Implementation of forensic tool I3A in Python programing language

Dragan Randjelović

Milan Čabarkapa

Jelena Mišić

Aleksandar Miljković

Aleksa Maksimović

Slobodan Nedeljković

Vojkan Nikolić

Abstract: Nowadays, in the computer era, the technology is a part of our daily lives. Therefore, the security of information stored on electronic media is one of the biggest challenges of the computer system security and integrity. Since the misuse of other people's data is a criminal offense, the security of data relates to both safety managers and law enforcement professionals and courts. The main problem of data security relates to the security of evidence. Namely, due to the specificity of computer crime detection, the classical methods of criminal forensics cannot be used for detection of this type of crime. Thus, it is necessary to develop a specific tool capable of responding to the new challenges and provide the evidence necessary for judicial proceedings. In this paper, we investigate the effect of various operating systems on the efficiency of forensic tools. The possibilities of both integrated and non-integrated tools of digital forensics are compared. The functions of the I3A and SIFT workstation forensic tools are considered, and short instructions for their use are presented. The effect of different operating systems on the I3A forensic tool is analyzed, and the obtained results are provided and discussed. The functionality of the I3A is tested on both Windows and Linux platforms (Ubuntu, Fedora, and Knoppix).

Keywords: computer networks, digital forensics, digital evidence, I3A, forensics tools efficiency.

Introduction

A large number of computer crimes is emerging almost every day worldwide. Following the technology development, this trendwill continue. Hence, people are more often the victims of a new type of crime, the modalities, and the "modus operandi" develops with a hitherto unseen dynamics. To successfully counter this type of crime, it is necessary to implement comprehensive prevention, and if that solutiondoes not produce the desired results, a key role in the discovery of the perpetrator and collection of the evidence of his guilt is now taken by a young discipline of forensics - digital forensics.

There are a lot of literaturerelatedto the digital forensicsfieldincluding the computer forensics, digital forensics, cyber forensics, etc. In the computer forensics, computer uses digital technology to develop and provide evidence forthe court and prove or disprove a claim. A

slightly different definitionof digital forensics is given by John Vacca, and according tohis opinion, a computer forensics involves the preservation, identification, extraction, and documentation of evidence stored on a digital computer. Moreover, in some cases, digital forensics is consideredseen as a science and as art that uses the IT knowledge and skills to assist in the resolution of any legal process. However, digital forensics is defined as a process of collecting, preserving, analyzing and presenting digital evidence. In most cases, the terms “computer forensics” and “digital forensics” are regarded as synonymous, but there is still some differencebetween them. Unlike computer forensics relatedto the collection of digital evidence stored on a computer (PC), digital forensics is a more general term and refers to all the devices that can carry digital data.

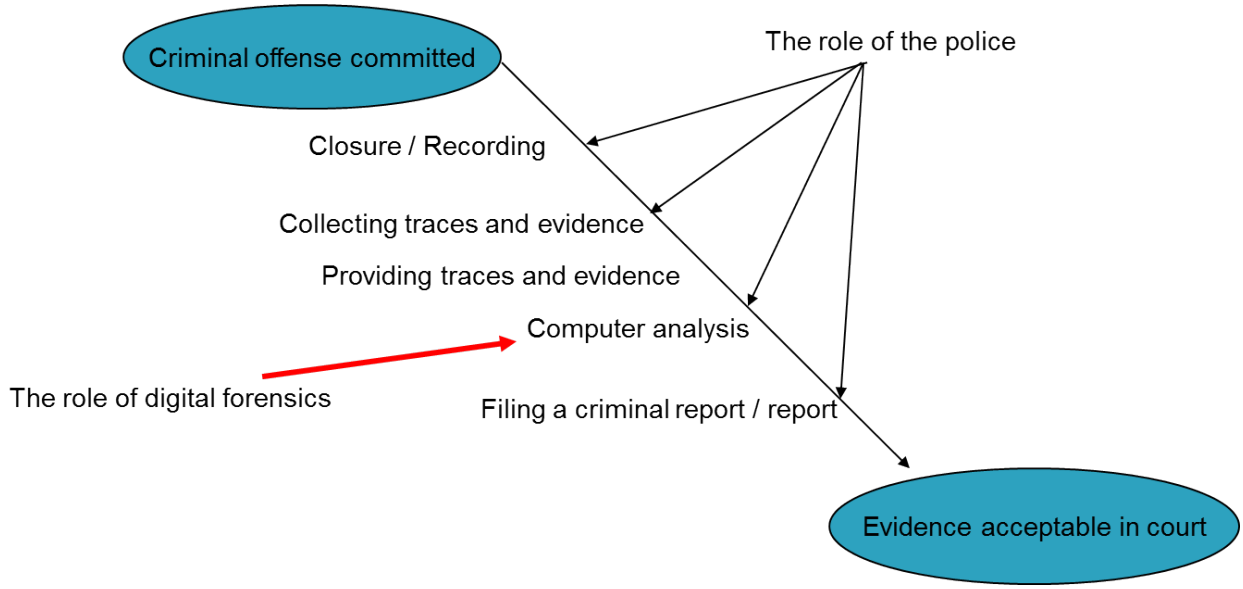

Image 1 – The role of digital forensics

When an incident occurs, the process of digital forensic investigations starts. Digital forensics is crucial for the successful detection and prosecution of criminals in the computer crimefield. When theprocedure of digital forensics analysisstarts, its duration must be conducted in accordance with the law because only in this way, evidence gathered in theprocess ofdigital forensics analysismay be valid in court. There is a general agreement in the literature on the sequence of procedures, but there are different opinions on the number of phases. In most cases, there are four stages, although there are cases where this number ofstages is three, five oreven seven.

The process of digital forensic investigation consists of four following stages:

- Acquisition, within the so-called bit-by-bit copy of data is made, andthis copy is called thedisk image.
- Searching, wherein the disk images are “start up” on a computer-elimination of files which does not represent the digital evidence.
- Analysis, wherein the interpretation of digital evidence is performed.
- Presentation of the results obtained from the previous stage.

According to one definition, digital evidence is defined as any information that is stored or transmitted using a computer and that supports or refutes the theory of how the offense is performed and who is theexecutor. Also, digital evidence can be defined as the data and information relevantto the investigation, which are stored or transmitted by electronic device in a digital form. In other words, digital evidence is any information in a digital format (consisting of 1 and 0), which is relevant to the legal proceedings such as various patterns of text, images, sound clip, video clip, or a combination of previous. The UNIX operating system has beenalreadyhighly-developedwhen the Windows operating system was introduced so that the majority of free tools and utilities are developed under the Linux.

Digital forensics tools

As widely known, digital evidence is stored within a computer system, so it is impossible to see the content without the help of appropriate forensic tools. Nowadays, there are many forensic tools including both the tools with only one purpose and those with a much greater range of options. The choice of a tool to use depends on the specifics requirements of the investigation. However, it is always desirable to choose atool that will contribute to the most to achieving the objective for which it is used. Forensic tools can be divided into several groups, but it should be noted that, according to the functions they perform, they donot strictly belong to one particular group. In the literature, in most cases, the tools are classified into commercial and non-commercial tools, i.e., those that are licensed and those that are open source.

The commercial tools are made mainly for the Windows platform. These tools have many modules integrated into a single program covering more areas of the process of digital forensic investigations. The main shortcoming of these tools is a high cost.

On the other hand, the non-commercial tools are free of charge. Moreover, they are running on Linux platform, and they usually incorporate all aspects of the process of digital forensic investigations. It should be highlighted thateven thoughthese tools are free, they can make a full investigation, i.e., provide all the features that the expensive commercial tools have. Besides, in the open source tools, thesource code is available for consideration and further customization, which makes them very functional.

The origins of computer forensic analysis are not related to the Windows operating system, which has achieved such popularity recently, but toUNIX, which represents the operating system developed in the early 1970s. The developers of UNIX preferred to create a relativelylarge number of small programs which canbe used together to perform more complex tasks rather than one program which cando everything, and it is from these small programs that the sophisticated commercial computer forensic packages available today have grown. The small programs are still found in modern versions of the UNIX operating system, and many are also available for Windows.

In addition to shortly described non-commercial software for digital forensics, it is necessary and obligatory to stress out so-called integrated forensics tools group, which integrate various (mentioned) non-commercial tools and their different combinations.

Importance of Help files

The help tools, although used by digital forensics experts, can be very complex to use because of a very specific nature of digital forensics science, which connects legal and security aspect with the aspects of computer, network and informational science; thus, working with these tools can be very challenging, even for digital forensics experts. Many digital forensic tools come with the integrated instructions, but these instructions either donot explain the procedure well enough or reference to the web pages and forums about that particular tool.

In many situations related to the digital forensics tools, it is necessary to analyze the given file or a computer data in a closed system, where the term „closed system“ denotes the system which because of security or safety reasons isnot connected to the Internet.

Besides, the digital forensics experts work is legally bound; namely, although some tool maybe offers a possibility to treat a file differently, the legal boundaries restrict the treatment demanding the procedure to be executed in a specified, clearly defined way. Therefore, it is necessary to integrate the instructions for the use andthe instructions related to the legal system in which they are used.

I3Atool

I3A is a bundle of forensic tools which can execute a series of processes of a digital forensic investigation.

Image 2 – Start screen of I3A software

I3A app was designed in the Visual Basic environment which is a composite part of the Visual Studio 2013 Ultimate package. Open source apps were used in their portable form – they are not installed withthe operating system or on the hard drive of a targeted computer, and in that way, they do not “pollute” it. By clicking on the button with the name of the app on its surface the desired program is run and simultaneously the help option is enabled, specifically Help files that have been localized in Serbian.

Digital forensic tool i3A can be run on a live system, by a USB drive, or by running a live version of Windows To Go system. None of the previously mentionedways to use theDigital forensic tool i3A does allow access or modification of media and data on it.

Every forensic softwaretool must be compatible with at least following generations of operating systems (OSs). I3A tool was tested working on computers with Microsoft Windows XP/7/8/8.1 and Windows 10 RC, in 32/64 bit versions.

Earlier it was impossible to install Microsoft operating systems on USB memory devices;thus forensic tools have been made for Linux platform. With the appearance of USB 3.0 SSD drive, it became possible to start a live version of operating system Windows 7/8 and 8.1 onit, along with the use of software packages integrated intoI3A tool. By using the Windows 8.1 operating system, we get a large base of drivers we can expand.

I3A is a group of forensic tools that can execute multiple steps of a forensic investigation. The dd acquiring tool secures a corrupt image creation [3]. Raw (dd) is a tool for image creation ofa physical or a logical drive, file or folder content which creates an uncompressed image and demands enough disc space. An autopsy is a forensic tool which performs the extraction and presentation of digital evidence. Extraction of digital evidence includes theanalysis and extraction of a relevant subset of the evidence data. Presentation denotes thedata management from an extraction tool to a more understandable and usable format.

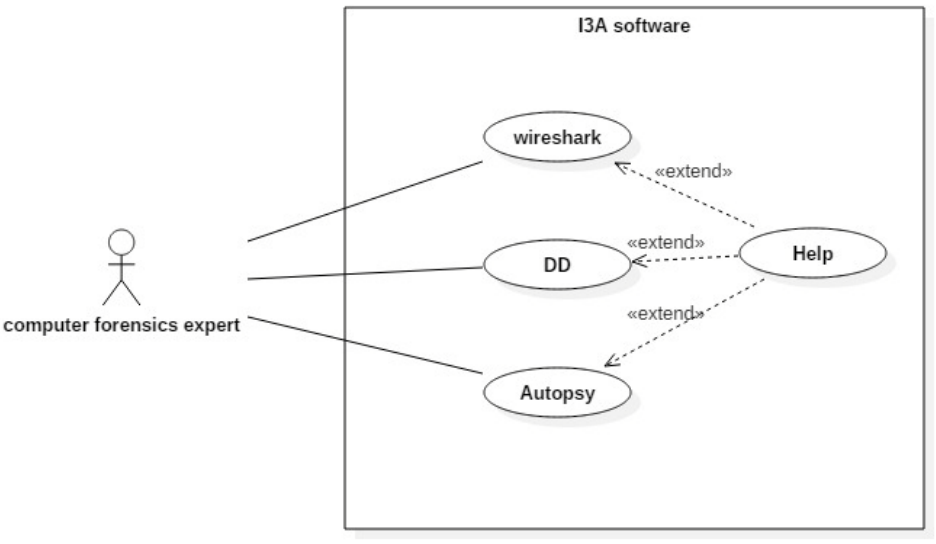

Image3 - Use Case diagram of I3A software

The diagram shows that when the i3A “computer forensics expert” is used, there is a possibility to chose one of the desired tools for forensic analysis. Besidesby running a portable version of the desired tool, there is a possibility toreview the manual for the exact tool. In that way, it is secured that auser, besides practical and fast forensic procedures, has an insight intothe proper ways of tool usage followed by the corresponding explanations and images.

Concrete forensic solutions for computer system analysis have been chosen because they cover a wide range of digital forensic subjects.

For the instruction manual, it is important that mentioned software hasa strong internet community and documentation that offer good explanations and directions for specific cases of application. Besides, to enable the customization of a certain part of the software for the necessity of automatizationand the specialty caused by the diversity of business process and investigation procedures, it is very important that eachof the integrated tools is within the frame of open source license.

An open-source software implies that the software source code is available with the open source license to all the users that can change, alter and improve its content. Namely, the open source programs come with the entire source code in theprogramming language they are developed in, so the program itself can be altered, which is not the case with paid software.

The main purpose of the open software is to make programs more understandable and available.

Inthe presented I3A tool, an integration of three open source tools has been executed, along with explanations for their usage in the app developed in Microsoft Visual Studio; thus, the I3A tool is consisted of:

- Wireshark,
- dd,
- Autopsy,
- Instruction manual.

In the forensics field, the ddis used to make an exact copy of data on the medium, which allows safe analysis of collected evidence. This program can copy and convert files, hard drives, compact disks, flash memories, disk fragments, etc.;itis used for missing data recovery and data backup copies. The dd copies bites of information from one place to another without knowing the structure of data. Therefore, in contrast to other programs for hard disc image creation, the dd copies everything from the hard drive, bit by bit, including the “slack” space and deleted files. The dd program providesa high level of simplicity and functionality with the minimal requirements for the performance of a device. The ddis an open source program, so it is completely free to download and use.

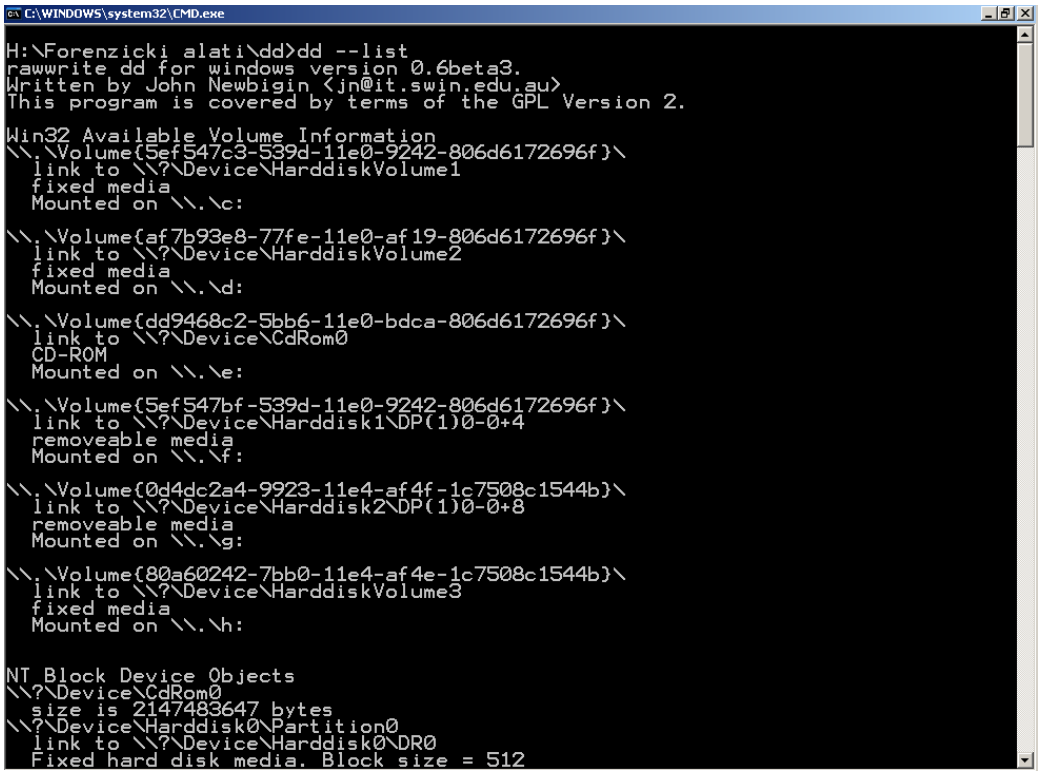

Image 4 – Example of command "-- List"

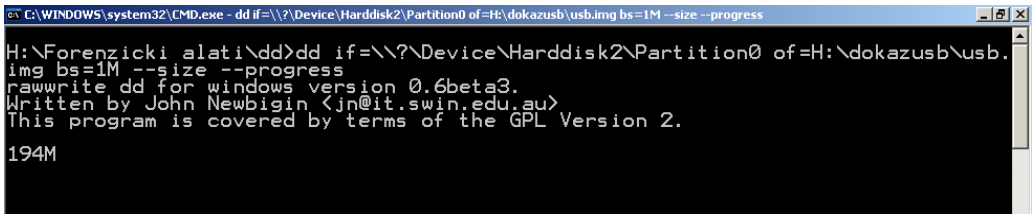

Image 5 – Procedure of creating an uncompressed RAW image

Similarly, Wireshark is a free program tool thatis included in the open source software group of programs. Besides a command console, Wireshark provides a graphic user interface which largely facilitates work and tool management. Wireshark supports all major network protocols and has the option to improve a new protocol in a way that made the number of supported protocols rise to more than a hundred. Wireshark is a software tool that “understands” the structure of different network protocols, so it is capable of presenting the data form packages specialized for different protocols in an easily understandable way to users. Wireshark uses a library of “pcap” code (Packet capture) for “capturing” packages, which means that it can capture packages only from networks supported by pcap (Ethernet, IEEE 802.11, ...). Since theWireshark is an open source tool, it is relatively simple to implement program add-ons for a new protocol.

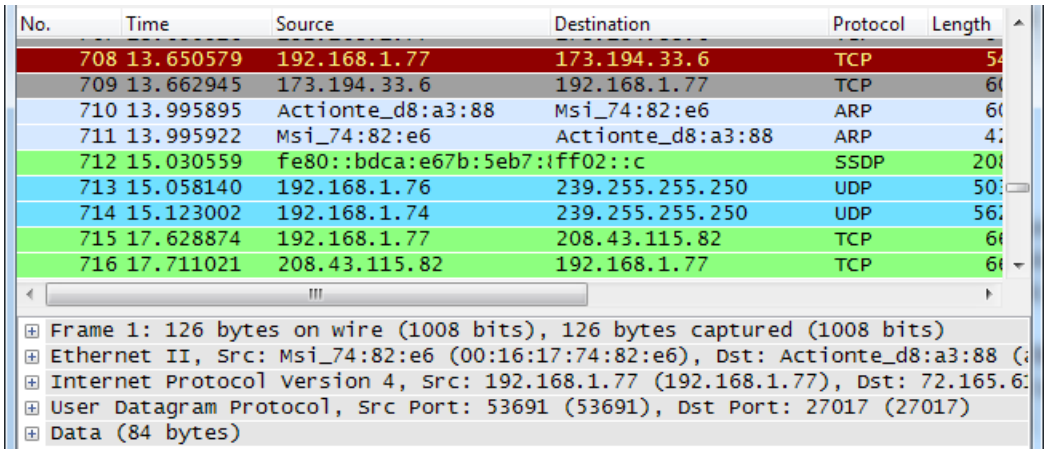

Image 6 – Data and different protocol overview

An autopsy is a tool that enables a user to analyze the hard drive or some other data storage device independently from the operating system. The user receives a review of all files, even the deleted ones. The tool enables a review of time-lapse of activity of a specific file, which is very important for analysis. The autopsy contains the functions that enable collection of the source data (images, discs, files) and their search, running the analysis module, overview of the results, content overview, and making a report. The autopsy is an expandable app, which means that despite it basically contains tools that cover most of the necessary functions, it allows customers to upgrade and alter source version of the program, which further enables improvement of the app characteristics by introducing the new ways of analysis or adapting the report according to personal customer needs. One of those added functions is the one that enables image creation and investigation of Android OS.

Image 7 – Welcome window with options

Inside the Help folder,there are the most important and the most frequent ways of the program uses, alongside with the detailed and image followed directions of the operation flow. Directions are made in the HelpNDoc tool, which is free open code. Moreover, it can make instructions in HTML, CHM, PDF, and doc format. In this project of three integrated tools within the i3A, a standard Windows CHM help file is used. The directions are followed by the images of program states created using the open code software PrintKey. The text is downloaded from the program sites and forums about the digital forensics and then translated inthe Serbian language.

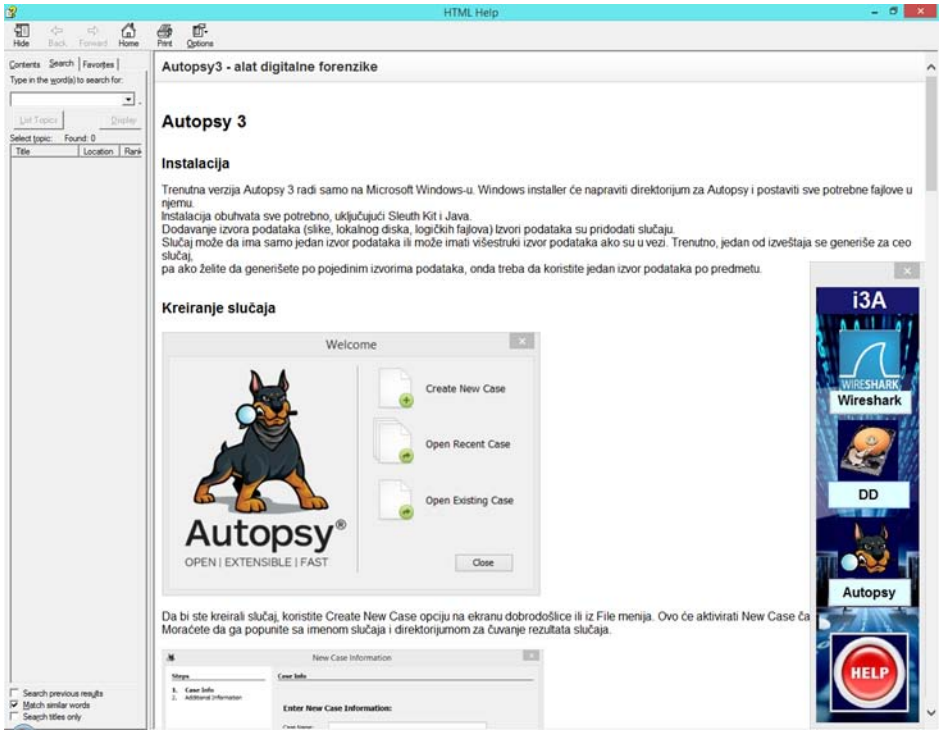

Image 8 – Help is complemented with images of state of the program, made using a free program PrintKey

Comparison of I3A forensic tool on Windows and Linux platform

To find the influences of Windows and different Linux platforms on integrated digital forensics tool I3A, it is necessary to compare theI3A versions for mentioned platforms. For thispurpose, we used a USB memory with the size of 512MB containing the documents "prazan.docx" and "proba.docx", which were previously deleted from the mentioned USB drive. In this particular case, we compare the time which is necessary that these integrated tools of digital forensics do their bit by bit recording media.

In the simulations, the time needed to perfume the analyses by mentionedtools was measured. For both integrated and non-integrated tools, the speed start tool, the speed of acquisition of the media, evidence download speed, the speed of file analysis, the speed of search by keyword, the speed of search by file types, the speed of “live” analysis and the speed of drafting the report were measured.

The period between the start time and finishing time of the i3A tool was measured. While iA3 tools start was instantly, it was much longer with the SIFT Workstation, about 1 minute and 25 seconds.

Also, the time needed to produce a report in Autopsy was monitoredand it was immediately. The measurement results are presented in Table 1.According to the results presented in Table 1, it can be concludedthat each of testedtools has its own advantages and disadvantages.

Table 1 Comparison of the most important technical characteristics of integrated and non-integrated tools[1]

Type of analysis	*SIFT Windows*	*I3A Windows*	*I3A Ubuntu*	*I3A Knoppix*
Start	1min.25sek.	Immediately	Immediately	Immediately
Acquisition	1min.52sek.	Seen 55 s	55 s	55 s
Loading Image	14sek.	14 s	14 s	14 s
Analysis of the file	19 sec.	19 s	17 s	18 s
Search by keyword	1 min.40 sec.	1 min 40 s	1 min 38 s	1 min 40 s
Search by file type	39 sec.	39 s	38 s	40 s
"Living" analysis	1min.24sek.	1 min 24 s	1min 24 s	1 min 24 s
Preparing reports	immediately	Immediately	Immediately	Immediately

The I3A tool is based and created in a way that using a GUI it is fully implemented in Windows environment, and it was tested on different Linux versions, therefore with different demands.

The functionality of I3A was tested onthe following Linux OS versions:Ubuntu, Fedora, and Knoppix, as live versions, together with desktop (Workstation) versions.

The Wine Developeris a component demanded for the operationof I3A app, and its additional package named the Wine-Mono, has the files necessary for I3A operationon the Linux platform. The runtime dot Net Framework 4, was also neededalong with an active internet connection for acquiring the right packages and installers.

Implementation of I3Ain Python programming language

The interpreted languages are programming languages that do not require an explicit compilation step.For instance, in the normal case, a C program has to be compiled before it is run, which is not the case with a JavaScript program. Therefore, JavaScript is sometimes called a "scripting" or interpreted language.

[1] Dragan Randjelovic, Damir Delija, Dragan Stojkovic, Marko Velickovic, Dragan Erlevajn, Comparing integrated and non-integrated digital forensics tools

Line betwean interpreted languages and compiled language is getting more and more blurry since compilation can be so fast with modern hardware and modern compilation techniques.For instance, the V8 represents the JavaScript engine in Google Chrome,which is widely used outside of the browsers, actually compiles the JavaScript code on the fly into machine code, rather than interpreting it.

Regardlessthe language is a "scripting" language or not, the performances dependmore on the environment than on a language. Namely, there is no reason to write a C interpreter and use it as a scripting language, or compile JavaScript to the machine code and store itin an executable file.

On the other hand, Python [2]represents an interpreted high-level programming language for general-purpose programming. It was created by Guido van Rossum and released in 1991. Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.

The diverse application of the Python language is a result of the combination of features which makethis language an edge over others. Some of the Python benefits are listed in the following.

- Presence of third party modules:

The Python Package Index (PyPI) contains numerous third-party modules that make Python capable of interacting with most of the other languages and platforms.

- Extensive support libraries:

Python provides a large standard library which includes internet protocols, string operations, web services tools and operating system interfaces. Many widelyused programming tasks arescripted into thePython standard library which reduces the code length significantly.

- Open source and community development:

Python language is developed under an OSI-approved open source license, which makes it free to use and distribute, including for commercial purposes.

Further, its development is driven by the community which collaborates withits code through hosting conferences and mailing lists and provides for its numerous modules.

- Learning ease and support available:

Python offers excellent readability and uncluttered simple-to-learn syntax which help beginners to utilize this programming language. The set of the code style guidelines, the PEP 8, provides a set of rules to facilitate the code formatting. Additionally, alarge number of users and active developers has resulted in a rich internet resource bank to encourage development and the continued adoption of the language.

- User-friendly data structures:

[2] https://www.python.org

Python has the built-in list and dictionary data structures which can be used to construct fast runtime data structures. Further, Python also provides the option of dynamic high-level data typing which reduces the length of needed support code.

- Productivity and speed:

Python has a clean object-oriented design, provides the enhanced process control capabilities, and possesses the strong integration and text processing capabilities, and its own unit testing framework, which all contribute to the improvement of speed and productivity. Therefore, Python is considered a viable option for building the complex multi-protocol network applications.

I3A is developed in such way that it could enable the digital forensics experts to automate their procedures in a simple way, by changing or adding script which would connect the work of these tools. Writing of these scripts would demand the knowledge of Python, but as it was already stated, one of the main advantages of Python is the simplicity of learning, as well as a strong community that supports this program language and offers many study material and improves the existing scripts.

Besides, Python popularity has enabled the manufacturers of these three software used in the I3A (wireshark, autopsy and dd) offer additional options of connecting using Python through some of the modules. The exact examples for that are:

- Pyreshark is a plug-in for Wiresharkthat allows other plug-in to be writteninPython.
- Autopsy-Plug-in is a repository of Autopsy Python Plug-in where there is a bunch of plug-ins already written in Python for Autopsy.

Two main issues of the development of I3A tool in Python programing language are how to make this app portable and the improvement of speed compared to the existing version of the tool.To makePythonscripts portable for the Windows platform,a portable Python distribution such aswinpython can be used. A user only needs to copy aportable distribution on the flash disk together with thePython script he wantsto run.The WinPython is a free open-source portable distribution of the Python programming language for Windows 7/8/10 for bothscientific and educational usage.

As for the Linux platform, the Python does not come on all GNU/Linux distros but is present on most of the popular Linux home user distributions, mostly because the application of Gnome desktop environment and KDE use Python 2.5+ interpreters. Since Python is integrated into the Linux system/environment from the beginning, Linux users feel easy to program in Python. However, it should be mentioned thatJava is still equally popular.

On the Mac OS, as well as the Linux platform, in most cases, Python comes pre-installed, which means that the integration with these systems is already executed, unlike the Visual Basic.

The Visual Basic programming language is faster then the Python becasuePython being as an interpreted language. The main reason interpreted languagesareslow is thatthey need to process each instruction before creating and executing a machine code.

Python is first compiled into the byte-code. However, the byte-code generated for a Python program is further interpreted, so every byte-code instruction is re-evaluated whenever it is executed.

Conclusion

Since the digital evidence is stored within a computer system, it is impossible to see the content without the help of appropriate forensic tools. There are a number of these tools, and in this paper, the I3A tool written in Visual Basic that offers solutions for some of the problems in the digital forensic field, is presented. The I3A tool has been initially written in Visual Basic programming language and it has been proven as successful in digital forensics tasks. However, due to the greater possibilities of instruction localization, possibility of automatization, as well as a possibility of more connectivity with additional modules of digital forensic tools, we suggest the implementation of I3A tool in Python programming language, which is a script language. Moreover, Pythonis not compiled, and it is relatively easy to learn. Besides, it would enable experts in the digital forensics field to automate their procedures in a simple way, by changing or adding script that would connect the work of several tools.

References

1. Dragan Randjelovic, Damir Delija, Dragan Stojkovic, Marko Velickovic, Dragan Erlevajn, Comparing integrated and non-integrated digital forensics tools
2. Dragan Randjelovic, Kristijan Kuk,Vladan Borovic, Dragan Mladenovic, Dragan Erlevajn, Testing of the operating system effect on the efficiency of forensics tool I3A
3. Python, May, 2018, https://www.python.org
4. AccessData, Jun 05, 2015, www.accessdata.com
5. Altheide, C., and Carvey, H. (2011). Digital Forensics with Open Source Tools. Massachusetts: Elsevier.
6. Brown, L. T. (2010). Computer Evidence: Collection and Preservation, Second Edition. Boston: Course Technology.
7. Casey, E. (2004). Digital Evidence and Computer Crime, Second Edition. London: Academic Press.
8. Carvey, H. (2009). Windows Forensics Analysis. USA: Syngress Publishing, Inc.
9. EnCase, Jun 05, 2015, www.encase.com
10. Forensic Focus, Jun 01, 2015, www.foren
11. Fradella, H.F., O'Neill, L., and Fogarty, A. (2004) The Impact of Daubert on Forensic Science, 31 Pepp. L. Rev. Iss. 2
12. Ignjatović, Đ. (1991). Pojmovno određenje kompjuterskog kriminaliteta. Beograd: Anali Pravnog fakulteta u Beogradu.
13. McClure, S., Scambray, J., Kurtz, G. (2006). Хакерске тајне: заштита мрежних система. (превод), Београд, Микро књига.

14. Milanovic, T., Kuk, K., Randjelovic, D., Čisar, P.(2015). Text mining techniques and identification of information by documents written (in Serbian) in High-end International Forum on Public Security Technology Informatisation, Shenyang, China, September 2015, pp. 575-583 .
15. Milosavljević, M., & Grubor, G. (2009). Digitalna forenzika - udžbenik. Beograd: Univerzitet Singidunum.
16. Milosavljević, M., & Grubor, G. (2009). Istraga kompjuterskog kriminala. Beograd: Univerzitet Singidunum.
17. Newman, C. R. (2007). Computer Forensics: Evidence, Collection and Management. New York: Auerbach Publications.
18. Pastore, M., & Dulaney E. (2007). Security + (prevod na hrvatski), Miš d.o.o.
19. Petrović, R. S. (2000). Kompjuterski kriminal. Beograd: Ministarstvo unutrašnjih poslova Republike Srbije.
20. Ranđelović, D., & Bogdanović, T. (2010). Alati za digitalnu forenziku, NBP - Žurnal za kriminalistiku i pravo, Vol. XV, No. 2, 25-47.
21. Randjelović, D., Delija, D., Popović. B. (2009). EnCase forenzički alat, Bezbednost 1-2, pp. 286-312.
22. Randjelović D. (2011). Poredjenje komercijalnih i nekomercijalnih alata digitalne forenzike i njihova upotreba Naucno tehnicka informacija, VojnoTehnicki Institut Beograd.
23. Randjelović, D. (2013). Visokotehnološki kriminal. Kriminalističko-policijska akademija, Beograd.
24. Vacca, R. J. (2005). Computer Forensics: Computer Crime Scene Investigation, Second Edition. Massachusetts: Charles River media.
25. Velickovic, M .(2016). Integrated digital forensics tools. Belgrade: Academy of Criminalistic and Police Studies.