Testing of the Hypothesis in the Research of Computer Incidents on the Basis of the Analysis of Attributes and Their Values

Igor Pantiukhin, Igor Zikratov, Anna Sizykh, Addy Christian Crosby Nii
ITMO University
Saint Petersburg, Russia
{zevall, zikratov}@cit.ifmo.ru, {anya.sizykh, jlcaddy4}@gmail.com

Abstract—**In this paper, we consider the verification of the hypothesis in the research of computer incidents study, based on the analysis of attributes and their values with post-incident computer equipment. The essence of the hypothesis constitutes the ability to study computer incidents in the types of computer memory; analyzing only the attributes and their values, without taking into account the contents of the data. The paper describes the process of forming the attributes and their values post-incident computer equipment and also, describing how to create a database of existing computer incidents. In performing these computational experiments, consists of assessing the possibilities of investigating computer incidents on the basis of the analysis of attributes and their values in the created database. Forming an opinion regarding the more informative details about the computer incidents by analysis of attributes and their values from the volatile memory, NVRAM (Non Volatile Random Access Memory), network traffic complex.**

I. Introduction

In recent times, observation show a constant increase in the number of computer incidents that occurred in the computer equipment. An important task is the study of computer incidents and in particular, the holding of post-incident internal audit. Under post-incident internal audit implies the recovery of events in information security incidents in the computer equipment. Post-incident internal audit aims to obtain data from the computer equipment and search in them for information on computer incidents that may be in non-volatile memory dumps, volatile memory, network traffic [1].

Most computer incidents that occur during the operation of computer equipment remain unexplored due to the lack of modern methods and approaches of integrated internal audit that meet the modern requirements of technology. At this point, the problem of post-incident internal audit with computer equipment data (non-volatile memory, volatile memory, network traffic) decide to separate from each other [2]. In the case of an internal audit data from computing means apart from each other a high probability of losing important data sets that will provide key information for the analysis of computer incident occurred. This is due to the fact that modern computer incidents may reflect personal information in the volatile memory, non-volatile memory, network traffic, and simultaneously in all [3-5].

For the process of post-incident internal audit of computer equipment growth of stored and processed data is a problem [6]. One possible solution is to investigate the incidents of computer-based analysis of the attributes and their values. An important feature is that the attributes and their values can be obtained both in the volatile memory and the non-volatile memory or network traffic. Also, it is possible to describe the relationship between attributes of different types of computer memory. Thus there is a possibility to get a comprehensive picture of the computer incident that occurred in the computer equipment. It should be noted that the approach based on the analysis of attributes and their values allows to describe the data not only within one computer equipment, but also in several.

The approach is based on the study of computer incidents using attributes analysis and their values as opposed to a full analysis of the data by means of computer equipment allows for carrying out the research of computer incidents in the conditions of constant growth of the volume of information processed and stored, the growth of the number of varieties of computer incidents. This approach will allow to reduce the time spent on the process of investigation of computer incidents, improve the accuracy of the information content and information about computer incidents through the use of different methods [7], [8]. In the study shows the test of the hypothesis of the possible incidents of computer-based analysis of the attributes and their values. Confirmation of the hypothesis computational experiments carried out research of computer incidents in the created database. This formed the conclusion to raise awareness of information about computer incidents through the use of attributes and analysis of their values from the volatile memory, non-volatile memory and network traffic simultaneously.

II. Attributes

A. *The process of formation of the attributes and their values with post-incident computer equipment*

By means of computer equipment, it implies a variety of devices such as personal computers, mobile phones, tablets and servers. All means of computer equipments have volatile memory, non-volatile memory, and there is the possibility of network traffic [9]. Before the process of formation of the attributes and their values, there is the need to get memory dumps. Let's consider the processes of obtaining dump with post-incident computer equipment and to furthermore, the formation of attributes and their values.

B. Obtaining and formation data attributes and their values from the non-volatile memory

Under the non-volatile memory medium is understood that after the disconnection of power or a power failure, it retains a set of stored information in it.

Examples of such memory can serve as a classic hard disk drives (HDD), SSD drives, Flash drives, EEPROM. Because of its features, physical design, software implementation, there are approaches to obtaining the maximum amount of information with the non-volatile memory medium. These methods are described in the documents in the following references [10-12]. There are two basic approaches for reading data from the non-volatile memory:

- Hardware. Under hardware, it is understood that accessing data to be read by standard software is impossible. Typically, in this case, use of hardware resources and data recovery techniques (PC-3000, DFL-DDP Data Recovery Equipment) [13].
- Software. This production method is a standard data and the most common. Typically, the storage medium is connected to the serial port and reading of information is performed through various software [14, 15].

To solve the problem of obtaining the attributes and their values to the non-volatile memory dumps used by third-party open-source library Hachoir and its components.

Hachoir is a universal platform for manipulating binary files, written in the Python language, it is independent of the operating system and is designed to study the existing files. There are times when using Hachoir library can not process the attributes of some files. In such a case each of these files must be treated separately by using the Linux system utilities and regular expressions. Based Hachoir and Linux system utilities, a table containing a set of attributes and their values has been generated for each file with post-incident dump of a non-volatile memory. An example of the Table I is shown below.

TABLE I. EXAMPLE OF ATTRIBUTES AND THEIR VALUES OBTAINED FROM THE NON-VOLATILE DATA MEMORY

HDD	
Path	/Windows/System32/DriverStore/FileRepository/prnep00a.inf_amd64_neutral_92a4c727cdf4c2f7/Amd64/EP0NH43R.DLL
File name	EP0NH43R.DLL
MIME type	application/x-dosexec
Endianness	Little endian
Creation date	2009-07-14 01:28:27
Permission	777
Format version	Portable Executable
Extension	DLL
Comment	CPU
…	…

C. Obtaining and formation data attributes and their values from the volatile memory

Considering volatile memory medium, the owner of the information, after power loss, loses the data stored. Computing tools can run on different hardware and software, so each of them can be various difficulties in accessing the volatile memory. Because of these features, retrieving data from the volatile memory can be difficult. There are the following ways to get data from volatile memory: Crush Dump, Kernel Modules, Hibernate File, IEEE1394, Hot boot acquisition, Cold boot acquisition, Virtual Machine Imaging, Hardware [16]. To maximize the amount of volatile memory data using these techniques, there are methods described in the following [17], [18]. For the formation of the attributes and their values was to develop software in Python using the Volatility framework library. Through software we have table of attributes and their values from the volatile memory. An example of the Table II is shown below.

TABLE II. EXAMPLE OF ATTRIBUTES AND THEIR VALUES OBTAINED FROM THE VOLATILE DATA MEMORY

RAM	
Offset (V)	“0xfffffa8000caf040”
Prevelegies	[2, SeCreateTokenPrivilege, “Create”]
PDB	“0x000000001d6a0000”
Process_path	“C:\Windows\system32\services.exe”
Exit-data-time	“2016-11-10 13:08:59 UTC+0000”
Dll_list	[“0x00000000772c0000”,”0x1a9000”,”0xffff”,”C:\Windows\SYSTEM32\ntdll.dll”]
Handles	[“0xfffff8a0012fd550”, “0x4”,”0x9”,”Key”,”MACHINE\SOFTWARE\MICROSOFT\WINDOWS NT\CURRENTVERSION\IMAGE FILE EXECUTION OPTIONS”]
Connection	[“0x3e32c2d0”,”TCPv6”, “:::49156”,”:::0”,”LISTENING”,”-”,”-”,”-”]
Security	[“S-1-5-18”, “Local System”]
…	…

D. Obtaining network traffic data.

Network traffic can be obtained from the places its aggregation in formats or in full dump packet headers format. Processing can be enclosed in its submission as a text csv file format, pcap, tcpdump. After that, it will be possible to handle for post-incident internal audit. The process of obtaining and analyzing network traffic are described in [19, 20]. Below are tables of attributes with values dump network traffic is given in Table III.

TABLE III. EXAMPLE OF ATTRIBUTES AND THEIR VALUES OBTAINED WITH THE DUMP OF NETWORK TRAFFIC

NET	
Time	03:00:38
Date	2016-10-02
Source	192.168.1.253
TTL	1
Destination	66.102.9.99
Source Port	1985
Protocol	HTTP
Destination Port	1985
...	...

III. Formation of Database of Existing Computer Incidents

To test the hypothesis of the use of attributes and their values, obtained from the computer equipment in the problems of the study of computer incidents, it is necessary to form the basis of existing computer incidents. To fill the database, the site malwr.com was chosen [21], which contains statistical data study various malicious files. Tracing in a sandbox allows you to keep track of the changes that makes malicious software. These changes are shown in the dumps memory as change the attributes and their values. Therefore, analyzing the statistics of the database can define attributes and their values, which refer to a computer incident. Existing computer database of incidents of data formed by parsing malwr.com site, and structure are shown in Fig. 1.

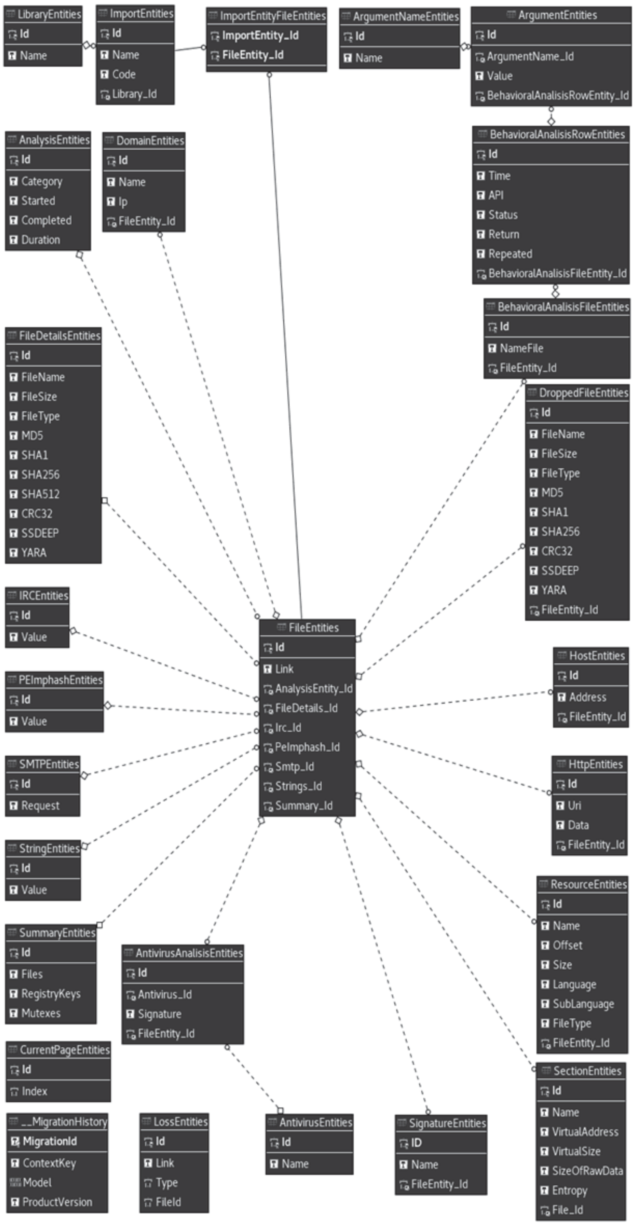
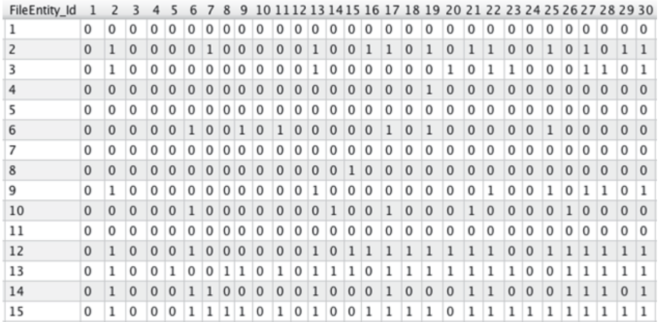


Fig. 1. The structural database of existing computer incidents

The formed database contains sections such as: Quick Overview, Static Analysis, Behavioral Analysis, Network Analysis, Dropped Files, Comment Board. The database stores all the attributes obtained during the verification file, but many of them are informative, for example, FileDetailsEntities (name, size, hash values), DroppedFilesEntities, or YARA (text and binary patterns contained in the examples of the family of malicious software) as initial values can be hidden. An example is the utility for GhostRat remote access (used in many documented attacks), in which the network connection begins with the magic word Gh0st (it is registered in the default settings) that allows you to monitor its work, but the source code Gh0stRAT are freely available, and it is means that the magic word can modify each user [22].

The study of attributes and their values in this research will be carried out in two ways: static analysis and behavioral analysis. Under static analysis, it refers to the study of attributes and values of the following database tables: PEImphashEntities [23], SectionEntities, ResourceEntities and ImportEntityFileEntities. Also, we refer to the static analysis indicators antivirus reaction to the checked file (AntivirusAnalisisEntities).

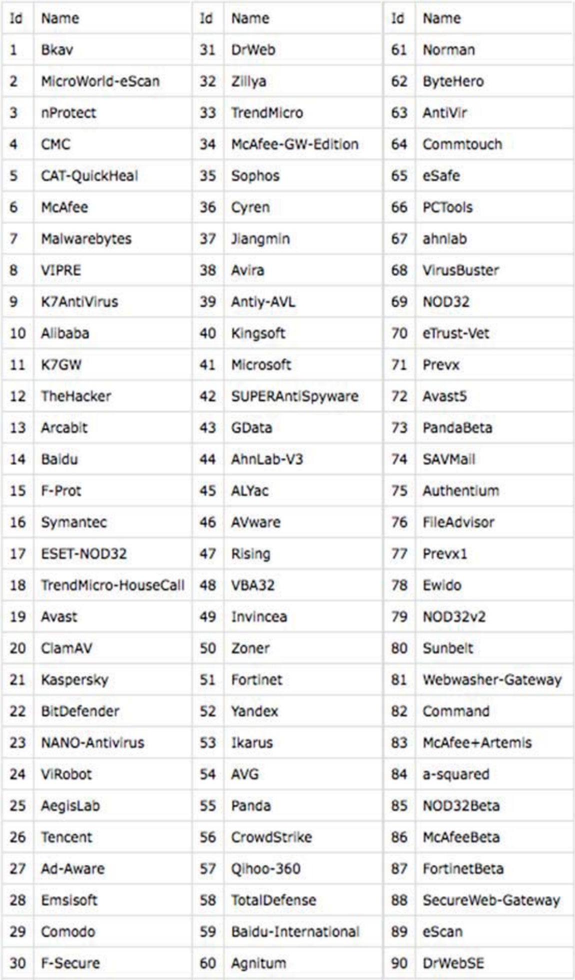
By analyzing the database, it became evident that important information for the study of computer incident response is an antivirus on a suspicious file. This follows the fact that the verification process is in the sandbox, where each of these objects is started and tested, including taking into account all requests and actions made by them in the system, and after this, a decision is made about the nature of the suspicious file. For clarity, we form a tuple response indicator values for each of the antivirus files. For this was written Python code that using standard libraries formed necessary to analyze the file with the extension csv. When checking each file, antivirus can answer "Clean" or for example "Win32.Trojan.Raas.Auto" or "Trojan.PDF.Phishing.DT", in the preparation of a tuple, it was decided that the value of the parameter equal to "Clean" will be replaced to "0" (safe), while all other values are regarded as "1" (malicious), an example of the result is shown in Fig. 2.



FileEntity_Id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	1	0	1	1	0	0	1	0	1	0	1	1
3	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	1	1	0	1
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1
10	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
13	0	1	0	0	1	0	0	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1
14	0	1	0	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	1	1	1	0	1
15	0	1	0	0	0	1	1	1	1	0	1	0	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1

Fig. 2. Part of the tuple with the parameters showing the antiviruses' reactions to suspicious files

In total, the database contains information research of suspicious files with 90 antivirus solutions. Fig.3 provides a list of the antivirus tools.



Id	Name	Id	Name	Id	Name
1	Bkav	31	DrWeb	61	Norman
2	MicroWorld-eScan	32	Zillya	62	ByteHero
3	nProtect	33	TrendMicro	63	AntiVir
4	CMC	34	McAfee-GW-Edition	64	Commtouch
5	CAT-QuickHeal	35	Sophos	65	eSafe
6	McAfee	36	Cyren	66	PCTools
7	Malwarebytes	37	Jiangmin	67	ahnlab
8	VIPRE	38	Avira	68	VirusBuster
9	K7AntiVirus	39	Antiy-AVL	69	NOD32
10	Alibaba	40	Kingsoft	70	eTrust-Vet
11	K7GW	41	Microsoft	71	Prevx
12	TheHacker	42	SUPERAntiSpyware	72	Avast5
13	Arcabit	43	GData	73	PandaBeta
14	Baidu	44	AhnLab-V3	74	SAVMail
15	F-Prot	45	ALYac	75	Authentium
16	Symantec	46	AVware	76	FileAdvisor
17	ESET-NOD32	47	Rising	77	Prevx1
18	TrendMicro-HouseCall	48	VBA32	78	Ewido
19	Avast	49	Invincea	79	NOD32v2
20	ClamAV	50	Zoner	80	Sunbelt
21	Kaspersky	51	Fortinet	81	Webwasher-Gateway
22	BitDefender	52	Yandex	82	Command
23	NANO-Antivirus	53	Ikarus	83	McAfee+Artemis
24	ViRobot	54	AVG	84	a-squared
25	AegisLab	55	Panda	85	NOD32Beta
26	Tencent	56	CrowdStrike	86	McAfeeBeta
27	Ad-Aware	57	Qihoo-360	87	FortinetBeta
28	Emsisoft	58	TotalDefense	88	SecureWeb-Gateway
29	Comodo	59	Baidu-International	89	eScan
30	F-Secure	60	Agnitum	90	DrWebSE

Fig. 3. The list of antivirus software

On the basis of data from the tuple are reflected in the graphs below, showing the dependence of the tuple of values from the reaction of the antivirus tools. Fig. 4 is a graph which shows the performance of each of the proposed antivirus on the example of 10,000 files. As seen in Fig. 4, generally only the first 60 antivirus software are informative.

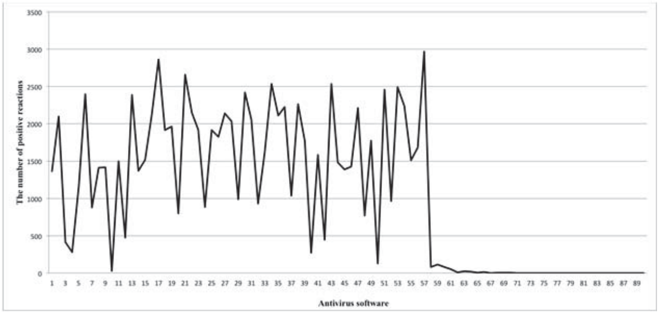


Fig. 4. The effectiveness of antivirus software

The database contains statistical information about the study, 180 thousands of suspicious files. But as can be seen from Fig. 5 and Fig. 6, which shows the response of the antivirus, statistically (samples do not cross) to 10000 and 2000, respectively, suspicious files, not all of them are defined as malicious at least 10 software from 90 which were considered.

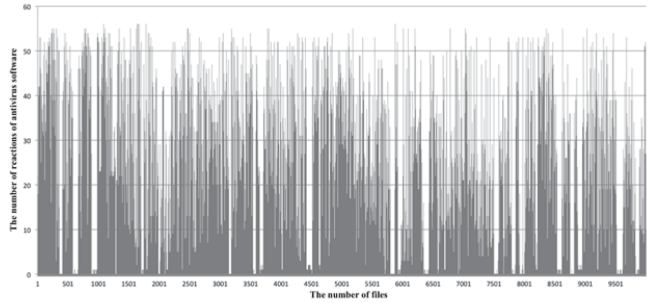


Fig. 5. Statistics of the response of antivirus software for 10000 suspicious files

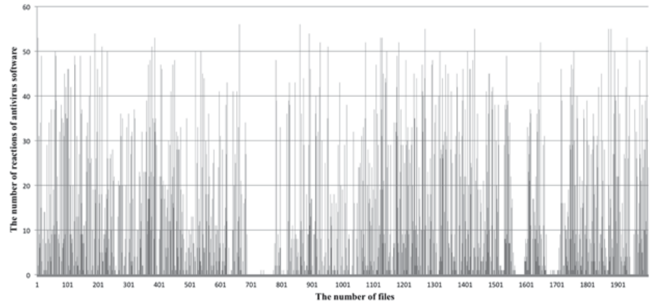


Fig. 6. Statistics of the response of antivirus software for 2000 suspicious files

Under the behavioral analysis is meant to obtain the attributes of a database table BehavioralAnalisesRowEntities, which are responsible for a system of data collection response to the actions of the file/process, for example, modify the registry values, or call the system libraries.According to data from BehavioralAnalisesRowEntities API function calls (for example, GetSystemTimeAsFileTime, RegOpenKeyExW, LdrGetDllHandle) by categories. Fig. 7 event (the API function which was called) is reflected X-axis, Y-axis - to which category the event.

Categories are responsible for the division of API functions on the basis of their influence and intervention. Total allocated 11 categories: hooking, windows, services, misc, device, synchronization, threading, process, file system, registry, system. For behavioral analysis it is convenient to make a tuple which contains the status of the implementation of API functions (success or failed) for each scanned file. Title success (denoted by 1) indicates that there was a successful implementation of the action API functions, such as NtWriteFile - opening a file. While the failed status (denoted by 0) means that the implementation of the action is denied, for example NtQueryAttributesFile - check file existence. Fig. 8 and Fig. 9 are graphs of the status of implementation of API requests functions from the function. Fig. 8 are considered 311 requests API functions, in Fig. 9 considered 155 requests API functions.

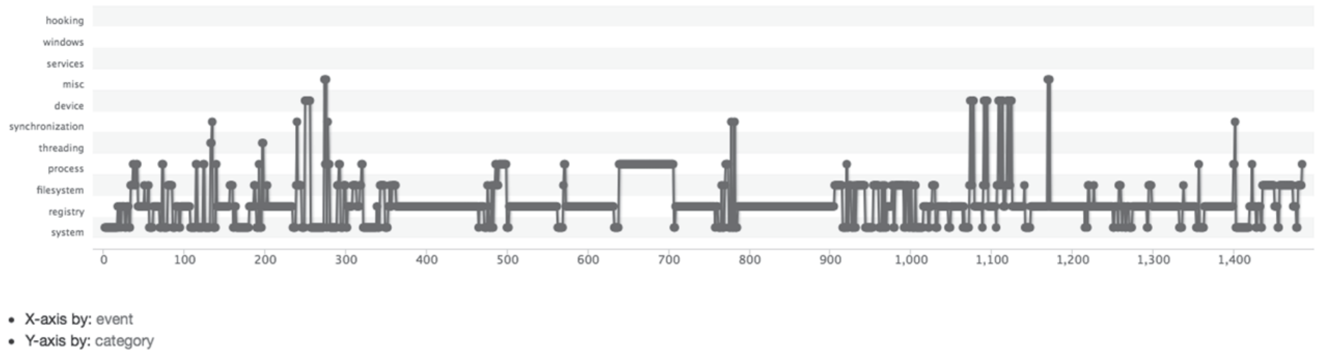


Fig. 7. Dependence of the API functions by category

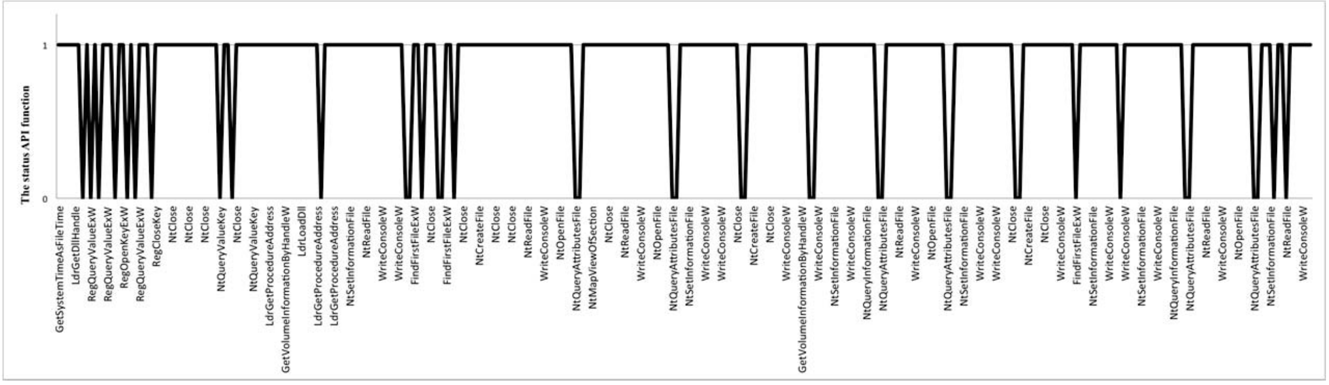


Fig. 8. Status 311 queries of API functions

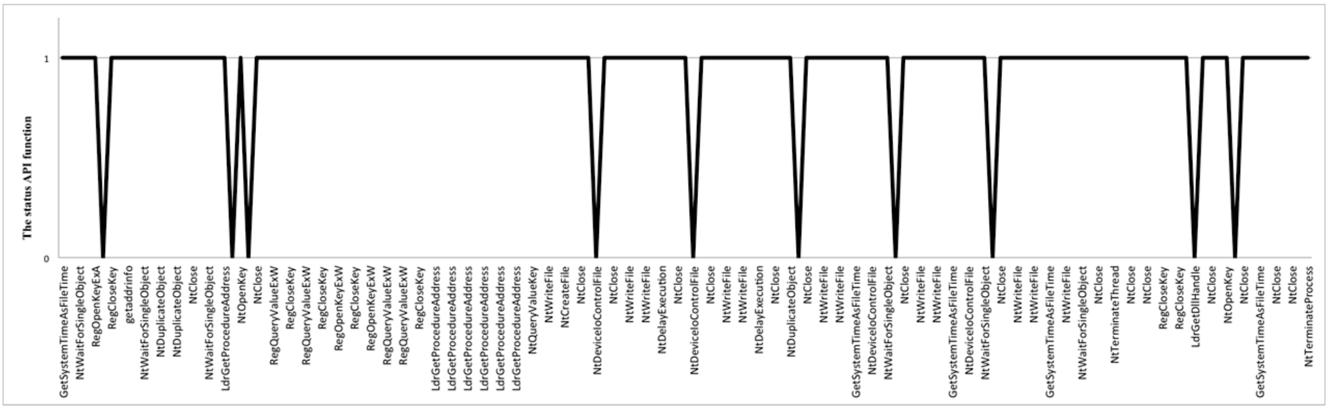


Fig. 9. Status 155 queries of API functions

As can be seen from Fig. 8 and Fig. 9 most of the API calls have ended successfully, and the fact that many API functions have undocumented features, even the mere receipt of information about a file may lead to an emergency stop system (NtSetInformationFile from ntdll.dll). On the status of the request and the purpose API functions, you can clearly see what action initializes each file of computer equipment, which will track and monitor unauthorized activity file/process in the system. Thus by examining static and behavioral tests it can be concluded that the following group of attributes can be used to study computer incidents:

1) PEImphash - PEImphashEntities.
2) Information about partition (name, virtual address, virtual size, size of raw data, entropy) - SectionEntities.
3) Resource information (name, offset size, language, sub-language, file type) - ResourceEntities.
4) Information on the import system libraries - ImportEntityFileEntities.
5) Indicators of response to antivirus on checked file - AntivirusAnalisisEntities.
6) Status queries API functions, given their purpose - BehavioralAnalisesRowEntities.

For a comprehensive analysis of computer incident, it is necessary to collect all these attributes together, which will give a complete picture of the impact of each file/process of computer equipment, and will help in the conduct of post-incident internal audit. Getting the maximum number of attributes and their values from the volatile memory, non-volatile memory and network traffic will allow for the acquisition of the greatest amount of information about computer incidents. Applying search techniques and on this foundation have a database can used to be carry out the study of computer incidents by analyzing the attributes and their values in memory dumps of post-incident computer equipment.

IV. Conclusion

In this paper we tested the hypothesis in the research of possible computer incidents based on analysis of the attributes and their values obtained with post-incident computer equipment. Database analysis showed the possibility of using the formed attributes and their values in a study of computer incidents that may be in dumps of volatile memory, non-volatile memory, network traffic.

The approach to the study of computer incidents on the basis of the analysis of attributes and their values has advantages in terms of constant growth in the number of varieties of computer incidents, increase in the number of volumes of stored and processed data. Investigation of computer incidents based on analysis of the attributes and their values, thus achieving reduction of the volume of processed information, and later, and time-consuming to investigate computer incidents post-incident computer equipment.

The study of computer incidents, based on the attributes of the analysis and their values has a practical application and can be used in the development of predictive systems of protection against computer incidents in the task of reducing time spent in the study of computer incidents in computer forensics in the development of automation systems in the investigation of computer incidents in large volumes of data.

Currently, the determination of computer incident is achieved by comparing the (search) post-incident attributes of computer equipment to the database. It is planned on the basis of the database obtained by applying machine learning techniques to build a system that will accurately, without using a computer database, to determine the incident data post-incident of computer equipment.

References

[1] I.S.Pantiukhin, I.A.Zikratov, A.B.Levina, “Graph-based post incident internal audit method of computer equipment”, *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, vol.16, no.3, 2016, pp. 506–512.

[2] J.T.Luttgens., M.Pepe and K.Mandia, *Incident response & computer forensics. Third Edition.* McGraw-Hill Education Group, 2014.

[3] PrivateCore official website, Physical Memory Attacks, Web: https://privatecore.com/resources-overview/physical-memory -attacks/.

[4] M. Bishop, *An Overview of Computer Viruses in a Research Environment.* Dartmouth Computer Science Technical Report PCS-TR91-156, 1991.

[5] H.Choi, H.Lee, H.Kim, “Fast detection and visualization of network attacks on parallel coordinates”, *Computers & Security*, vol.28, no.5, 2009, pp. 276-288.

[6] I.A.Zikratov, I.S.Pantiukhin, A.S.Sizykh, “The method of classification of user and system data based on the attributes”, *Open Innovations Association and Seminar on Information Security and Protection of Information Technology (FRUCT-ISPIT)*, 2016 18th Conference of. – IEEE, 2016. – pp. 404-409.

[7] I.A.Zikratov, I.S.Pantiukhin, I.E.Krivtsova, N.K.Druzhinin, “The method of elf-files identification based on the metric classification algorithms”, *Open Innovations Association and Seminar on Information Security and Protection of Information Technology (FRUCT-ISPIT)*, 2016 18th Conference of. – IEEE, 2016. – pp. 397-403.

[8] I.V.Yurin, I.S.Pantiukhin, “Testing the hypothesis of creating a digital polygraph based on video and audio data”, *Vestnik gosudarstvennogo universiteta morskogo i rechnogo flota imeni admirala S.O. Makarova,* vol.3, no.31, 2015, pp. 202-209.

[9] A.S.Tanenbaum and T.Austin, *Structured Computer Organization. Sixth Edition.* Published by Prentice Hall, 2012.

[10] L.Volonino and R.Anzaldua, *Computer forensics for dummies*. John Wiley & Sons, 2008.

[11] M.G.Solomon, D.Barrett and N.Broom, *Computer forensics jumpstart*. John Wiley & Sons, 2011.

[12] B.Nelson, A.Phillips and C.Steuart, *Guide to computer forensics and investigations*. Cengage Learning, 2014.

[13] Chris Kasperski, *Data recovery. A practical guide.* Published by BHV-Petersburg, 2006.

[14] G.E.Sienkiewicz, *Art data recovery*. Published by BHV-Petersburg, 2011.

[15] P.A.Tashkov, *Data Recovery 100%.* Published by Piter, 2008.

[16] M.Burdach, *Physical memory forensics*. USA: Black Hat USA, 2006.

[17] M.H.Ligh, A.Case, J.Levy and A.Walters, *The art of memory forensics: detecting malware and threats in windows, linux, and Mac memory*. John Wiley & Sons, 2014.

[18] A.Case, G.G.Richard, “Memory forensics: The path forward”, *Digital Investigation*, 2017.

[19] S.Davidoff and J.Ham, *Network forensics: tracking hackers through cyberspace*. Published by Prentice hall, 2012.

[20] N.Meghanathan, S.R.Allam and L.A.Moore, “Tools and techniques for network forensics”, *International Journal of Network Security and its Applications*, vol.1, no.1, April 2009, pp. 14-25.

[21] Malwr official website, malwr, Web: https://malwr.com.

[22] Norman, “The many faces of Gh0st Rat”, Web: http://download01.norman.no/documents/ThemanyfacesofGh0stRat.pdf

[23] FireEye official website, Tracking Malware with Import Hashing, Web: https://www.fireeye.com/blog/threat-research/2014/01/tracking-malware-import-hashing.html.