

Fuzzy Crime Investigation Framework for Tracking Data Theft based on USB Storage

Ahmed M. Neil
Information Systems
Department, Faculty of
Computers and Information
Mansoura, Egypt

Mohammed Elmogy
Information Systems
Department, Faculty of
Computers and Information
Mansoura, Egypt

A. M. RIAD
Information Systems
Department, Faculty of
Computers and Information
Mansoura, Egypt

ABSTRACT

Since the lives of the persons are on the edge after being convicted in digital crimes. The main goal of digital forensics is to extract accurate evidence which determines whether the convict is guilty or not. The recent challenge is due to the big size of data that the investigator may deal with. These data stored in unnoticeable tiny devices such as USB sticks which may lead to a muddled decision because of the tediousness of the investigation. Fortunately, in Windows Operating systems, all users' transactions are stored in a central point which is known as Windows Registry. It stores all hardware and software configurations, user activities, and transactions. Therefore, digital forensics based on Windows registry is considered as a hot research field. This paper presents a proposed framework for digital crime investigation based on Fuzzy logic. It helps the investigator in the decision making phase about the evidence. This deals with the extracted evidence from relevant Windows Registry keys. Also, tracking the usage of USB devices for data theft was presented. Finally the proposed framework was tested on a simulated case study.

Keywords

Computer forensics, Crime investigation, Fuzzy logic, Data theft based on USB storage

1. INTRODUCTION

The enormous increase of data size and the need to transfer it safely and efficiently has become today demand for wide segment of people. Consequently, it has resulted in producing more powerful, faster, and bigger capacity portable data storage devices such as USB. The usage of the USB devices has become more popular in many fields such as workplaces, education institutions, etc. This is because of their small shape, large storing capacity, and data transferring rate. This encourages many people to use them to store and transfer their data, such as e-mails, corporate documents, third party sensitive data, etc. Due to the small size of USB's, users can carry them around without being noticed. Besides, this can sometimes make them suitable to carry out malicious activities, such as steal sensitive data and injecting malicious code from or to the system [1, 2].

A malicious activity is not sorted only on reading or writing a file from or to USB device. But it's far beyond that such as playing a pornography video, wiping corporate sensitive data stored in a local partition through bootable anti forensics USB stick, etc. Therefore, it is very difficult to acquire and analyze data from these tiny flash devices or drives [3]. Basically, some of them do not always permit the successful acquisition and recovery of all the data that have been stored on the devices [4]. In addition to that, the legitimate access to data or the digital resource that the insider gains allow him stealing

data which is very tricky to detect [5-7]. But what if a crime has took place and the forensics experts have to extract evidence from these devices for proofing. Is there any other way, which incase the investigators could not extract evidence from the USB device, to attain evidence from. To use the USB drive, you have to plug-in first it into computer through the USB port. Through the operating system the user can explore the content of this USB drive. Windows system automatically tracks user window viewing preferences [8]. It stores this activity in some areas such as Windows Registry. Since, the plethora information, such as user accounts, typed URLs, shared network, and run command history, are stored in it [9- 12]. The investigator can extract some useful data relevant to the case under investigation, such as the last accessed web site, the type of plugged in device, and some other data. These data need to be extracted, analyzed, and evaluated from forensic point of view. As a result, Windows Registry forensics is considered as a hot research field due to the huge amount of evidential information which it contains. This paper is organized as follows. Section 2 presents an overview on Windows Registry Structure. Section 3 spots the light on how to retrieve some USB device evidence from Windows Registry keys. Section 4 lists some respected related work. Section 5 describes the main building blocks of our proposed fuzzy crime investigation framework. Section 6 illustrates a stimulated case study. Section 7 demonstrates the usage of fuzzy logic in the investigation process. Finally, a set of conclusions are presented in section 8.

2. WINDOWS REGISTRY STRUCTURE

Windows Registry is constructed of five root keys (hives). They begin with 'HKEY' (an abbreviation of Handle to a Key). But really only two of these are 'real' Registry keys. The others are aliases or branches within one of the two hives as [13- 18] Windows Registry stores these data on a disk in several hive files. Just like a file system, Registry hive files contain clusters of data. In which these keys are dependant in functions. A lot of information discussed the basics of Windows Registry and how to analyze various Registry data, settings in [19-21]. However, there is another area in the registry that it plays a vital role in the investigation process. This area called "Shell bag" which contains Windows Explorer, such as folder window size, position on the screen and others. Shell bag can be found in the following Registry keys including two locations in each user's profile such as:

- HKEY_USERS\<USERID>\Software\Microsoft\Windows\Shell
- HKEY_USERS\<USERID>\Software\Microsoft\Windows\ShellNoRoam

The former key is used to store information related to remote folders and the latter key is about the local folders'

information [8]. Next an overview on other respected related work will be demonstrated.

3. RETRIEVING USB DEVICES EVIDENCE FROM WINDOWS REGISTRY KEYS

Windows Operating systems have a built-in method to audit USB events based on the changes to the registry. However, to implement in-depth auditing of USB device events; we must first understand what Windows records in the registry [14]. In

Windows Registry, when the user plug-in the USB device into the USB port, the Windows OS store automatically this activity in a relevant key in Windows Registry called USBSTOR key. To explore the content of this key; the investigator should be able to access the suspect regedit program to analysis Windows registry. Then, he should navigate to Computer\HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR key in the registry. Figure 1 illustrates the above mentioned steps to access USBSTORTYPESET TEXT.

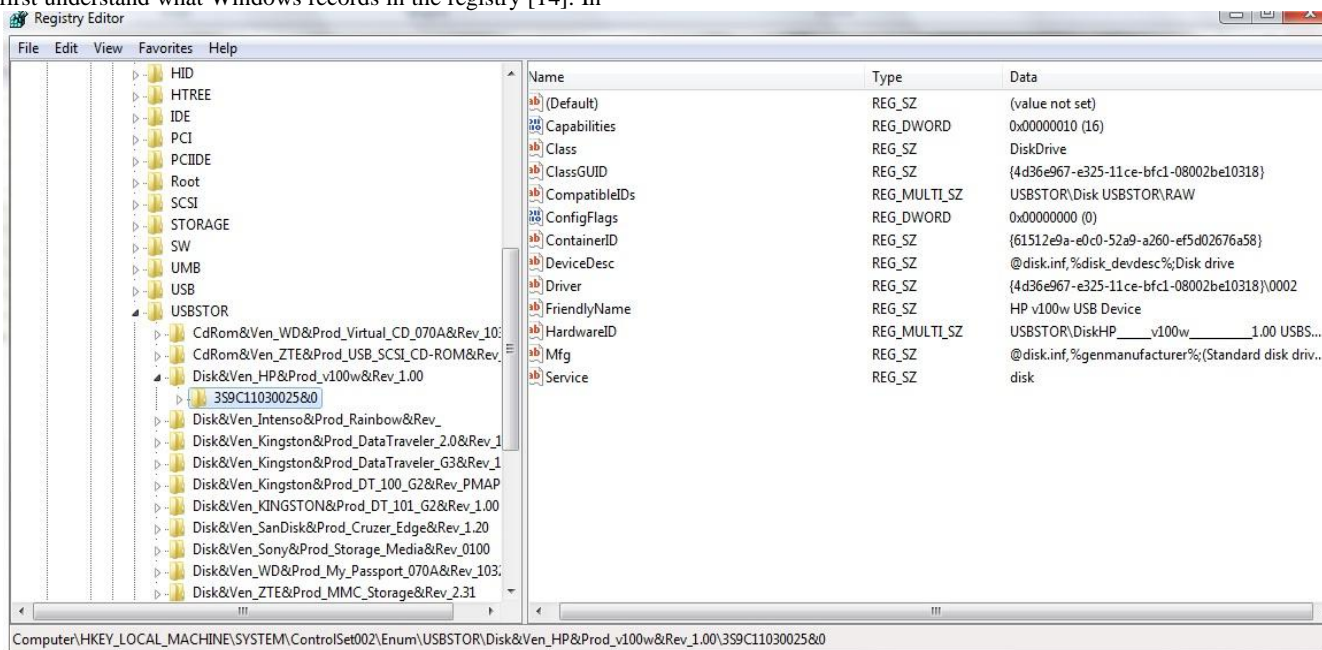


Fig. 1: Windows Regedit program shows the plugged in UBS sticks through USBSTOR Key from Windows registry.

Therefore, the investigator can now see that there are many USB devices had been plugged in to this computer. Among of these devices there is a USB stick under the vendor name HP v100w.

4. RELATED WORK

There is no doubt that digital forensic is hot research topic nowadays. This is because of the growth of digital crimes a rate that hits our daily bases infrastructure. Digital forensic is an active academic research area to improve procedures for solving digital crimes [22]. But some researcher's claims that there are some characteristics affect the investigation process. Such as the physical shape of the USB devices used in the crime. For example, Ashley et al. [23] stated that the tiny and adaptable nature of these devices make forensics investigators' tasks even more complex. As a result criminals use flash memory technologies to commit their illegal activities. As carry out valuable information or secrets that have been stolen from a business on USB flash. Basically, to obtain some evidence relevant to the case under investigation. The investigator has to use the proper tools that help him to extract this evidence. Not that only, due to the advancement in the digital crime. That enables the criminal to use multi session or multi operating system. Thus the investigator should use a sophisticated tool that can deal with this situation. Niandong et al. [24] proposed a system which can classify most kinds of attack types (91.5% correct classification rate on average) and also provided analyzable and comprehensible information for forensic experts. Chung et al. [25] introduced a new forensic

system based on several open source software to reduce cost and they enhanced autopsy's graphic interface. Moreover, they created a live DVD/USB for analyzing Microsoft Windows and Unix/Linux file systems. Additionally, they collected the volatile information of system by using live-analysis, which avoid lost of data due to shutdown of machine. Tanushree et al. [26] have illustrated in their work some methods for registry keys examination and generation through some areas relevant to the Windows Registry. In which through these keys, the investigator can trace the user activity regarding Data theft crime via USB. An approach in fuzzy logic based on an experimental result of a system.

5. PROPOSED FUZZY CRIME INVESTIGATION FRAMEWORK

Digital Forensics concerns to prove whether the accused person is guilty or not through Evidence. They should be collected and examined in a certain way. To solve a myth so the court of law can make a decision. That decision which consequent a person's life for being guilty or not. Thus the need for a science to help the decision makers to produce accurate decision was a must. As a result, we sought to use Fuzzy logic in this work to find out a technique to solve one of many digital crimes. Fuzzy logic which is a branch of artificial intelligence. It deals with reasoning algorithms used to emulate human thinking and decision making in machines [27]. The importance for this approach is to tackle the problem of the development of day to day technology. That led to the advancement in the malicious activity methods.

Therefore, enhancing intelligence-based approaches for law-enforcement and intelligence-gathering organizations became a necessity, particularly supported by the emergence of a new interdisciplinary research domain, the computational forensics. Computational methods find a place in the forensic sciences in three ways. First, they provide tools for the human examiner to better analyze evidence by overcoming limitations of human cognitive ability. Thus, it can facilitate the human examiner case work. Secondly, they can provide facilitation on large volume of data which are not human ability [28, 29]. All Fuzzy methods, such as fuzzy logic, Fuzzy sets, etc., play a vital role in learning complex data structure and classifying them to make an intelligent decision. This means, that the computational methods represented in fuzzy logic, can be a key player in any digital forensics field. Next a proposed fuzzy logic sets which are implemented on some crimes with their potential evidence will be presented. As shown in Figure 2, we proposed a fuzzy inference engine as a component in our model. After the evidence analyses phase in the final stage. The functionality of this component is to take outcome data from the analysis phase and test it on the fuzzy logic features. Then after the fuzzy data processing take place. Writing up the report including the findings should be done. This model was applied on different case scenarios which are shown in Figure 3.

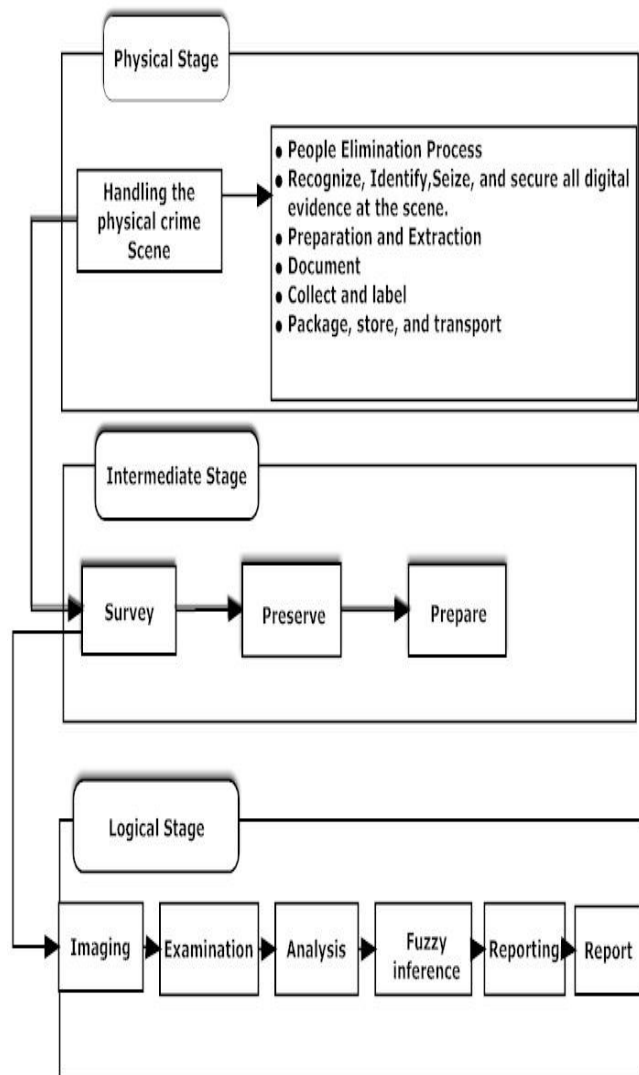


Fig.2: The proposed steps of extracting and analyzing the digital evidence

6. A CASESTUDY INVESTIGATION FRAMEWORK

In our previous work we proposed an enhanced framework to be used in any investigation process [30]. Since the investigation process might be a tedious work depending on the nature of the crime. Besides not all crimes are not equal neither in nature or the way it committed. The authors have divided their model into two stages Physical and logical. The purpose beyond this division is to focus more on each stage for accurate result.

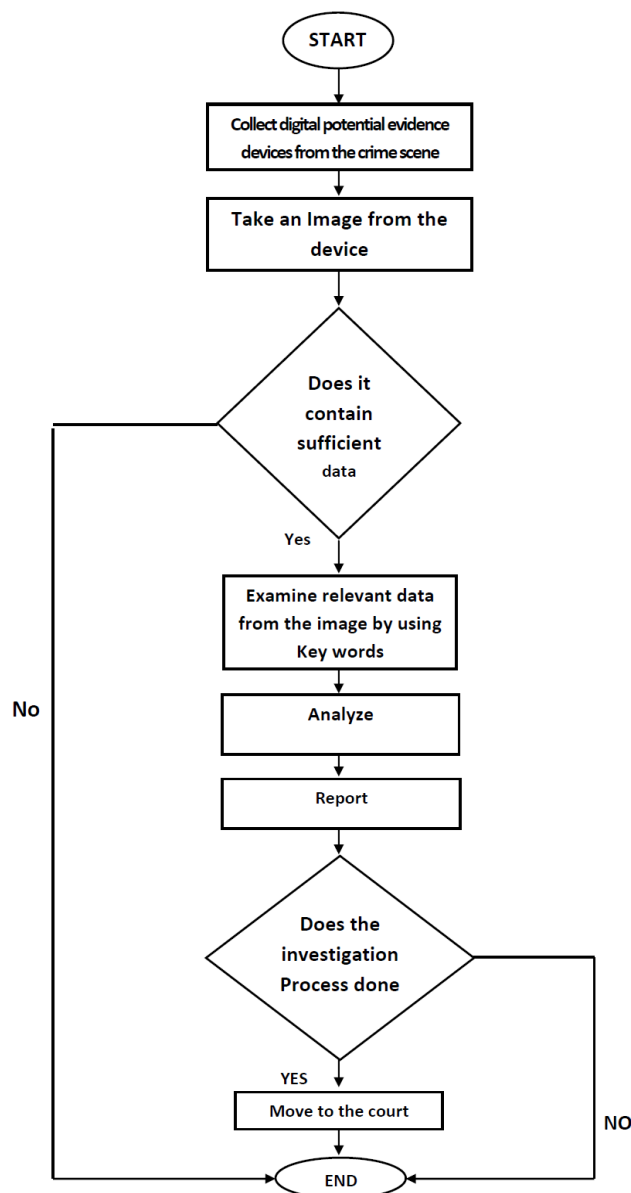


Fig.3: The flow chart for a general steps required for the investigation

1. Survey: the examiner should match the evidence received into the lab with the documentation taken in the physical crime scene.
2. Preserve: the examiner should unpack the evidence securely especially the fragile evidence and work on the evidence according to the priority.
3. Prepare: the examiner have to prepare the suitable tools to use in the examination process (i.e. imaging tools).

4. Imaging: The examiner will take an image of the current state of the digital device such computer, PDA, Mobile phones.
5. Examination: the examiner should review the image and use key words with techniques to extract evidence within the image.
6. Analysis: The examiner reviews the examination results for their value in the case.
7. Fuzzy interface engine: The investigator inserts the crime variable to expert system to help him/her in the decision making.
8. Move to the court: For testimony.

7. IMPLEMENTATION AND A CASE STUDY

Following is a case study stimulated on a machine where the crime took place. This crime start with a suspect who plugged his USB sticks in his colleague machine to steal his Marketing plan. After the suspect stole the marketing plan, he made some slight modifications on it and then handed it to the GM. The owner claimed that this plan seems to be like his own work. The suspect denied that and he said it cannot be because he handed the plan first. The GM asked a third party investigator to proof if that happened or not. The following steps were done by the investigator:

- 1- Take the suspect and the victim machines in custody.
- 2- Ask the claimer about the file and partition name on the hard disk.
- 3- Take an image from both the claimer and the suspect device.
- 4- Prepare the suitable tools to analyze the images.
- 5- Start to dig in the Windows Registry looking for the relevant keys to the crime under investigation. Such as:
 - HKEY_LOCAL_MACHINE\SYSTEM\ControlSet00x\Enum\USBSTOR\<device_class>\<device_unique_id>\
 - HKEY_LOCAL_MACHINE\SYSTEM\ControlSet00x\Control\DeviceClasses\{\<disk_devices_GUID>\<device_class#\<device_unique_id#\<disk_devicesGUID>\>\

- HKEY_LOCAL_MACHINE\SYSTEM\ControlSet00x\Control\DeviceClasses\{\<volume_devices_GUID>\<STORAGE_RemovableMedia#ParentId_Prefix#\<volume_devices_GUID>\
- HKEY_LOCAL_MACHINE\SYSTEM\ControlSet00x\Enum\Storage\RemovableMedia\<ParentID_Prefix>\

In order to know which file was moved, copied or opened. The investigator should go to
\\Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\

If the Marketing plan was saved in a PDF file format the following key can show that:

HKEY_CURRENT_USER\Software\Adobe\Acrobat Reader\5.0\AVGeneral\cRecentFiles

- 6- In case the investigator found something, then writing up a report should be done and submit it to the authority.

8. USING FUZZY LOGIC IN THE INVESTIGATION PROCESS

As stated before that using fuzzy logic in computer forensics. Is to help the investigators in the decision making process.

In our proposed frame work a fuzzy inference component was presented. Fuzzy inference can be defined as a process of mapping from a given input to an output, using the theory of fuzzy sets [27].

We used Mamdani inference engine which can be considered as the most commonly used fuzzy inference technique. It is performed in four steps: fuzzification of the input variables, rule evaluation, aggregation of the rule outputs, and finally defuzzification. We implemented the proposed component through examining four inputs and two outputs that includes six rules. Figure 4 shows overview of the Mamdani inference engine of our system.

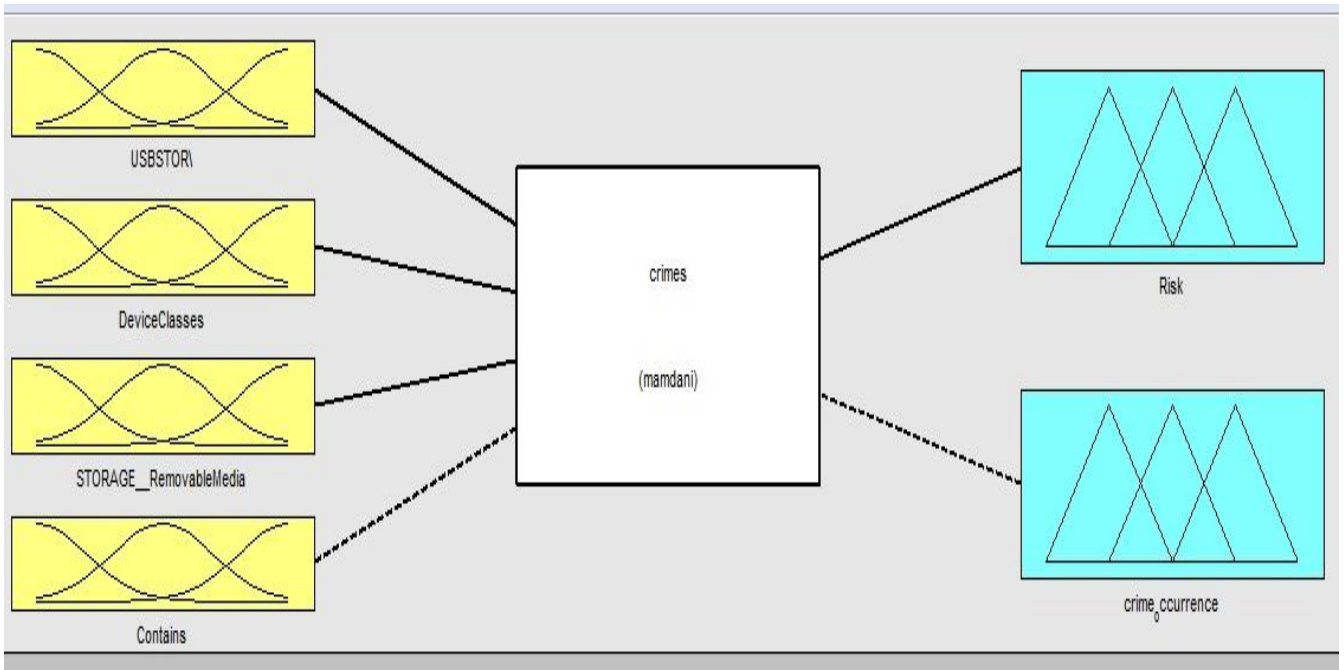
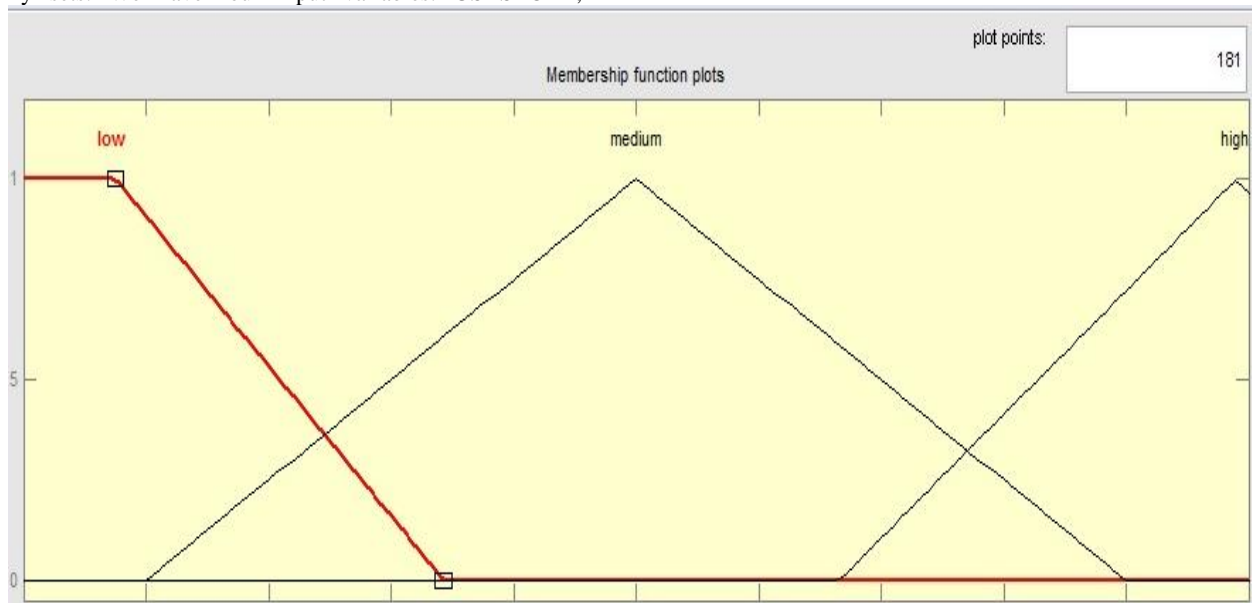


Fig.4 Membership functions input variables

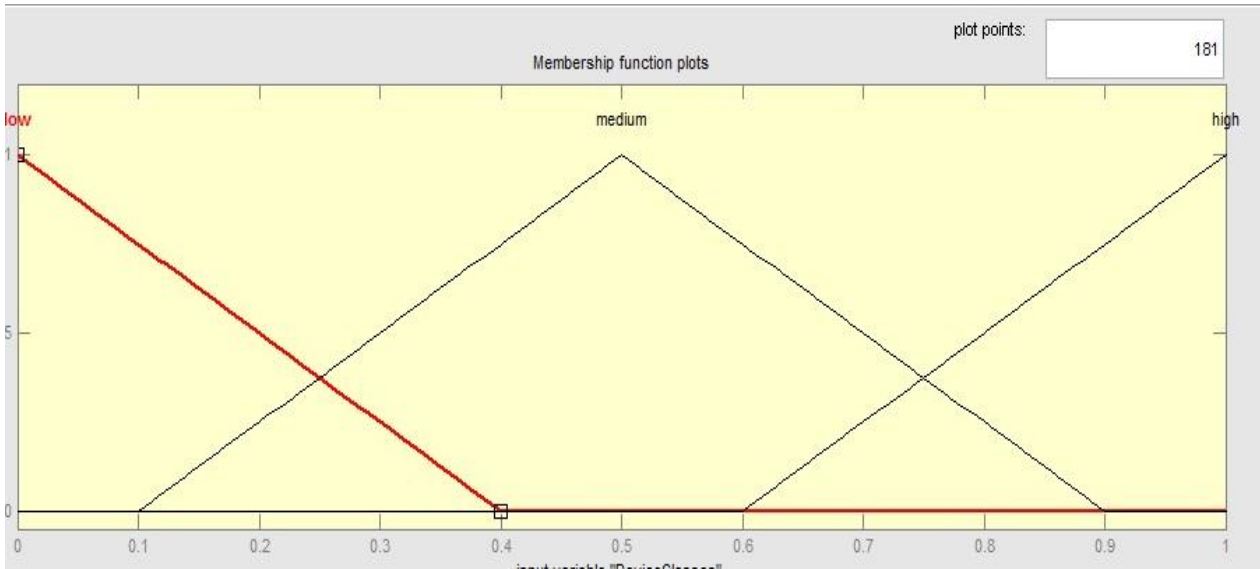
8.1 Fuzzification

The first step is to take the crisp inputs and determine the degree to which these inputs belong to each of the appropriate fuzzy sets. We have four input variables: USBSTORE,

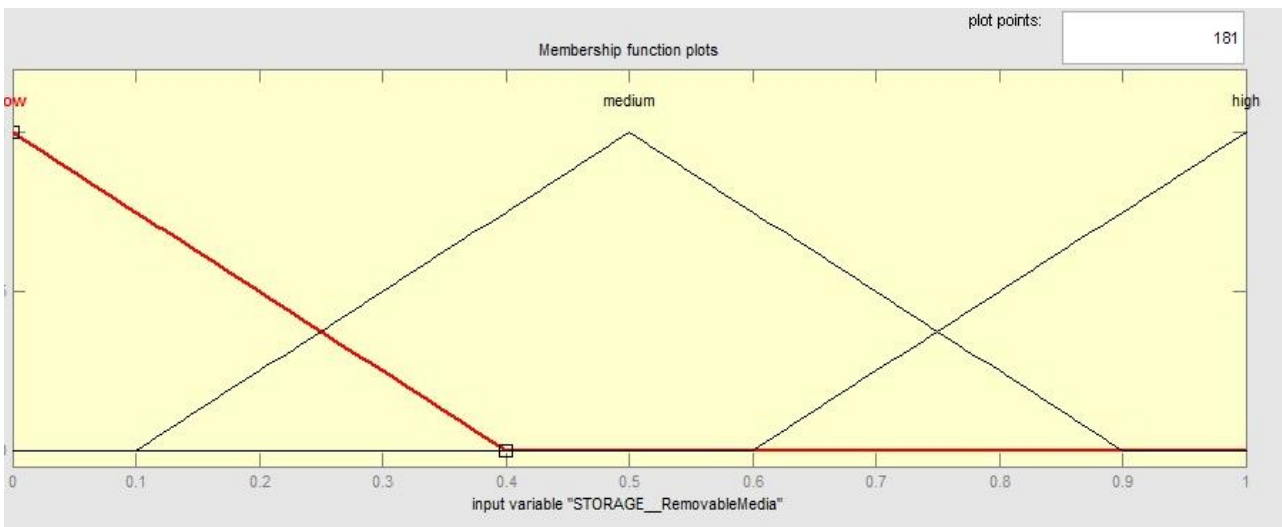
DEVICECLASSES, STORAG_REMOVABLE MEDIA, and Contains. We have two outputs: Risk and Crime occurrence. Figures 5 and 6 show the input and output variables.



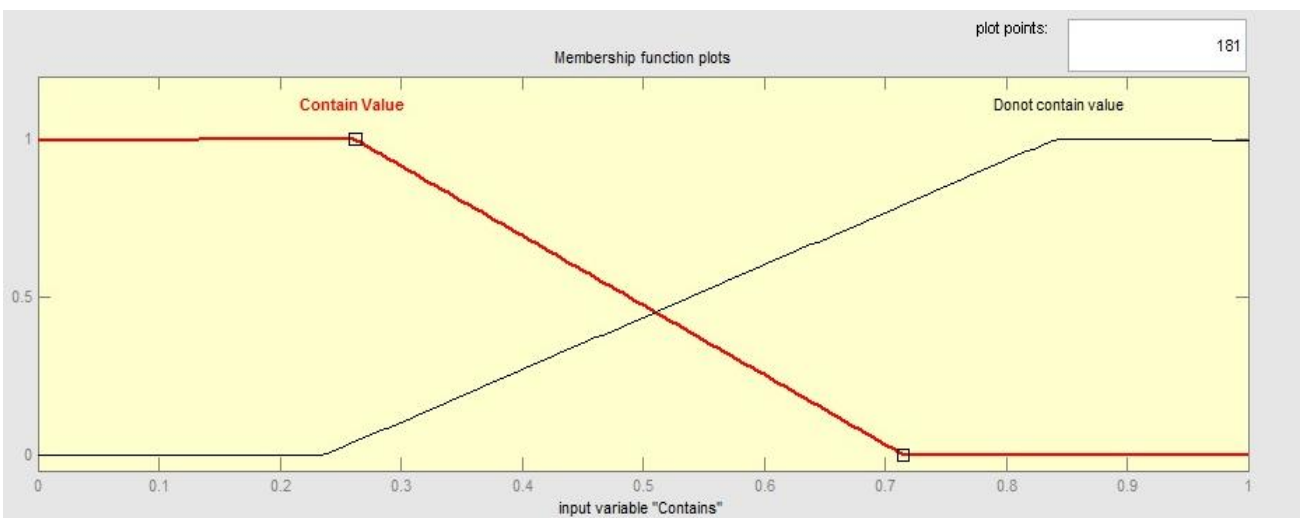
(a)



(b)



(c)



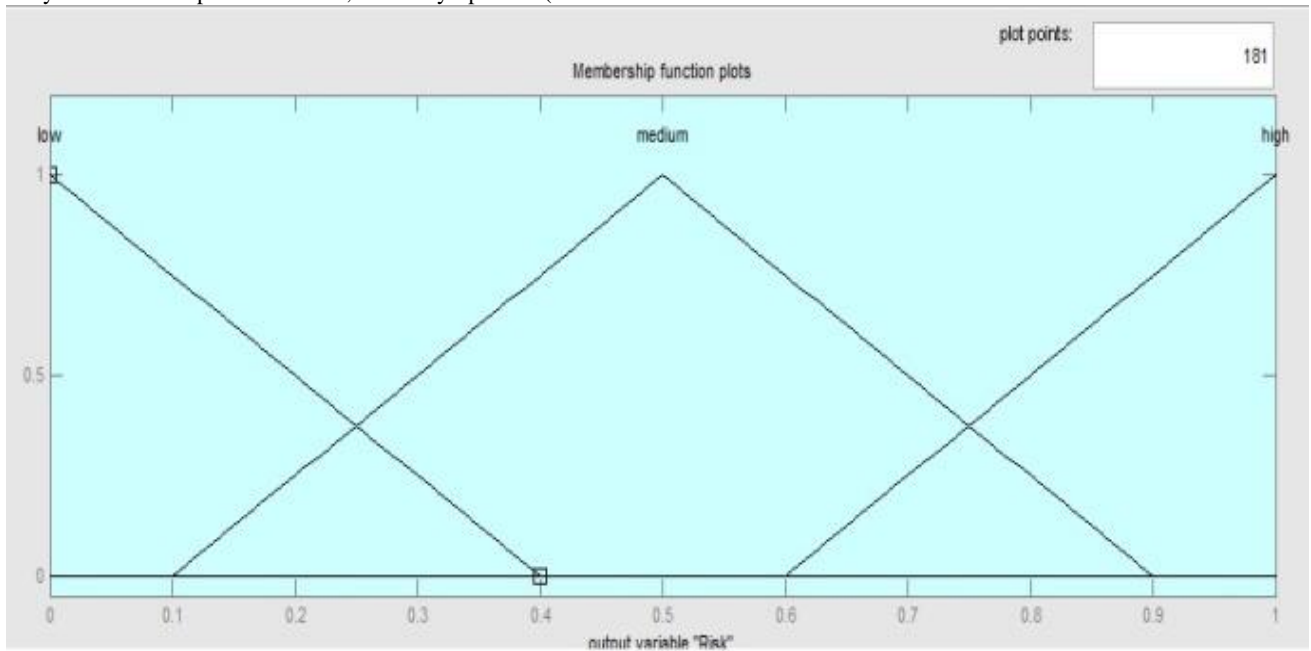
(d)

Fig.5: Input variables after fuzzification stage. (a) Usbstor Key. (b) DEVICE_CLASS Key. (c) STORAGE_REMOVABLE_MEDIA Key. (d) Contain.

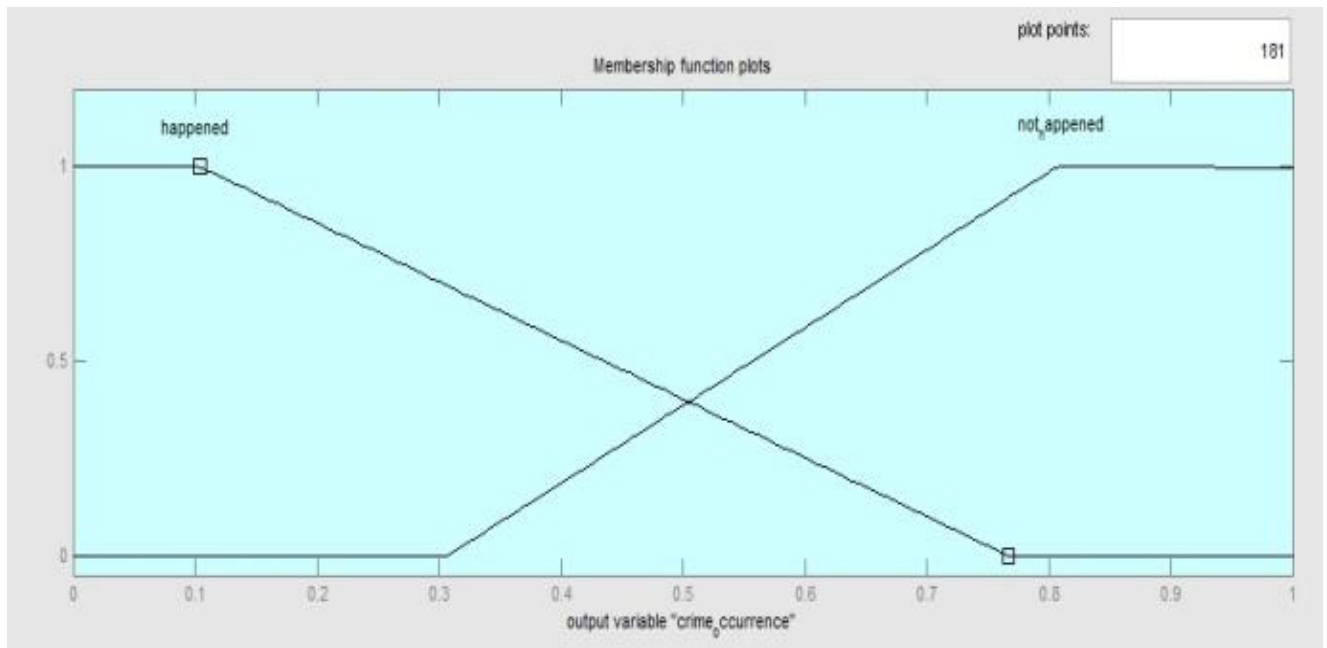
8.2 Rule evaluation

The second step is to take the fuzzified inputs USBSTOR =0.5, DEVICE_CLASSES = 0.5, STORAG_REMOVABLE MEDIA= 0.5, Contains= 0.5, and apply them to the antecedents of the fuzzy rules as shown in Figure 7. If a given fuzzy rule has multiple antecedents, the fuzzy operator (AND

or OR) is used to obtain a single number that represents the result of the antecedent evaluation. To evaluate the disjunction of the rule antecedents, we use the AND fuzzy operation as shown in Table 1.



(a)



(b)

Fig. 6: Output variables after fuzzification stage. (a) Risk output. (b) Crime occurrence output.

Table 1: Fuzzy logic rules

1. IF (USBSTORE IS low) AND (DEVICEClasses IS low) AND (STORAG_REMOVABLE MEDIA IS low) AND (Contains IS Donot_Contain value) THEN (Risk IS low) AND (crime_occurance IS not happened) (0)
2. IF (USBSTORE IS medium) AND (DEVICEClasses IS medium) AND (STORAG_REMOVABLE MEDIA IS medium) AND (Contains IS Contain_value) THEN (Risk IS medium) AND (crime_occurance IS happened) (1)
3. IF (USBSTORE IS High) AND (DEVICEClasses IS high) AND (STORAG_REMOVABLE MEDIA IS high) AND (Contains IS Contain_value) THEN (Risk IS high) AND (crime_occurance IS happened) (1)
4. IF (USBSTORE IS high) AND(Contains IS

- Contain_value) THEN (Risk IS high) AND (crime_occurance IS happened) (1)
5. IF (DEVICEClasses IS high) AND (Contains IS Contain_value) THEN (Risk IS high) AND (crime_occurance IS happened) (1)
6. IF (STORAG_REMOVABLE MEDIA IS high) AND(Contains IS Contain_value) THEN (Risk IS high) AND (crime_occurance IS happened) (1)

8.3 Aggregation of the rule outputs

Aggregation is the process of unification of the outputs of all rules. Figure 7 shows how the output of each rule is aggregated into a single fuzzy set for the overall fuzzy output. The output test has produced that Risk = 0.5, and crime occurrence = 0.257.

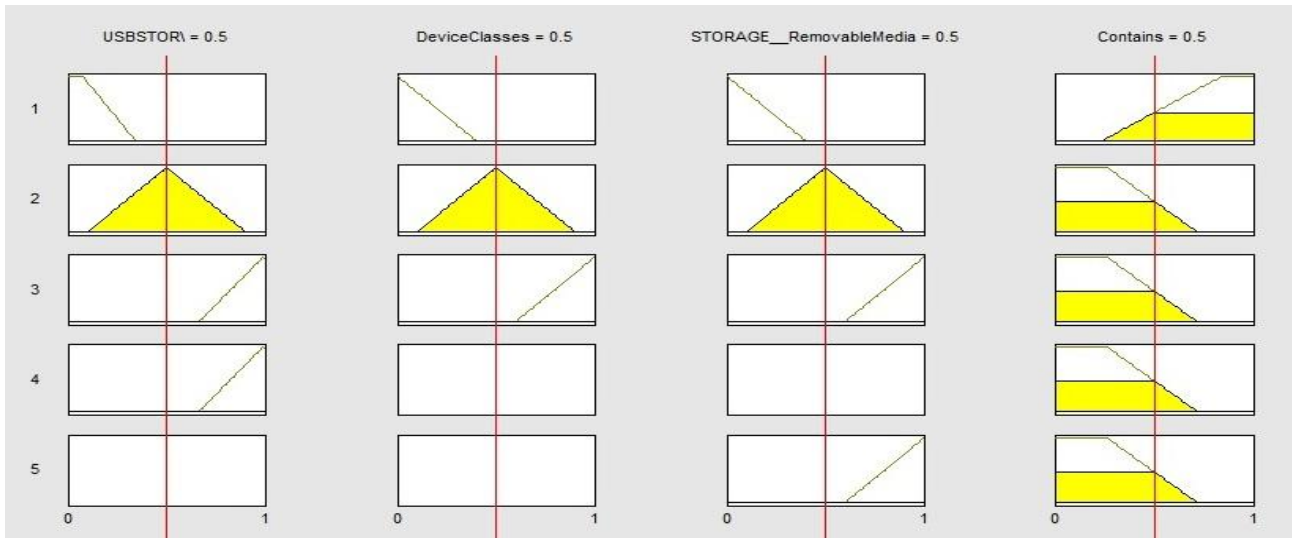


Fig.7 Rules representation for each input variable

8.4 Defuzzification

The last step in the fuzzy inference process is defuzzification. Fuzziness helps us to evaluate the rules, but the final output of a fuzzy system has to be a crisp number. The input for the defuzzification process is the aggregate output fuzzy set and the output is a single number. There are several defuzzification methods, but probably the most popular one is the centroid technique. We used centroid technique to find the point where a vertical line would slice the aggregate set into two equal masses. Figure 8 shows the output from defuzzification stage and after calculating the centroid.

9. CONCLUSIONS

The advancement in digital devices and the growth usage number of these devices in many daily bases fields. This led to computer crimes also to increase. And allowed cyber criminals to maliciously attack vital computational infrastructure to obtain or misuse the information illegally. After a crime occurred in a computer device, an investigation

process should take place to reveal what happened based on some evidence. Since any computer device operates through and operating system. Besides, one of the most well known and used operating systems is the Microsoft Windows. In Windows platform, there are many areas that help the investigator participating in the case analysis. Windows Registry is one of the excellent sources for potential evidential data in Windows OS. Since the plethora information, such as user accounts, typed URLs, shared network, and run command history, stored in it.

In this paper; we introduced a proposed expert system framework based on fuzzy logic IF-THEN rules. In which, we applied it on a case scenario to extract some potential evidence from a USB device. The expert system can evaluate the risk of any crime and determine whether it took place or not based on the founded evidence and the fuzzy rules. The purpose of this approach is to assure the accuracy of examining evidence. It helps the investigator in analysis process as shown the block diagram.

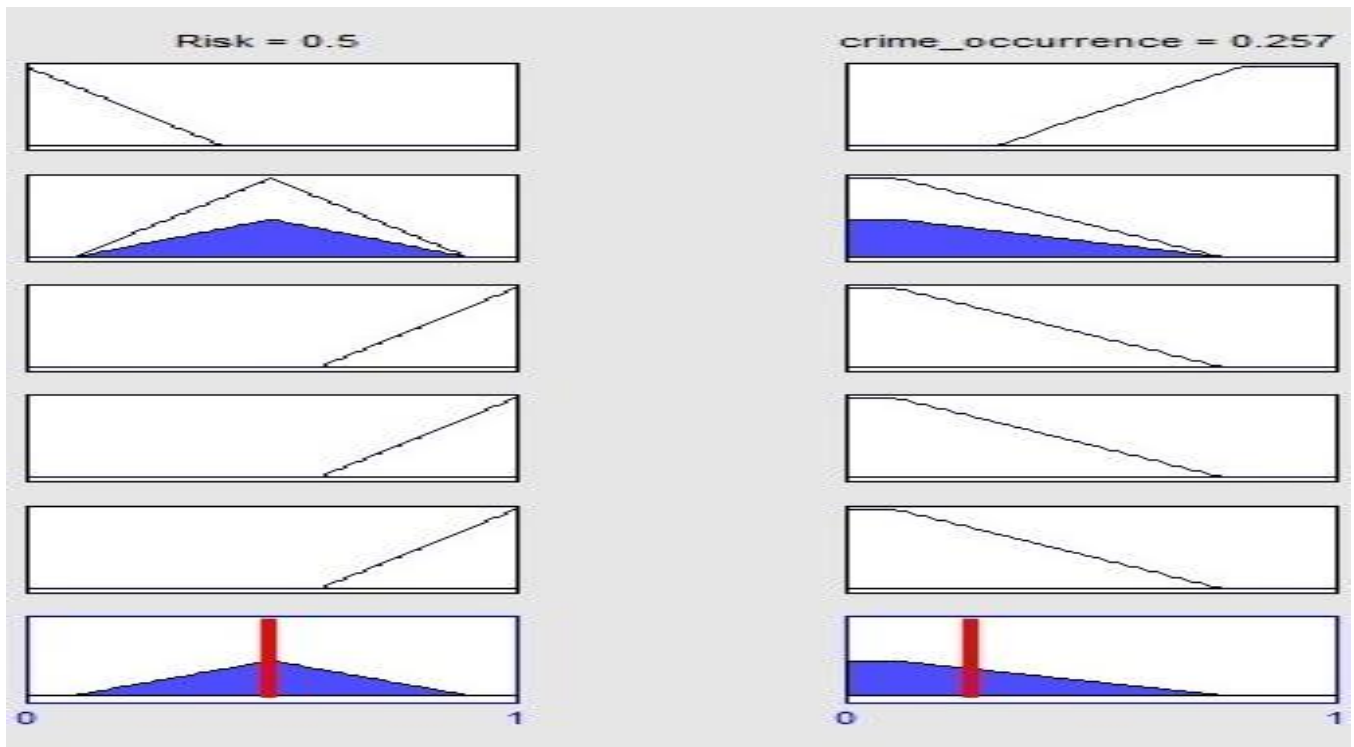


Fig.8 output rules representation for risk and Crime occurrence evaluation

10. REFERENCES

- [1] Victor Chileshe Luo (2007), 'Tracing USB Device artifacts on Windows XP operating system for forensic purpose', in Australian Digital Forensics Conference, ed., Edith Cowan University, 1-10.
- [2] Mathieu Gorge: USB and other portable storage device usage: Be aware of the risks to your corporate data in order to take pre-emptive and/or corrective action, Elsevier Vol 2005(Issue 8), Pages 15-17, (2005)
- [3] Krishnun Sansurooah : A forensics overview and analysis of USB flash memory devices, in Australian Digital Forensics, Edith Cowan University, Perth Western Australia, 99-108, (2009.)
- [4] Marcel Breeuwsma, Martien de Jongh, Coert Klaver, Ronald van der Knijff and Mark Roeloffs 'Forensic Data Recovery from Flash Memory', SMALL SCALE DIGITAL DEVICE FORENSICS JOURNAL , Vol. 1, No. (1), pp: 1-17, (2007)
- [5] Yang Yu, and Tzi-cker Chiueh: Display-Only File Server: A Solution against Information Theft Due to Insider Attack.', 1-9, (2004)
- [6] Kevin Hillstrom, Hillstrom Laurie Collier: Encyclopedia of Small Business. Magee D. Monique ed.. Farmington Hills, MI, USA, Gale Group/Thomson Learning, THIRD EDITION, Vol. 1 & 2, pp:1-1230, 2007
- [7] Grier Jonathan: Detecting data theft using stochastic forensics, Digital Investigation, Vol. 8. – pp: 71-77, 2011
- [8] Yuandong Zhu, Pavel Gladyshev, and Joshua James 'Using shellbag information to reconstruct user activities', Digital Investigation Vol. 6, pp: 69-77, (2009)
- [9] Zhenhua Tang Hong Ding, Ming Xu, Jian Xu: Carving the Windows Registry Files Based on the Internal Structure', in The 1st international conferences on information science & engineering, ed., The 1st international conferences on information science and engineering, pp: 4788- 4791, (2009)
- [10] Salvatatro J. Stolfo, Eleazar Eskin, Katherine Heller, Shlomo Hershkop, Andrew Honig, and Krysta Svore :A Comparative Evaluation of Two Algorithms for Windows Registry Anomaly Detection, pp: 1-34. (2005)
- [11] Youngsoo Kim, Dowon Hong: Windows Registry and Hiding Suspects' Secret in Registry, Technical report, Electronics Telecommunications Research Institute (ETRI)/Convergence Security Group, pp: 393 - 398, (2008)
- [12] Ewa Huebner ,and Frans Henskens :The Role of Operating system in Computer Forensics', Acm Sigops Operating Systems Vol. 42, No.(3), pp:1-3. (2008)
- [13] Jolanta Thmassen: Forensics Analysis of Unallocated Spaces in windows Registry Hives', Master's thesis, University of Liver pool, pp:1-63 (2008)
- [14] George J. Silowash, and Christopher King 'Insider Threat Control: Understanding Data Loss Prevention (DLP) and Detection by Correlating Events from Multiple Sources', Technical report, Carnegie Mellon University, pp:1-17, (2013)
- [15] Chad Steel: Windows Forensics the Field Guide for Corporate Computer Investigation, John wiley & sons, (2006)
- [16] Peter Hipson. Mastering Windows XP Registry, SYBEX

- Inc. (2002)
- [17] Abhijeet R. Patil: Forensic Analysis Of Windows XP Registry', Club Hack Magazine, Vol. 1, pp: 1-32, (Feb 2010)
- [18] Jerry Honey Cutt : Microsoft Windows Registry Guide, Microsoft Press, (2005)
- [19] Lih Wern Wong: Forensic Analysis of the Windows Registry, Technical report, School of Computer and Information Science, Edith Cowan University, pp. : 1-13, [<http://www.forensicfocus.com/downloads/forensic-analysis-windows-registry.pdf>] last access in (26/6/2012)
- [20] Derrick J. Farmer: A Forensic Analysis of the Windows Registry, Technical report, Champlain College Burlington, Vermont, pp: (1-17) [http://eptuners.com/forensics/contents/A_Forensic_Examination_of_the_Windows_Registry_DETAILED.pdf] last accessed in (28/2/2012),
- [21] Haoyang Xie, Keyu Jiang, Xiaohong Yuan, and Hongbiao Zeng: Forensic Analysis of Windows Registry Against Intrusion', International Journal of Network Security & Its Applications (IJNSA) Vol.4, No.2, pp. 121-134, (2012.)
- [22] Carvey Harlan: The Windows Registry as a forensic resource', Digital. Investigation Vol : 2 No.3, pp: 201--205. (2005)
- [23] Ashley Brinson, Abigail Robinson, and Marcus Rogers: A cyber forensics ontology: Creating a new approach to studying cyber forensics, Digital Investigation, El Sevier, pp: 37-43, (2006)
- [24] Niandong Liao, Shengfeng Tian, and Tinghua Wang 'Network forensics based on fuzzy logic and expert system', Computer Communications, Vol., 32, No. (17), pp: 1881 – 1892, (2009)
- [25] Chung-Huang Yang, and Pei-Hua Yen: Fast Deployment of Computer Forensics with USBs, International Conference on Broadband, Wireless Computing, Communication and Applications, IEEE, 413-416. (2010)
- [26] Tanushree Roy, and Aruna Jain: Windows Registry Forensics: An Imperative Step in Tracking Data Theft via USB Devices, (IJCSIT) International Journal of Computer Science and Information Technologies Vol. 3, pp: 4427- 4433, (2012)
- [27] Michael Negnevitsky: Artificial Intelligence, A Guide to Intelligent Systems, Addison-Wesley, (2005).
- [28] Kilian Stoffel, Paul Cotofrei, and Dong Han 'Fuzzy Methods for Forensic Data Analysis', IEEE International Conference of Soft Computing and Pattern Recognition, pp: 23-28, (2010)
- [29] Franke, Katrin and Srihari, Sargur N. : Computational Forensics: An Overview, in 'Proceedings of the 2nd international workshop on Computational Forensics', Springer-Verlag, Berlin, Heidelberg, pp. 1—10, (2008)
- [30] Ahmed M. Neil, Mohammed Elmogy, and A. M. RIAD: A Proposed Framework for Crime Investigation Based On Windows Registry Analysis, Journal of Engineering and Applied Science, Faculty of Engineering, Cairo University, Vol. 60, No. 1, February 2013.