2019 IEEE 9th International Conference on System Engineering and Technology (ICSET), 7 October 2019, Shah Alam, Malaysia

Digital Forensics Investigation Reduction Model (DIFReM) Framework for Windows 10 OS

Yazid Haruna Shayau
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
Serdang, Malaysia
Gs48608@student.upm.edu.my

Aziah Asmawi
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
Serdang, Malaysia
a_aziah@upm.edu.my

Siti Nurulain Mohd Rum
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
Serdang, Malaysia
snurulain@upm.edu.my

Noor Afiza Mohd Ariffin
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
Serdang, Malaysia
noorafiza@upm.edu.m

Abstract— **The advent of the digital age, globalization and automation has made life easier for people and businesses. However, the ubiquitous use of digital devices and the Internet also heightens the risk and incidents of cybercrimes. Under these circumstances, Digital Forensics has become a critical countermeasure. The ISO/IEC 27001 (Information security standards published jointly by the International Organization for Standardization – ISO and the International Electrotechnical Commission-IEC) provides guidance on identifying, gathering/collecting/acquiring, handling and protecting/preserving Digital Forensic evidence for use in court. The most challenging and important part of Digital Forensic Investigation (DFI) is data examination. Knowing the data created by the Operating System (OS) or user beforehand would ease the process. Unfortunately, most of the time, such details are not available to facilitate investigation. The examination phase is the most challenging for an investigator; in Microsoft Windows OS (Operating System). Investigators have to go through terabytes of system data, most of which are OS and application files irrelevant to the investigation from a suspect's computer. To address the problem highlighted above, this research proposes a data reduction model (DIFReM) and a tool which will not only help the investigator in identifying modified system files but also has the ability to detect files inserted into system directories and also be able to verify integrity using hashing. In the end, this research will provide the investigator with a more effective and efficient digital forensics tools.** (*Abstract*)

Keywords— Digital Forensics, Digital Forensics Investigation, Digital Forensics Models, Reduction, Windows 10 (*keywords****)***

I. Introduction

Nowadays, the adoption and use of all sorts of digital devices (Internet of Things – IoT) has caused people to share their sensitive personal and financial information on interconnected devices. The adverse effect of this phenomenon is that it gives cybercrime perpetrators easier access to endless stream of devices making it possible to commit crimes from across the globe thereby making it difficult to detect or charge the perpetrator in a court of law.

Preventive measures are only effective to a limited extent. That is why digital forensics is important to aid in unraveling digital crimes. The process usually begins with a warrant that leads to search and confiscation of digital evidence which will be investigated for trails left behind by the perpetrators using accepted methods and tools. Trails like the migration of data which may be an image, a file, an email, a transaction history or even activity log. Digital forensic investigations strictly demand use of not only professional tools, techniques and know-how but also industry-accepted standards and professionalism during the entire process beginning from request of a warrant to presentation of findings in court.

This paper describes the framework for the DIFReM as a tool for that will not only help the investigator in identifying modified system files but also the ability to detect files inserted into system directories and also be able to verify integrity using hashing. The remainder of the paper is organized as follows: Section II provides a background on digital forensics, while Section III elaborates on related works. Section IV contains our proposed model (DIFReM) which will hold a theoretical framework that addresses enhancing digital evidence collection and acquisition for Windows 10 OS. Drawn conclusions and suggestions for future work are presented in Section V.

II. Background

Cybercrimes continue to be a major challenge to the Internet community. As such, there will be high demand for digital forensic investigations to aid in tackling the menace, bringing culprits to book and compensating victims by serving justice.

A. Digital Forensics

The primary purpose of digital forensics may be explained as the processes of discovery, protection, collection, analysis and presenting legal electronic evidences which are thought of as potential evidences. Digital forensics aims to discover digital evidence for different types of cases ranging from identification of a digital intruder to resolution of a murder case. In digital forensics, the main objective is not to directly expose a person as guilty or innocent. It aims to provide numerical evidences to forensics unit in a different way as complete and impartial interpretation of the evidence. Conclusion on the culpability of the suspect lies on the

978-1-7281-0758-5/19/$31.00 ©2019 IEEE

judicial authorities by using the evidence presented by the forensics investigator as a result of his/her investigation of the evidence obtained through judicial processes. There are other areas of digital forensics which include data recovery, data annihilation, data conversion, encryption and decryption; finding undercover files and identifying criminals with the help of IP numbers.

B. Digital Forensics Processes

Processes which culminate to arriving at legal electronic evidence from investigated electronic evidence is called Digital Forensics Phases. Digital Forensics Phases are shown in Fig. 1. For digital forensics, the starting point is a realization of a crime or incidence due to a report, suspicious records, and sign of intrusion, alteration denunciation of an individual or crime case. Those who respond first to a crime scene are responsible for its security and that of the evidences. Therefore, first responders and digital forensics investigators should be properly trained on protocols of identifying a crime scene (taking pictures and videos) beforehand. They should be well trained in securing and protecting the crime scene. This shall be followed by collection of evidence, protection of evidence, analysing evidence, reporting and presenting the evidence. The Fig. 1 below shows the Digital forensics phases.

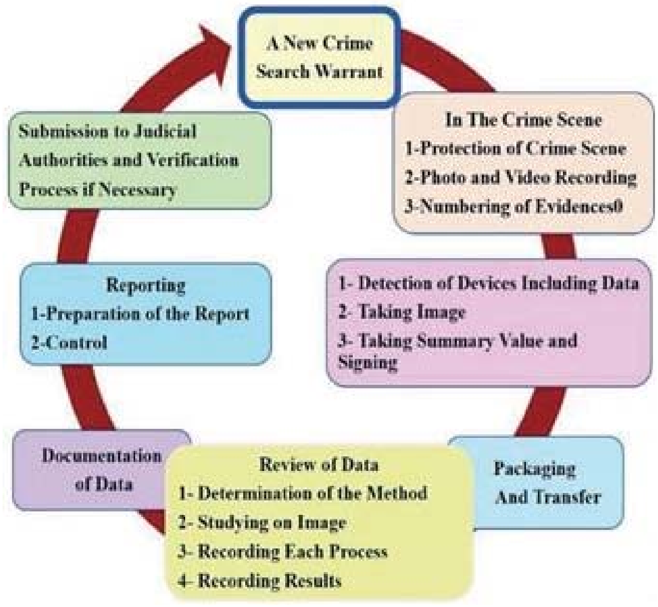
A New Crime
Search Warrant
In The Crime Scene
1-Protection of Crime Scene
2-Photo and Video Recording
3-Numbering of Evidences0
1- Detection of Devices Including Data
2- Taking Image
3- Taking Summary Value and Signing
Packaging And Transfer
Review of Data
1- Determination of the Method
2- Studying on Image
3- Recording Each Process
4- Recording Results
Documentation of Data
Reporting
1-Preparation of the Report
2-Control
Submission to Judicial Authorities and Verification Process if Necessary

Fig. 1. Digital Forensics Phases

C. Digital Forensic Investigation

As the adoption of computer technology keeps growing over the years, computer forensics, similarly coined as digital forensics, became a professional field. The surge in computer use and cyber crimes leads to increasing demand for this skillset and as such it became not only a skill but specialised area in academics and law enforcement. Crimes committed using digital means and requiring digital forensics investigation ranging from misdemeanours like cyberbullying, email harassment, illegal access of personal files, corporate espionage to serious crimes like spying, murder and also terrorism. The judiciary, corporate bodies, governments, private investigators, network managers and entities now seek the services of computer forensics experts to aid in investigating and solving civil and criminal cases.

D. Digital Forensics Investigation Models

There have been several attempts at creating a digital forensics investigations models that can be accepted universally but it has not been fruitful owing to the diversity of the entire computer forensics field. Alternatively, unique models are designed for specific areas of computer forensics, like Cloud Forensics, Android Forensics and IoT Forensics among others.

E. Reduction Approach

For the challenges faced by investigators in going through hundreds of thousands of files, the most frustrating part is not knowing where specifically the suspect has hidden, modified or removed a crucial pertinent artefact from. This among others is due to the fact that the investigator cannot have complete knowledge on the total volume or nature of files that are involved in the target computer. This can be solved by embracing a unique feature of the Microsoft Windows OS which is that most of the files are;

1. Same for all similar operating system both in size and hashes
2. Only very few system files are modified (during use). The rest are needed to function without the system or user modifying them throughout the course of use.

This uniqueness is what we want to harness by creating a reduction model which will use a database of untouched clean system files to use as a benchmark for comparison with the suspected OS. A few reduction models have been proposed but to the best of our knowledge, a similar approach has not been done for Windows OS. Most indexed approaches were geared towards document search or in-text investigations aimed at deducing similarities between word files or the contents in them. Our choice for windows 10 Professional 64bits is because it is the basic level distributed on new laptops or PCs as such, is among the most widely used consumer OS at present.

III. Related Works

A. Digital Forensic Data Reduction and Data Mining Framework

[1] reviewed the challenges in digital forensics with regards to dealing with voluminous size of data which may affect timeframe to conclude an investigation. In their study, they highlighted the consequence of growing storage size as well as “big data” and proposed a Digital Forensic Data Reduction and Data Mining Framework which is employed in different phases of a digital forensic examination. The proposed framework does not replace the standard phases and procedures that define a full-blown forensics analysis. They analyzed their design alongside a normal forensics process so as to identify the unique phases where the reduction procedure is employed and also a review phase which distinguishes between the reduced and full analysis. This helped establish a standard digital forensic framework

which can be compared to the reduced framework so as to evaluate the cost-effectiveness among other needs by an investigator during the investigation by reducing volumes of data to be perused at each stage of the forensics investigation process. The framework is applied in relation to the seven needs identified by the researchers; faster collection, reduced storage, timely review, intelligence, research, knowledge management, archive and retrieval. They further proposed that the type of data to be collected and investigated should be clearly thought with focus on data or area where pertinent information is expected to reside.

[2] specified a procedure for selective imaging address the challenges related to collection of full forensic images especially from large storage by culling the data to image at the collection phase. Legally approved procedure prefers a full forensic imaging so as to guarantee integrity and also minimize adulteration. Nonetheless, due to time constraints the investigator may in the first instance decide to create a bitstream image (Full disk image) after which proceeding to select the isolated part for investigation. This can guarantee availability of other parts to be perused in the future if needed. With this, their proposed framework will still maintain its full imaging and analysis phases, but adopt a reduced collection and review phases which will aid in reducing the full analysis rather than replacing it.

[3] proposed that a solution to the volume of data challenge is to strategically choose a subset of data instead of an entire bitstream copy and the subset could include portions of unallocated space. Nevertheless, discussions were made insisting on the need to design a framework to be adopted. An example can be taken from the fact that a subset data like files from Microsoft Windows Internet Explorer Internet history 'index.dat' or history files and folders from other browsers may contain pertinent information required for solving a case even though having smaller size than other files like unallocated clusters, or 'Pagefile.sys' memory paging files. From this, they deduced that collecting and preserving Internet history files while ignoring unallocated clusters, will help reduce the volume and guarantee that pertinent information which may be crucial to an investigation will not be lost. Windows OS has lots of files of different types which may be important during investigation such as Log Files, Windows Registry Files, Windows Desktop Search database files, Prefetch files, email archival files and word documents. The reduction process is done by postulating that by not capturing and preserving the entire data, there is a danger of losing crucial information which may hinder the attempt of conducting full analysis of the collected data.

[4] proposed the concept of Digital Evidence Bags as an approach of collecting a variety of digital evidence while retaining information relating to the source and location of the data subset. [5] worked on a notion of a Sealed Digital Evidence Bag, which was coined from traditional evidence bags. Contemporary commercial forensic tools provide the investigator the capability to define the imaging files that should be adopted as the data subset into logical evidence files. [6] proposed Forensic Feature Extraction (FFE) and Cross Drive Analysis methods. FFE is defined as a scan of a disk image for email addresses, message information, date and time information, cookies, social security and credit card numbers [6]. Information obtained from scanned device is subsequently saved in XML format for analysis and comparison. However, new techniques may prove difficult to be applied to original or historical data for interpretation. Although ubiquity and advancement in both knowledge and technology makes it possible to obtain extra data from historical dataset that were previously unknown or not accessible due to limitation. For example, Windows Registry analysis methodologies include newly discovered areas for locating information [7].

B. Digital Forensic Data Reduction by Selective Imaging

[1] proposed an Operating System-geared process of decreasing both volume of data and time by only obtaining information from identified area where pertinent information is expected to reside thereby not only speeding up the process but also simplifying it for the investigator. Expectedly, Operating Systems have a wide range of data types and system files which will also be part of data at the collection phase. Some of those files may include: word documents, registry files, log files, web browser history files, email containers, Spreadsheets and different Windows framework files among others.

The proposed reduction process can be executed concurrently with a complete forensic examination and subsequently it may be required to retain the evidence device or storage media for need of new information after the analysis phase. They advised that although the main device may still be available for preservation or collection of data, a full imaging should be done in case further investigation is needed in the future. Assumingly, the authors deduced that reduction technique will breed the risk of the evidence not being in the reserved data subset. This makes it impossible for reduction to be risk-free when compared to standard methods which embrace the entire data regardless of volume or storage size. A digital forensics triage methodology scans files and interprets the information, which reduces the capacity to execute more examinations with other tools. This technique will preserve the initial files deducted and also provide a platform to access and investigate subsequently without constraint or restrictions to a specific method or tool and can also be investigated with triage or other softwares.

IV. DIFREM FRAMEWORK

Unlike standard Digital Forensics investigation stages, for DIFReM model as depicted in Fig. 2, the most important stage takes place before the investigation begins. This is the point where a clean unadulterated installation is indexed and a library is created for all files and their locations together with hash and header values of these files. This database is kept as a benchmark for all analysis to be carried out during the reduction and investigation of the suspect's system.

The first step of this important stage is to install a clean copy of Windows 10 Professional 64-bit edition. No file or application will be added and action will be taken on it that can make any changes to default post installation condition including registering or activating the OS. It will also not be connected to the internet or an external device.

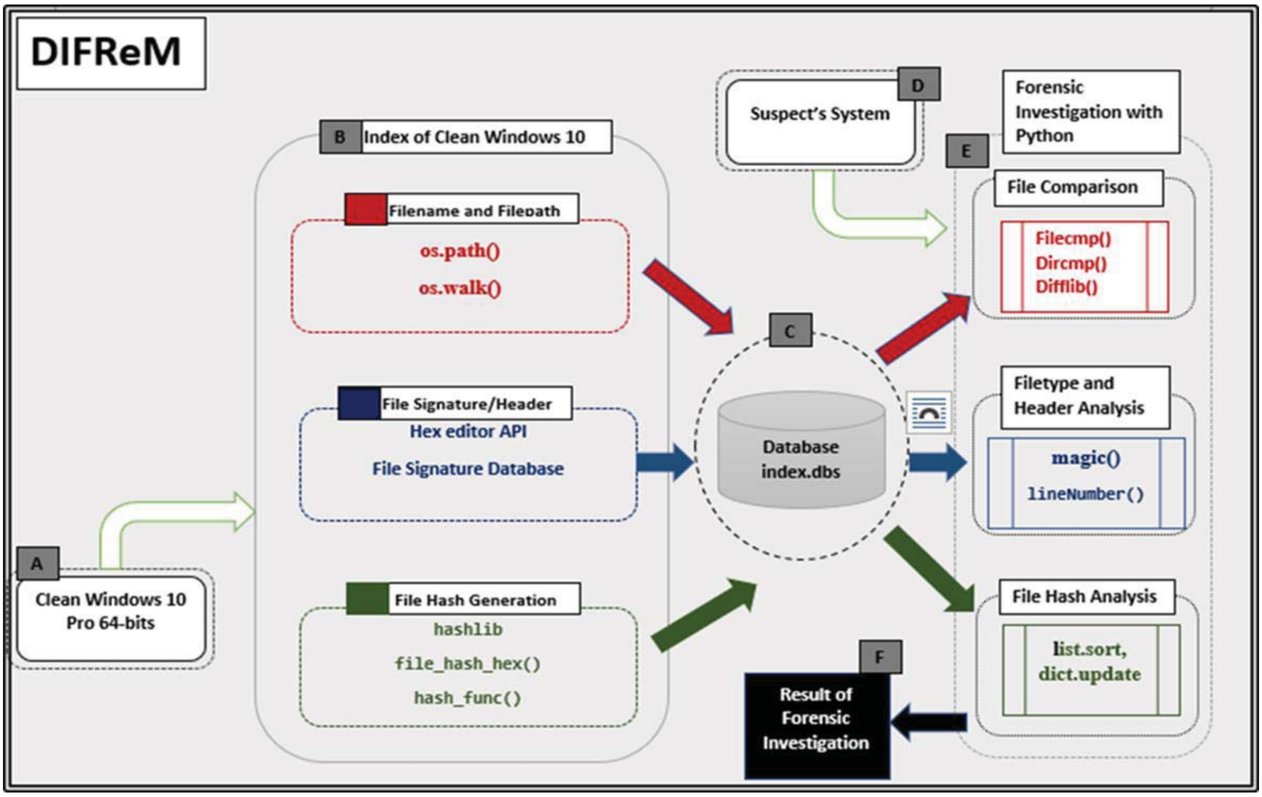

Fig. 2. DIFReM Framework

Afterwards an index will be made of all the files in the Operating System's C:/ drive including default User files. This index will later be used by an algorithm to crosscheck and isolate OS files as described in Fig. 3 for ease of investigation by a Forensics Investigator.

When the investigator runs the application, she/he is prompted on desired investigation to be carried out i.e., File signature check, Separation of OS and Non-OS file and OS files integrity check. The investigator will define the algorithm that the application will employ for that investigative process. Alternatively, the investigator can choose to run all options for a comprehensive analysis as described in Fig. 4 and Fig. 5.

An algorithm will be used to compare the suspect's Windows installation against an index of a clean unadulterated installation. Thus, an investigator will be alerted if unknown files have been added, removed or transposed in the suspect's Windows installation directory. Also, personal user files will be separated, giving the investigator a subset to work with while the tool executes the comparison and checks.

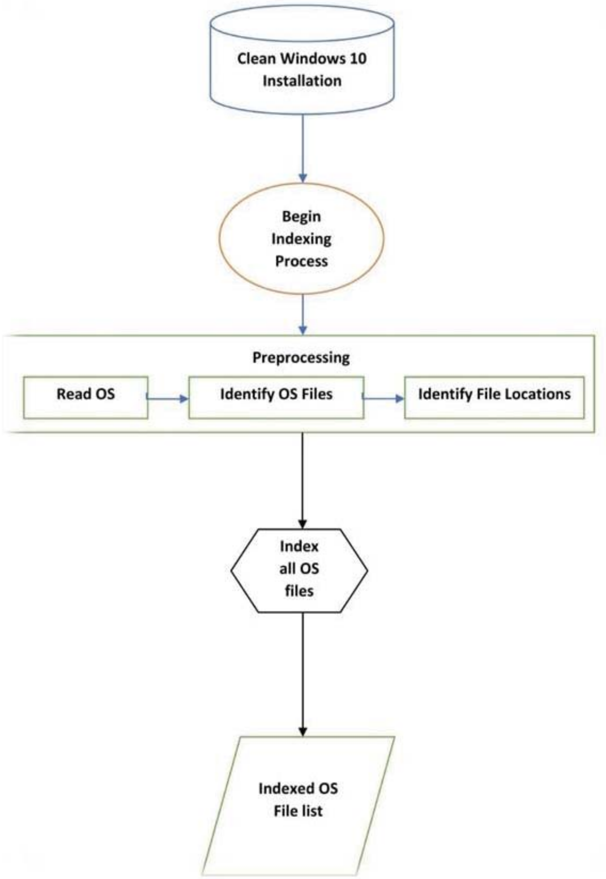

Fig. 3. Pre-processing Windows 10 64-bit OS files index

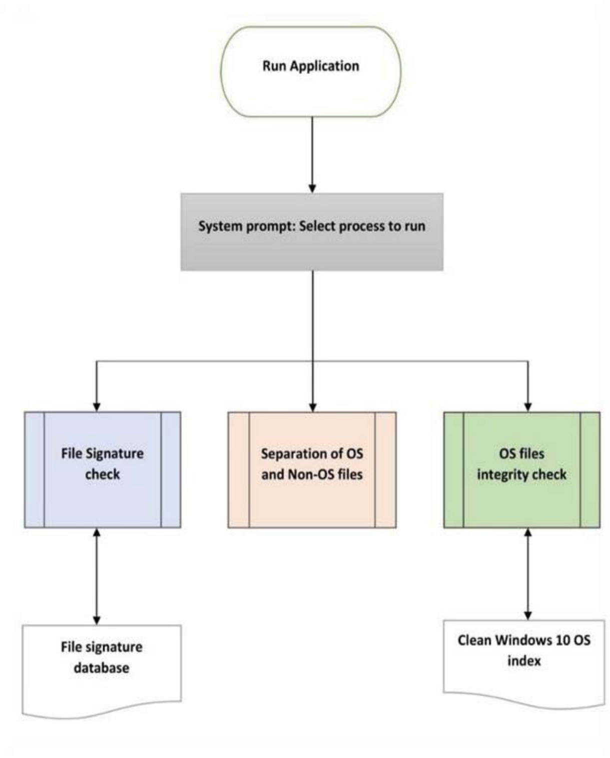
Run Application
System prompt: Select process to run
File Signature check
Separation of OS and Non-OS files
OS files integrity check
File signature database
Clean Windows 10 OS index

Fig. 4. DIFReM components flowchart

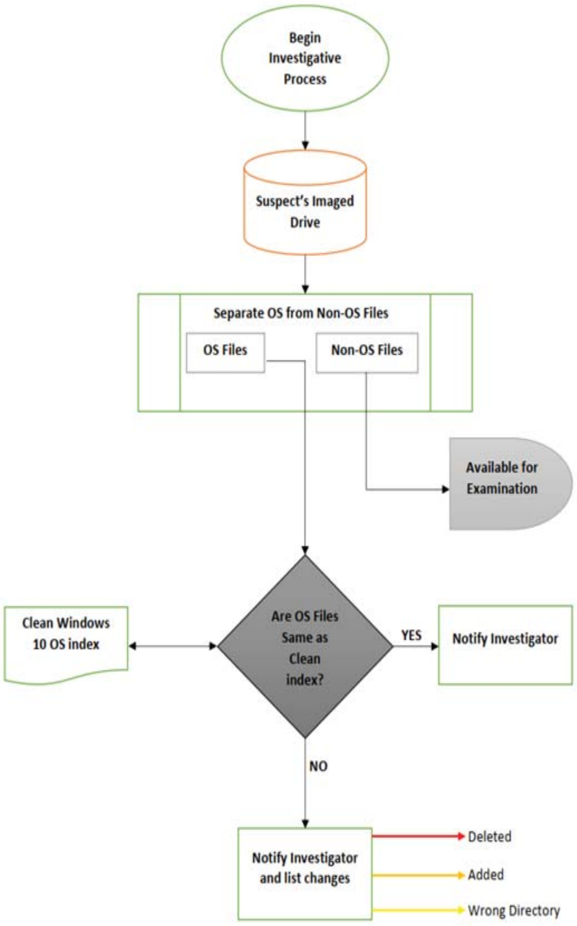
Begin Investigative Process
Suspect's Imaged Drive
Separate OS from Non-OS Files
OS Files
Non-OS Files
Available for Examination
Clean Windows 10 OS index
Are OS Files Same as Clean index?
YES
Notify Investigator
NO
Notify Investigator and list changes
Deleted
Added
Wrong Directory

Fig. 5. Comparing suspect's OS files to index for investigation

An algorithm will use a File Signature library to compare files with their extensions to detect if the extension has been changed so as to mask the file. This will be achieved by first reading a file's header, then checking the bits and comparing to our database of bits vs extension type. This will tell the investigator if the extension is the same as indicated in the header or if there is a discrepancy. Also, the entire file header of the two files will be compared to see if any other modification has been done to the file header as illustrated in Fig. 6.

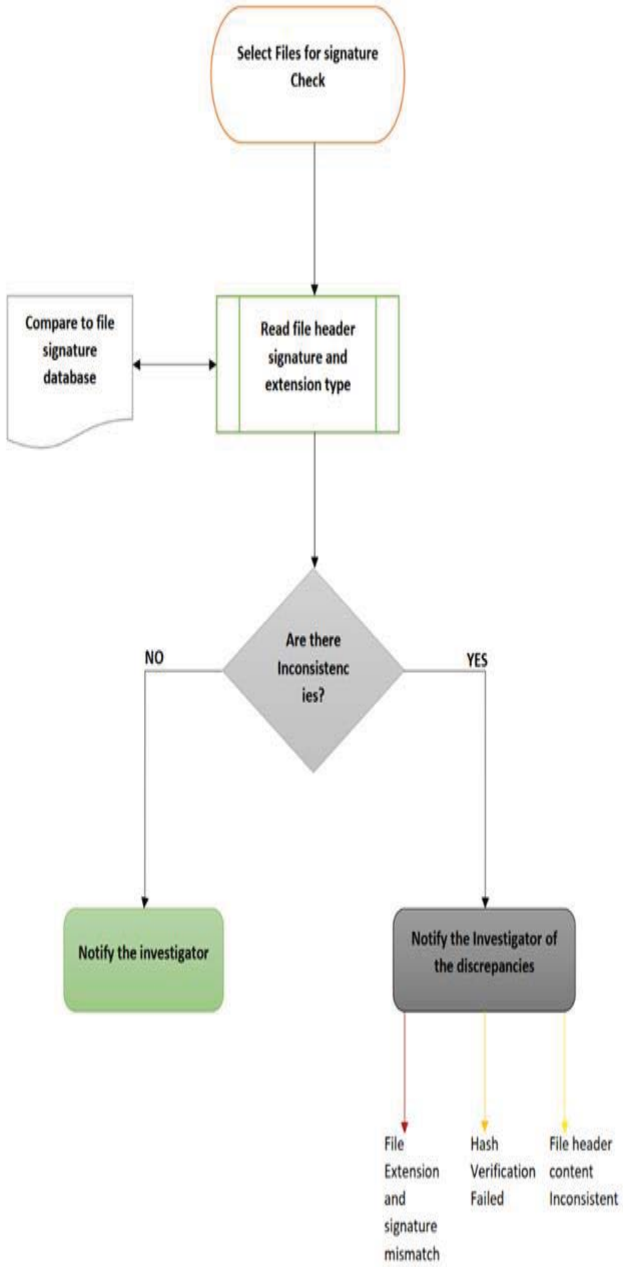
Select Files for signature Check
Compare to file signature database
Read file header signature and extension type
Are there Inconsistencies?
NO
YES
Notify the investigator
Notify the Investigator of the discrepancies
File Extension and signature mismatch
Hash Verification Failed
File header content Inconsistent

Fig. 6. Filetype verification using file signature in file header

With these two libraries; file index and hash algorithm database of the Microsoft Windows system files, we can now comparatively automate comparison between our libraries and files on suspect's system thereby greatly reducing the workload of the investigator while she/he concentrates on user files and post-installed applications.

The implementation of the DIFReM tool will not only help the investigator to identifying modified system files but also the ability to detect files inserted into system directories by automation with the ability for batch operations to ensure the investigator can concentrate on personal files of the suspect's system while the tool does the selected forensic operations.

V. Conclusion

Cybercrimes continue to be a challenge to the internet community. As such, there will be high demand for digital forensic investigations to aid in tackling the menace, bringing culprits to book and compensating victims by serving justice.

Yet, challenges still remain and will continue evolving ubiquitously with technology. Challenges like; high speed and volumes, explosion of complexity, privacy-preserving investigations, legitimacy, rise of antiforensics techniques and development of standards. With all these challenges, it is imperative to not only design generic forensic tools but all popular OS at least should have a standard tool designed solely for its investigation. This could greatly help speed up and simplify the task for the investigator especially at this age of ultra-fast Internet, high storage capacity and robust anti-forensic tools.

This reduction process will undoubtedly make the investigative examination phase of digital forensic process faster and easier thereby enabling the investigator to get to the bottom of the case within acceptable timeline by the court or security agency ensuring that not only does the culprit could be apprehended but also justice is not delayed.

In this vein, we hope that this tool geared towards Microsoft Windows 10 Professional operating system will open a path towards OS-defined forensic tools which will definitely be a delight to many investigators.

Acknowledgment

This work is supported by PUTRA Grant Scheme (IPM), Universiti Putra Malaysia under Project No. GP-IPM/2019 Vote: 9676100.

References

[1] D. Quick, & K. K. R. Choo, ”Big forensic data reduction: digital forensic images and electronic evidence.,” *Cluster Computing*, *19*(2), 2016, Pp. 723–740. https://doi.org/10.1007/s10586-016-0553-1

[2] E. Kenneally & C. Brown, “Risk sensitive digital evidence collection. Digital Investigation 2(2): 101–119, 2005.

[3] N. Beebe, ”Digital forensic research: The good, the bad and the unaddressed,” in Pollitt M & Shenoi S (eds), Advances in digital forensics: 17–36, 2009.

[4] P. Turner, ”Unification of digital evidence from disparate sources (digital evidence bags),” Digital Investigation 2(3): 223–228, 2005.

[5] B. L. Schatz & A. Clark, “An open architecture for digital evidence integration,” in AusCERT Asia Pacific Information Technology Security Conference. Refereed R&D Stream. 21–26 May 2006. Gold Coast, Queensland

[6] S. Garfnkel, “Forensic feature extraction and cross-drive analysis,” Digital Investigation 3: 71–81, 2006.

[7] H. Carvey, ”Windows registry forensics: Advanced digital forensic analysis of the Windows registry,” Burlington, MA: Elsevier, 2011.