analogous to a group of people crowding the entry door or gate to a shop or business, and not letting legitimate parties enter into the shop or business, disrupting normal operations.

Data

Data are any facts, numbers, or text that can be processed by a computer.

Information

The patterns, associations, or relationships among all this *data* can provide *information*. For example, analysis of retail point of sale transaction data can yield information on which products are selling and when.

Knowledge

Information can be converted into *knowledge* about historical patterns and future trends. For example, summary information on retail supermarket sales can be analyzed in light of promotional efforts to provide knowledge of consumer buying behaviour. Thus, a manufacturer or retailer could determine which items are most susceptible to promotional efforts.

Data Warehouses

Data warehousing is defined as a process of centralized data management and retrieval. Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access and analysis.

References

[1] Kent, K., et al., *Guide to Integrating Forensic Techniques into Incident Response, NIST Special Publication 800-86*. 2006, National Institute of Standards and Technology, Gaithersburg, MD.

[2] Palmer, G., *A Road Map for Digital Forensic Research*. 2001, Digital Forensic Research Workshop (DFRWS): Utica, New York.

[3] Carrier, B., *A Hyphotesis-based Approach to Digital Forensic Investigations*, in *Center for Education & Research in Information Assurance & Security*. 2006, Purdue University: West Lafayette. p.190.

[4] Casey, E., *Digital Evidence and Computer Crime*. Second ed. 2004 Elsevier Academic Press.

[5] Stephenson, P., *A Comprehensive Approach to Digital Incident Investigation*, in *Elsevier Information Security Technical Report*. 2003, Elsevier Advanced Technology.

[6] Oghazi, P., B. Pålsson, and K. Tano. *An attempt to apply traceability to grinding circuits*. in *Conference in Mineral Processing*. 2007. Luleå, Sweden.

[7] Golan, E., et al., *Traceability in the U.S. Food Supply: Economic Theory and Industry Studies*, in *Agricultural Economic Report* 2004.

[8] Clayton, R., *Anonymity and Traceability in Cyberspace.*, in *Computer Laboratory, Darwin College*. 2005, University of Cambridge. p. 189.

[9] Siti Rahayu, S., Y. Robiah, and S. Shahrin, *Mapping Process of Digital Forensic Investigation Framework.* International Journal of Computer Science and Network Security, 2008. **8**(10): p. 163-169.

[10] Carrier, B. and E. Spafford, *Getting Physical with the Digital Investigation Process.* International Journal of Digital Evidence, 2003. **2**(2).

[11] Ciardhuáin, S.Ó., *An Extended Model of Cybercrime Investigations.* International Journal of Digital Evidence, 2004. **3**(1).

[12] Beebe, N.L. and J.G. Clark. *A Hierarchical, Objectives-Based Framework for the Digital Investigations Process*. in *Digital Forensic Research Workshop*. 2004. Baltimore.

[13] Siti Rahayu, S., et al., *Advanced Trace Pattern For Computer Intrusion Discovery.* Journal of Computing, 2010. (6): p. 1-8.

[14] Prosise, Chris, et al. Incident Response and Computer Forensics, Second Edition. McGraw-Hill Osborne Media, 2003.

[15] Schiffman, Mike, et al. Hacker.s Challenge 2: Test Your Network Security & Forensic Skills. McGraw-Hill Osborne Media, 2002.

[16] Schweitzer, Douglas. Incident Response: Computer Forensics Toolkit. Wiley, 2003.

[17] Zalewski, Michal. Silence on the Wire: A Field Guide to Passive Reconnaissance and Indirect Attacks. No Starch, 2005.

[18] United States National Institute of Justice Technical Working Group for Electronic Crime Scene Investigation. Electronic Crime Scene Investiga-tion: A Guide for First Responders, July 2001.

[19] Ricci I. S. C. (2006) *Digital Forensics Framework that incorporate legal issues*. Available from www.sciencedirect.com Accessed on October 20, 2010

[20] Perumal S., (2009) Digital Forensics Model Based on Malaysian Investigation Process, IJCSNS Vol. 9 No. 8 Available from www.sciencedirect.com

[21] Salemat S. R. Yusof R. Sahib S. (2008) *Mapping Process of Digital Forensic Investigation Framework.* International Journal of Computer Science and Network Security Vol. 8 NO 10 Available from www.sciencedirect.com

[22] Garfinkel S., (2010) *Digital forensics research: The next 10 years. Digital Investigations 7.* 2010 S64-S73. Available from www.sciencedirect.com Accessed on August 20, 2010

[23] Ricci I. S. C. (2006) *Digital Forensics Framework that incorporate legal issues*. Available from www.sciencedirect.com Accessed on October 20, 2010

[24] R. Ahmad, Z. Yunos, S. Sahib, and M. Yusoff. (2012). *Perception on Cyber Terrorism: A Focus Group Discussion Approach*. Journal of Information Security, Vol. 03, No. 03, pp. 231-237. DOI: 10.4236/jis.2012.33029

[25] Prasad, K., (2012). *Cyberterrorism: Addressing the Challenges for Establishing an International Legal Framework*. In Proceedings of the 3rd Australian Counter Terrorism Conference, December 3-5, Novotel Langley Hotel, Perth, Western Australia.