OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

STUDY ON NEED OF DATA SANITIZATION AND SANITIZATION TECHNIQUES FOR MEMORY DEVICES

Deepika Chauhan[1] , Dr. Pratosh Bansal[2]
IT, IET DAVV, Indore, Madhya Pradesh, India[1,2]
chauhandeepika522@gmail.com, pratosh@hotmail.com

***Abstract*: The fright for security and privacy has increased significantly with the new updates which have occurred in the practice of network, electronic devices such as PDA, storage, digital I/O and Smart Cell Phones. With the passage of time, a device can have a large volume of stored data for a single user which may be of no use also. In addition to this, internet files, cache, cookies, images, video & audio files etc. may overriding the large space in memory of devices. So when a user or an organization changes or upgrades existing device, there is great need to delete existing data permanently. Criminals also delete files to destroy evidences. Deleting the old files, folders from the disk or from recycle bin or by disk formatting is not a guaranteed way to remove any data or file from the digital device. Data can be recovered with freely available large number of software. This imposes great threat to privacy of innocent user and the same time provides help to forensic investigators. Sanitization of a device means before changing or upgrading any storage media, absolutely removing complete data from it. Data can't be recovered with any data recovery software from sanitized device. However data sanitization is also considered as one of the method of anti-forensic. A study on need of sanitization of different memory or storage devices and the options which are applicable for sanitization has been carried out. The paper highlights various methods, tools and techniques, for hard drives and flash drives, for removing residual data and sanitizing them.**

Keywords: *Digital Forensic, Data Sanitization, Disk Wiping, Secure Eraser, Masking, Shuffling.*

---∴∴∴---

I INTRODUCTION

Digital Forensic is a branch of criminal forensics that deals with rendering and analysis of digital evidences, gathered from digital devices, using combination of different tools and technologies, in a systematic manner [1 2, 3]. Further, it is the applications of the law from science which collects preserves and analyzes the scientific evidences have been obtained during investigation [3, 4]. Nowadays, World is surrounded by the digital devices and hence the digital forensic investigation might have plenty of applications [5].

Evidences are collected from the degree of the investigation performed by going through the devices which are distributed to its all the sub-branches to refer to the mutual references for the digital devices involved. Presently, Microcomputer forensic has become a key part for the investigation. Forensic reference experiments and mobile stylistic allegory forensics can be used for investigation. For the same, examination of imaging of devices can be done to obtain fair results.

Computers might explain a 'scene of a crime' for instance by hacking [5, 6] or dissent of job attacks or they will help create a measure within the systematization of emails, documents or discretionary files which were involved in criminal activities like:

(a). Intellectual back forty stealing
(b). Industrial undercover work
(c). Employment disputes
(d). Fraud investigations
(e). Forgeries
(f). Bankruptcy investigations.
(g). Inappropriate electronic mail and net act by the whole of regard to within the employment place.
(h). Regulative compliance.

Primarily, forensics deals with all of the barrages and hit or miss of lurking proof. *"Computer forensics for the come up with the mixture of components of process and technology to save and equal taste from PC systems, networks, wireless information technology, and computerized information devices from one end to the other in a manner*

that's presentable as principle from one end to the other a purpose of law." Lawyers will challenge the proofs of validity once the status reaches into court [7, 8, 9].

The special forensic science branch called *Digital forensics* handles digital data retrieved from the digital devices like Disks, mobiles, smart electronic devices, PCs, laptops, etc. *Digital forensic* is basically the practices to collect, analyze & reporting of digital information & to support obtained from any digital device which is to investigated and preservation of the digital evidences and information for the future purpose which might be admissible legally [10, 11]. Mainly, the Digital forensics collects support by the documents like emails, histories, internet files, metadata about files like the previous modification time of update, use and modification. Advanced legal science is the branch of legitimate science which oversees automated media considered as cutting edge wrongdoing scene examination [12, 13]. Hence, study of all these gives us directions for analyzing, collecting, interpreting, protecting, and confirmation for reporting received through electronic devices. The aim of the mechanism is to save any type of the evidence in one of its most kind umbrella while playing with an examination by puting together, observing & confirming the pushed information for changing past events with the final objective [14, 15, 16]. One important point can be noted here that the confirmation is called *Digital Evidence* which is accumulated through cutting edge media.

II DIGITAL FORENSIC PROCESS

Investigation process of Digital forensic majorly includes the following phases [17].

1) ***Identification:*** Basically, It is a combination of three processes which encompasses validity of incident, generate list of operations required and generate flow of actions to be performed depending on knowledge.
2) ***Authorization:*** In this phase, an approval has been taken for the investigation to be done.
3) ***Preparation:*** In this phase, identify the required tools, generate steps to be followed, create investigation plan, identification of operation required and task allocations.
4) ***Securing and evaluating the scene:*** Under this, investigate the entities or devices and characters at the sight of crime and determine further possibilities, securing equipment, identifying and protecting evidence by conducting interview.
5) ***Scene Documentation:*** This is a very important phase in which, Information gathered from above mentioned steps are converted into documented proof by creating a record of documents which consists of photographs, notes, conditions, location information of devices, components that are removed or not, sealed and unseal able component, evidence bags and other facts.
6) ***Evidence collection:*** In this phase, collect all the evidence present physically such as digital device, analog evidence such as password, hand written notes, printouts, computer manual or any other referral physical entity related to crime.
7) ***Packaging, transportation and storage:*** After the evidence collection, generate feasible conditions for the evidence collected so that no alteration of information can be done such as protection of media and equipment during transfer, avoiding extreme pressure or exposure to conditions unfavorable for the device, effect of vibrations, magnetic source, elasticity etc. and maintenance procedure for storage and reception, chain of custody, copies of evidence and inventory for storage of media.
8) ***Initial inspection:*** This phase is the combination of several tasks such as device identification, tool selection, algorithm or process selection and expected outcomes.
9) ***Forensic imaging and copying:*** Under this process, physically remove hard drive from computer and generate a digital image for investigation and capture its behavior which can be done with many software tools available in the market. It is a combination of physical and logical acquisition.
10) ***Forensic examination and analysis:*** In this phase we are using different forensic science tools and techniques for processing which include creation of cryptographic hash values and its filtration with the help of hash libraries, viewing and exporting file and compound files (e.g. email) expansion, metadata extraction, indexing and searching.
11) ***Report and Presentation:*** In this phase we finally document the proofs obtained during analysis in which we cover procedure of investigation, findings, bookmarks, log files, notes. We generate the final conclusion based on the facts for further processing in court [18].

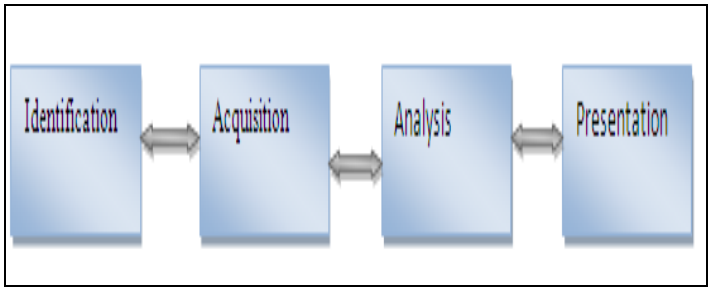

Figure 1: Process of Digital Forensic investigation [18, 20]

Figure 1 shows a typical Digital forensic investigation process which includes the major following 04 phases:

1. Identification
2. Acquisition
3. Analysis
4. Presentation

III ANTI- FORENSICS

Anti-forensics are the processes which are not in favor of computer forensics. In *Computer forensics* investigations, the purpose is to mine the facts or collect the digital evidence about any incidence or device which is then transformed as a proof for the justifiable purposes whether on contrast anti-forensics are the methods which works on the hiding of these raw facts. It is stated that the purpose of the anti-forensics are to alter or hide an electronic device so that it can become helpless in legal proceedings [15]. These methods make use of software utilities to delete files, delete registries, alter the timestamp of digital devices, deleting or altering logs using file/folder/volume encryption on a drives, uses bootable flash drives, CDs to alter data on digital device [15, 16].

IV TYPES OF ANTI-FORENSIC METHODS

There are fundamentally four methods which are used for anti-forensic purposes such as *(1) data hiding, (2) artifact wiping, (3) trial obfuscation* and *(4) attack against the computer forensics tools.*

4.1 Data hiding

A method called *Data hiding* is used to hide the sensitive data from any digital device which may exist there but in hidden layout. It is based on the principal of steganography. In this, digital information in any form can be stored in variety of carrier files such as executable/video/audio/image/ files. Data can also be veiled in slack space or on free space available on drives. We can hide metadata of many files on MBR (master boot record), unused space by drivers registers for device and tables, system area for protection, hard drives partitions in hidden form, BIOS itself & closed sessions on compact disks [17, 18, 19].

4.2 Artifact wiping

Second important method is *Artifact wiping* which is based on the principle of overwriting the digital media. Several software utilities which are sometimes known as data destructor/wiper/eraser exist. These are the software based utilities which are applied to completely overwrite the data present in any digital device [19]. These are the tools which delete slack space, log files, registries, clusters, cache files, browser history, and certain files of operating systems. They are available in open source or professional versions as: BC wipe, DBAN, ERASURE etc.

4.3 Trail obfuscation

Trail obfuscations are the methods for dealing with IP and MAC spoofing. Denial of Service (DoS) and Distributed Denial of Service (DDoS) Attacks are used in network for address spoofing. Onion routing so-called is used to make traffic analysis impossible [19]. Investigation of emails can be confused by falsifying the source of email's origin. This can be done using several techniques such as plant headers falsely, open proxies through Simple Mail Transfer Protocol (SMTP) Protocol and Secure Shell (SSH) servers' tunnel anonymously. Trail obfuscation can be accomplished with the help of wiping and/or changing log files at the server, event system files and date files [19].

4.4 Attack against the computer forensics tools

The last and very important category of anti-forensic methods are '*Attack against the computer forensics tools'* in which direct attack on the process of forensics can be applied which causes bigger harm in the process. In this, any phase of digital forensics can be altered using the utilities or attacks available in the market. On many of the computer forensics tools, the successful attacks include iLook FTK, Sleuth Kit, Encase & Win Hex. Codes which commit fraud by FAT directories, NTFS file tables, antinodes around for years, programs which write into the file slack, change the file signatures and flip bits for detection of evade hash [19].

V DATA SANITIZATION

Data sanitization lies under Anti- Forensic methods which either erases the complete records of information or displays masked information to protect it from unauthorized access. It is the procedure of purposely, for all time, and irreversibly expelling or the information put away on a memory gadget. A gadget that has been effectively disinfected has no leftover information notwithstanding when information recuperation is endeavored with cutting edge measurable instruments. Sanitization forms incorporate utilizing a product utility that totally deletes the information, a different equipment gadget that interfaces with the gadget being cleaned and eradicates the information, as well as an instrument that physically crushes the gadget so its information can't be recouped [18, 20]. Sanitization procedures are generally utilized by Government bodies where there is a need to keep up mystery among the authoritative records. Commonly they utilize a depended procedure which manages altering substance or disguising the delicate data in the first record rather, called *Data Redaction*, makes utilization of perceptive and realistic procedures which make unconscious of the data for the general crowd however on examination watchfully; the mystery piece might be uncovered [20]. Data sanitization can be defined as: *"Data sanitization is the process of deliberately, permanently, and irreversibly removing or destroying the data stored on a memory device. A device that has been sanitized has no usable residual data and even advanced forensic tools should not ever be able recover erased data. "*

VI NEED FOR DATA SANITIZATION

Data sanitization is very important. The following are the few important examples which are helpful in understanding the need of data sanitization:

(a). Selling and purchasing of digital devices.
(b). From one to another course, pass possession of treasure.
(c). When your choice goes for end up business?
(d). Warranty replacements of Hard disk or its update.
(e). Erase data completely from digital devices so that it could be reuse.
(f). Reallocate and reuse the cloud services provided.

VII DATA SANITIZATION TECHNIQUES

Data sanitization techniques are the methods which are used to destruct, wipe or delete data permanently from any memory device. They are sometimes referred as wiping standards which are different for different countries which are followed by some specific pattern or steps to make data unrecoverable [21, 22, 23].

Following Data Sanitization Techniques (Figure 2) have been proposed for removing content from hard disk drive.

1. Nulling out
2. Masking data
3. Substitution
4. Shuffling records
5. Number variance
6. Encryption/decryption
7. Degaussing
8. Physical destruction
9. Overwriting

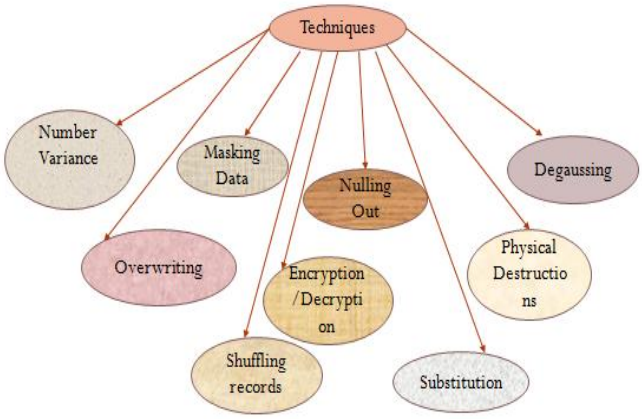

Figure 2: Data sanitization Techniques [20]

7.1 Nulling out

Nulling out is generally used for the database application. In this, the column from relational table is deleted by replacing the column data with NULL value to make the data invisible. But this is not as much of acceptable method as it does not fulfill the realistic world's requirement.

Example: If any customer is having any account and we put Null in fields such as name, address, contact number then it's difficult to write and test customer account maintenance form [23].

Result: The technique *nulling out* is useful in certain and very specific conditions but rarely due to the entire Data Sanitization strategy.

7.2 Masking data

Masking data is a technique in which data is replaced by some special symbols or characters which are called Mark characters (such as an X). This method provides an effectively way for differentiating the content as well as it preserves the basic structure, formatting and fonts of the report as it is.

Example: The common example of this type of data is credit card number.

4562 4213 6742 5467
4482 6766 5676 7888
4556 5667 5687 6778 [24]

And the information would appear after the masking operation as:

4562 XXXX XXXX 5467
4482 XXXX XXXX 7888
4556 XXXX XXXX 6778

7.3 Substitution

Substitution is the technique in which we randomly replace or fill the data with some random data which look similar to the stored information but completely irrelevant to that information [24].

This technique is effective in presenting the view of information and gives a similar look to the user as the data stored previously that is irrelevant to that information [25, 26]. Drawback of this method is long list of replacement data has to be maintained.

Example 1: To sanitize the contact details of customer we have to prepare a list of phone number which is very cumbersome task.

Another drawback of this technique is that it's hard to find the substitution data in large quantity for data sanitization.

Example 2: If we have million numbers of addresses of customers then alternative addresses for substitution require large number of addresses which is a cumbersome process and not feasible one.

Result: Conclusion is that it saves the look and feel of data but not efficient for the sanitization.

7.4 Shuffling records

Shuffling records is somehow similar to substitution. Here, generally, we shuffle the data between different locations [25, 26]. Shuffling from column to row and row to column can be done until there does not exist correlation between the existing data.

Drawback of this technique is that it is not efficient as there are chances of existence of original data in any form which may be recovered [26].Another limitation of algorithm is used for shuffling. If pattern of shuffling is identified then it is easy to recover the data.

Example: In the shuffling, identifying the shuffling pattern is easy and if we shuffle between any patterns of row and column. It's possible to recognize the swapping pattern and number of times the pattern repeats.

Another limitation is it is less effective for small volume of data.

Result: This method is easy to implement and fast to process but less effective for sanitization.

7.5 Number variance

Number variance technique is good for numeric kind of data with simply put the percentage value to any existing value [26]. It uses column method by some random percentage algorithm to interchange the values. This technique has lots of advantages such as masking of original data with numeric method by keeping all situations into a limit.

It can be better explored by example. Consider a column of sales data and there is need to mask them [25, 26]. A random variance of N% can be used to replace them for masking purpose. Here, it would generate some values with higher or lower range but all will be in the limit.

Result: This technique is an occasional technique and can only use for selective data type. It can't be used for alphabetic data like name or gender.

7.6 Encryption/Decryption

Encryption/Decryption techniques can be stated as the cryptographic approach. These are viable and are used to maintain the data security [26]. Data confidentiality deals with protection of data in the network. These techniques are used in many cryptographic algorithms for secure communication between the sender and receiver and also for secure information exchange. As confidentiality is a serious concern in networks, to accomplish the security in network we are applying symmetric or asymmetric cryptographic algorithm. The purpose is to use the same approach which maintains confidentiality at the end of storage. It states that an encryption can be used to store data into secure manner [26, 27]. This algorithm converts all targeted data into cipher text mode. At the time of recovery it can be explored by decryption.

Result: Confidentiality and privacy completely depends on the strength of algorithm key and encryption technique.

7.7 Degaussing

Degaussing is the process of eliminating a residual data on any digital device by applying magnetic fields. The term got its name after Carl Friedrich Gauss, a German mathematician. This is generally used on magnetic tape. It can also be used on USB, Hard drives, Smart phones or floppy disk. In this method, the device is demagnetized.

Result: Complete sanitization may not possible, Costly for modern drives. No verification mechanism available for identification of sanitization. Some part of the device possibly will get damaged permanently [27].

Table 1: Comparative study of all data sanitization techniques [26, 27, 28]

TECHNIQUE	DESCYPTION	ADVANTAGES	DISADVANTAGES
Nulling out	Replacing column data with NULL value.	• Prevent the data element visibility	• Fail in case of any application logic validation. • Not possible when reverse engineering is used.
Masking data	Replacing column data with masked character.	• Preserve the look and feel. • Very efficient in specific invariable format.	• Complex • Less effective
Substitution	Random replacement of data with similar but irrelevant data	• Quit powerful. • Preserve the look and feel.	• High efforts are required to identify substituted data and procedure for substitution.
Shuffling records	Random moving of data in row and column	• Fast and simple • Preserve look and feel.	• Ineffective for small data. • Original content is present. • Recovery is possible if algorithm is identified. • Work for numeric values only.
Number variance	Modification of each value of a column by real value with random percentage.	• Use for financial and date driven data fields. • Leave a meaningful range. Use for numeric data,	• Should be used with another sanitization option.
Gibberish generation	Eliminate all embedded reference to real data.	• Used for sanitize formless data such as notes.	• Not widely accepted.
Encryption/ decryption	Converting data into unreadable format or vice versa.	• Real data will be available to that person who has key.	• Security depends on strength of key • Man in middle attack possible. • Destroy look and feel.

7.8 Overwriting

Overwriting is an approach in which data blocks are replaced by some other blocks of meaningful data. This can be implemented from file level up to complete drive level [27].

Result: It securely deletes all data with feedback mechanism for the verification .It has two drawbacks, first time consuming process and creation of bad sectors.

A Comparative study of all data sanitization techniques have been shown in the Table 1.

VIII DATA SANITIZATION ALGORITHMS

The following data sanitization algorithms are based on the principle of *overwriting* techniques used for data sanitization. The number of passes varies from one algorithm to another [27, 28].

8.1 Guttmann

Guttmann algorithm is the software based method developed by Peter Guttmann in 1996.This method uses the principle of overwriting for the memory devices [26]. This uses a complex algorithm of 35 passes. It uses random characters for first 4 passes and last 4 passes after that it uses some complex pattern for passes from 5 to 31.This algorithm is designed for different types of hard drives. This algorithm ensures that none of data could be recovered if any device would be erased. Its performance is best against the Magnetic Force Microscopy. Using this method on latest drive may over kill because of the technology used in new drives which sometimes creates "Bad sectors" [27, 28].

8.2 HMG IS5 (UK)

HMG IS5 an originally found data sanitation method for various data sanitization software used in the United Kingdom. This will overwrite any file by 3 passes with the help of following steps:

Pass-1: Write a Zero ("0")
Pass-2: Write a One ("1")
Pass-3: Write any character randomly & verify the written character.

This is the fastest method for sanitization as it goes through only 3 passes for completely sanitizing the drive [27, 28]. But this is not useful for older of hard drives.

8.3 US department of defense sanitizing - DoD5220.22-m (7 passes)

The algorithm was accepted by the Department of Defense (DoD) of United State for sanitization purpose [27, 28]. Its implementation required three passes for complete sanitization of the drive.

Pass-1: Write a Zero ("0") & verify the written character.
Pass-2: Write a One ("1") & verify the written character.
Pass-3: Write any character randomly & verify the written character.

8.4 CANADA-CSEC ITSG-06

The sanitization method *CSEC ITSG-06* from CANADA, in which data sanitization is performed with the help of three (03) passes in the following ways:

Pass-1: Write a Zero ("0") or One ("1")
Pass-2: Write the complement of previously written character.
Pass-3: Write any character randomly & verify the written character.

IT Security Guidance 06, originally defined in Section 2.3.2: It is Clearing and Declassifying Electronic Data Storage (EDS) Devices were published by Communication Security Establishment Canada (CSEC) [27, 28].

8.5 AR 380-19

The sanitization method *AR 380-19,* in which data sanitization is performed with the help of three (03) passes specified below. This method is sometimes incorrectly used by the data destructor programs in which they skip the verification phase [27, 28].

Pass-1: Write any character randomly.
Pass-2: Write a character specifically.
Pass-3: Write the complement of character written specifically and verify the same.

8.6 PFITZNER

Pfitzner method uses 33 passes for implementation. This method uses random character for the process of sanitization [26, 28].

8.7 SCHNEIER

The sanitization method for data, *Schneier,* which is responsible for implementing secure data erasure with the help of the passes mentioned below. This method was created by Bruce Schneier. This method is sometimes modified by the programmers [27, 28]

Pass-1: Write a One ("1").
Pass-2: Write a Zero ("0").
Pass-3: Write a stream of the characters generated randomly
Pass-4: Write a stream of the characters generated randomly
Pass-5: Write a stream of the characters generated randomly
Pass-6: Write a stream of the characters generated randomly
Pass-7: Write a stream of the characters generated randomly
Some support verification after first and last pass.

8.8 RANDOM DATA

Random Data is a pass based method implemented by writing approach happened into multiple pass. It can be known as do it yourself method because it permits customization during the sanitization [27, 28].

8.9 WRITE ZERO

Write Zero is NULL based method implemented by writing Zero value as every instance [27, 28]. It can be known Zero method because it makes all data as Zero and they can't be

viable or accessible after the sanitization.

8.10."3+7+3" BEYOND DOD STANDARDS ALGORITHM (13 PASSES)

"3+7+3" is a thirteen pass sanitization method developed by USA defense department. First three passes use random patterns for replacement and next seven passes use standard passes developed by DOD. Last three passes again use random patterns for replacement. It is compressed but slow method for replacement [27, 28].

8.11. GOST R 50739-95(RUSSIA)

GOST R 50739-95 technique was developed by Russia and used to destruct the files by overwriting the existing information into storage media or HDD [26, 28].

It can be implemented by two ways can be listed as below.

Writing a Zero for sanitization

Writing a random character for sanitization

8.12.NCSC-TG-025(US NATIONAL SECURITY AGENCIES)

NCSC-TG-025 is similar to above method it is also software based sanitization method used to destruct the original content for security purpose. The basic objective of this method is to replace the software relevant files into HDD [28].

8.13 NCSC-TG-025 WIPES

NCSC-TG-025 WIPES can be implemented by two ways can be listed as below [28]

Writing a Zero for sanitization and perform verification

Writing a One for sanitization and perform verification

Writing a random character for sanitization and perform verification.

8.14 AFSSI-5020(US AIR FORCE)

AFSSI-5020 software was developed by US Air force to erase all secure content. Similar to previous method, it can be implemented by following ways [28].

Writing a Zero for sanitization and perform verification

Writing a One for sanitization and perform verification

Writing a random character for sanitization and perform verification.

8.15 AR 380-19(US ARMY)

AR 380-19 software was developed by US Air force to erase all secure content. Similar to previous method, it can be implemented by following ways [28].

Writing a random character for sanitization and perform verification

Writing a specified character for sanitization and perform verification

Writing a specified character with compliment for sanitization and perform verification.

8.16 US NAVY-NAVSO P-5239-26

US NAVY-NAVSO P-5239-26 software is developed by US Navy to erase all secure content. Similar to previous method, it can be implemented by following ways [28].Writing a random character for sanitization and perform verification

Writing a specified character for sanitization and perform verification Writing a specified character with compliment for sanitization and perform verification.

8.17 RCMP TSSIT OPS-II (CANADA)

RCMP TSSIT OPS-II is the Canadian technique to erase all private content. It can be achieved be several ways which can be listed as follows [28].

Step 1: Writing a Zero for sanitization

Step 2: Writing a One for sanitization

Step 3: Writing a Zero for sanitization

Step 4: Writing a One for sanitization

Step 5: Writing a Zero for sanitization

Step 6: Writing a One for sanitization

Step 7: Writing a Random Character for sanitization

8.18 VSITR (GERMANY)

VSITR is the German technique same as Canadian approach to erase all private content. It can be achieved by severe ways which can be listed as follows [28].

Step 1: Writing a Zero for sanitization

Step 2: Writing a One for sanitization

Step 3: Writing a Zero for sanitization

Step 4: Writing a One for sanitization

Step 5: Writing a Zero for sanitization

Step 6: Writing a One for sanitization

Step 7: Writing a Random Character for sanitization.

8.19 NZSIT 402(NEW ZEALAND)

NZSIT 402 technique is developed by New Zealand and also defined in NZICT Manual by government [26, 28].

Step: Writing a Random Character for sanitization

8.20 ISM 6.2.92(AUSTRALIA)

ISM 6.2.92 technique is developed by Australia and performs excellent performance for drive less than 15GB size [28].

Step: Writing a Random Character for sanitization.

8.21 GIBBERISH GENERATION

Gibberish Generation technique is based on the concept of referencing. It considers that if Sanitization will be pointless and we remove the selected data without masking or deleting the referenced data. For Example if algorithm is going to delete customer value there is needed to remove all referenced value such as transaction bill, memo etc Thus, this method not only removes specified data but also remove all relevant data can be state as reference data [27, 28].

Result: It is an occasional method and can be used for selected cases such as inventory system.

IX DESIGNING ISSUES IN DATA SANITIZATION

Before applying data sanitization over any memory device, we have to determine the different parameters which

might produce issues during the sanitization. Here we discuss about some issue which we carried in mind before starting the sanitization over any device:

(i) Level of confidence Identification

Different sanitization confidence levels have been provided through different disk sanitization techniques from higher to lower:

(i)Highest confidence - Physical destruction

(ii)Entire zeroing media

(iii)Affected file and its record zeroing

(iv)Lowest confidence-Deleting the affected file &its record.

(ii) Low level search

This is needed to verify that data is removed successfully or not.

(iii) Security

Complete sanitization is practically not possible so if there exists, a chance of residual which may contain some sensitive information.

(iv) Number of passes

Number of passes of an algorithm required to complete the sanitization, is an intricate matter, because with the increase in number of passes the probability of damage of media happen.

***(v)* Time**

How much time is required to run any algorithm?

(vi) Damage

There might be chances of permanent distortion of storage media during sanitization.

(vii) Cost

As the number of passes increases, Cost would be another issue.

(viii) Security level

Sensitivity and security level Assessing for the stor ed data.

(ix) Sanitization type

Selection of an appropriate media sanitization method based on the need.

(x) Sanitization method

For the type and media, selection of an appropriate media sanitization method [20].

Table 2: Data Sanitization Methods [28]

S. No.	TYPE	DESCRIPTION
1.	Disposal	Without sanitization, abandonment of media. Suitable if due to a loss of privacy of information will have no problem within the working of the organization.
2.	Clearing Information	Privacy for the information for the attack by keyboard is protected here. Overwriting is an important & acceptable method of it.
3.	Purging Information	Privacy for the information for the attack in the laboratory is protected here. On a disk drive, Executing the secure & erase firmware command and degaussing are very useful and all time acceptable purging method. Degaussing method is not very useful and effective on optical media storage devices like CDs, DVDs etc.
4.	Destroying Information	The main objective of it is to completely destroy the media. Multiple methods which include crumbling, burning, pulverize, melting and shredding can be accomplished. For destroying Optical media (e.g., CDs, DVDs) pulverizing, shredding or burning must be used.

X APPLICATIONS

Nowadays, Technology grows; every sector is digitalized used in digital devices for efficient working. Due to higher use of digital device requirement of sanitization will be increasing in all the sectors. Data Sanitization cover wider zone for their application as they are used in various sectors for the security concerns. The broad applications are mentioned here where applications of data sanitization are used.

(a). Government organization/Fiduciary Institutes
(b). Military organization
(c). Financial institute
(d). Pharmaceutical
(e). Data storage Institute
(f). Dot-Com

XI DATA SANITIZATION TOOLS

Data sanitization tools are sometimes referred to as data destructor/Erasers or wipers. These are the software based utilities which we are using for the sanitization of data. These are available in different formats and use different algorithms for the sanitization. They generally make use of the principle of overwriting the data. They are used to completely delete data

from any memory device. When we delete any file or remove it from the recycle bin, we actually delete only the reference of that file present in the allocation table of system's file system, not the actual file which can be easily recovered by some recovery tools available in market

Data destructors are the software based methods which utilize independent or combination of techniques defined for the data removal which permanently delete data from any memory Device

If you are planning to format or delete complete from any digital device, you will use any described algorithm based tools for deletion. There are four methods (Shown in Table 2) of media sanitization. Each of them is appropriate according to situations and provides levels of variable protections for information confidentiality available in the storage media [28].

XII CATEGORIZATION OF DATA SANITIZATION

In the form of strategic similarity, to express your views or ideas, *Categorization* is the technique. It is helpful in generalizing the sanitization branches into similar kinds of fields. In other words, we can say that it can be listed in the way as approach for data sanitization is deployed. The lists of all categories are as following and same is in figure 3:

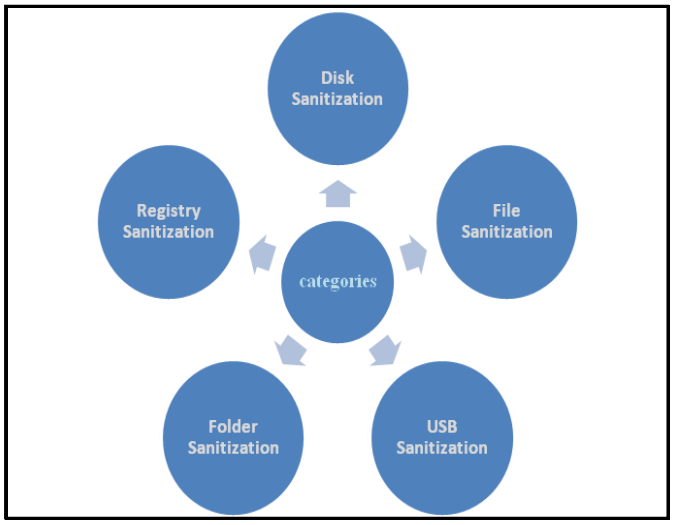

Figure 3 : Categories of data sanitization

(i) Disk sanitization

Disk sanitization is mainly responsible for sanitizing the secondary memory devices or the hard disk drives.

(ii) Registry sanitization

To remove clean, the registry history and its relevant information, *Registry sanitization* helps.

(iii) File sanitization

To remove, the history of files and their all relevant references, File sanitization method is responsible.

(iv) USB sanitization

To clean, the USB drives, removable devices and flash drives, *USB sanitization* method helps.

(v) Folder sanitization

A kind of modeling or collection data structure called *Folder* is that which help us in integration of multiple files and the folders in a single place. This is a very special kind of categorization which classifies all the methods and techniques which may have capability of removing all folders/sub folders and stored files completetly

XIII CONCLUSION

In this paper, comparisons have been done on the different techniques used for data sanitization along with their advantages and disadvantages. Overview of data sanitization tools have been given to apply it on different categories of digital media. This provides fine points concerning some open source tools available for sanitization of flash drives. The sanitized drive using the tools doesn't contain any residual data on the tools. Issues and parameters required before selecting any methodology for sanitization have also been discussed. Generally, all the tools use the method of *Overwriting* for the sanitization purpose as it is the best option for the recent version of drives. There may be a chance of *Bad Sector* creation when working with older version of drives due to number of Passes.

REFERENCES

[1] Chris Pogue, Cory Altheide and Todd Haverkos. "Unix & Linux Forensic Analysis Toolkit", edition-first published by Syngress Publishing, Inc. & Elsevier, Inc.,2008, Burlington-USA.

[2] Brian Carrier. "File System and Forensic Analysis", published by Pearson Education, Inc, 2005, New Joursey-USA.

[3] Paul Bakker. "Search Tools, Indexed Searching in Forensic Images", published by Sleuth Kit Informer 2004. Available on http://www.sleuthkit.org/informer/sleuthkit-informer-16.html#search.

[4] Simson L. Garfinkel. "Digital forensics research: The next 10 years", published by Digital Forensic Research Workshop-Elsevier Ltd.

[5] Blog: "Access Data. Forensic toolkit overview", http://www.accessdata.com/Product04_Overview.htm?ProductNum¼04; 2005.

[6] Saltzer Jerome H and Frans Kaashoek. "Principles of Computer System Design: an Introduction", 2009.

[7] Nicole Beebe. "Digital Forensic Research: The Good, the Bad And The Unaddressed".

[8] T. Abraham, R. Kling and O. de Vel. "Investigation profile analysis with computer forensic log data using attribute generalization" processing of 15th Australian joint conference on arterial intelligence,2002.

[9] E. Huebner, D. Bem, F. Henskens and M.Wallis. "Persistent systems techniques in forensic acquisition of memory, Digital Investigation" vol. 4(3-4), pp. 129–137, 2007.

[10] G. Dorn, C. Marberry, S. Conrad and P. Craiger. "Analyzing the impact of a virtual machine on a host machine, in Advances in Digital Forensics", V, G.

Peterson and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 69–81, 2009.

[11] Nance, K, and D J Ryan. "Legal Aspects of Digital Forensics: A Research Agenda", 2011 44th Hawaii International Conference on System Sciences, 2011.

[12] K. Bailey and K. Curran. An evaluation of image based steganography methods, International Journal of Digital Evidence, vol. 2(2), 2003.

[13] N. Beebe and J. Clark. A hierarchical, objectives-based framework for the digital investigations process, Digital Investigation, vol. 2(2), pp. 147–167, 2005.

[14] Deok-Soo Kim. "Piecewise Power Basis Conversion of Dynamic B-Spline Curves and Surfaces", Advances in Geometric Modeling, 01/29/2004.

[15] Richard Austin (2007). "Digital forensics" available at www.snia.org /1/9/2015.

[16] Eoghan Casey and Benjamin Turnbull (2011). "Digital Evidence on Mobile Devices" available at http://booksite.elsevier.com/9780123742681.

[17] YunusYusoff, Roslan Ismail and Zainuddin Hassan. International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 3, June 2011 "common phases of computer forensics investigation models" available at http://airccse.org/journal/jcsit/0611csit02.pdf. 25/11/2015.

[18] "Forensic Examination of Digital Evidence: A Guide for Law Enforcement" available at https://www.ncjrs.gov/pdffiles1/nij/199408.pdf 1/12/2015.

[19] Gary C. Kessler. "Anti-Forensics and the Digital" Investigator"http://www.garykessler.net/library/2007_ADFC_anti-forensics.pdf access on 8/10/2015.

[20] "Data Sanitization" available at www.stanford.edu / 28/8/2015.

[21] A Net 2000 Ltd. "Data Sanitization" available at /http://www.orafaq.com/papers/data_sanitization.pdf access on 10/12/2015.

[22] Net 2000 Ltd. "Techniques of sanitization" available at www.datamasker.com /4/9/2015.

[23] Kate Holmes (26 FEB, 2016). "Understanding the Impact of Anti-Forensics Techniques" available at http://www.ftitechnology.com/resources/blog/understanding-impact-anti-forensics-techniques.

[24] Tim Fisher (April 28, 2016). "Data Sanitization method" available at https://www.lifewire.com/data-sanitization-methods-2626133 access on 30/12/20015.

[25] "Data sanitization methods" (updated 9-Nov-2016) available at https://www.irs.gov/uac/media-sanitization-methods access on 3/1/2016.

[26] "Techniques of data sanitization" available at https://www.datamasker.com/datasanitization_whitepaper.pdf access on 13/1/2016.

[27] Kara Nance and Daniel J. Ryan. Proceedings of the 44th Hawaii International Conference on System Sciences - 2011 "Legal aspect of digital forensics" available at https://www.computer.org/csdl/proceedings/hicss/2011/4282/00/10-04-03.pdf access on 24/12/2015.

[28] "Degaussing" available at www.techtarget.com access on /24/6/2016. "Sanitization methods" available at http://googleweblight.com access on 15/7/2016.