Framework for Detecting Ransomware from Document files

Akshay Chauhan
M.E (IT)
L.J Institute of Engineering and Technology

Krunal Panchal
Assistant. Professor. PG Department
L.J Institute of Engineering and Technology

***Abstract* - Along with the rapid growth of new science and technology, the functions of our devices become more and more powerful. In spite of that, everything has two aspects. Devices bring so much convenience for people and also bring the security risks at the same time. Malicious application has become a big threat to the mobile and computer security. Thus, an efficient security analysis and detection method is important and necessary. As a new type of malicious software, ransomware is one of the biggest security threats in recent years. A ransomware encrypts the victim's valuable data, or denies the access to the device and asks for ransom money. Paying the ransom money, however, may not guarantee recovery of the data being encrypted or locked. To protect users, a detection method should be able to detect ransomware in the user's real-time environment and make it difficult for the ransomware to avoid detection.**

Keywords* – *ransomware; malware detection; malware analysis; behavioural detection;

I. INTRODUCTION

In general, a ransomware is a type of malware that after infecting a computer system, restricts the access to the system or its resources, and demands a ransom paid to the creator(s) of the malware for the restriction to be removed [2]. The attackers managed two ways to encrypt the information people have. The first one is through e-mails and establishing themselves in the networks. The second, by exploiting vulnerabilities in public web servers. This last one happens after traversing the network, using legitimate tools to identify and infect hundreds of computers (GMS, 2016) [1]. Ransomware is a type of malware which encrypts different types of files and/or disables computer functionality. Then, it requests a payment from victims in order to send the decryption key to restore system files and capabilities.

II. INTRODUCTION TO VARIOUS RANSOMWARE DETECTION METHODS

There are mainly two methods used to detect ransomware.

(1) Signature Based Detection Method
(2) Anomaly Based Detection Method

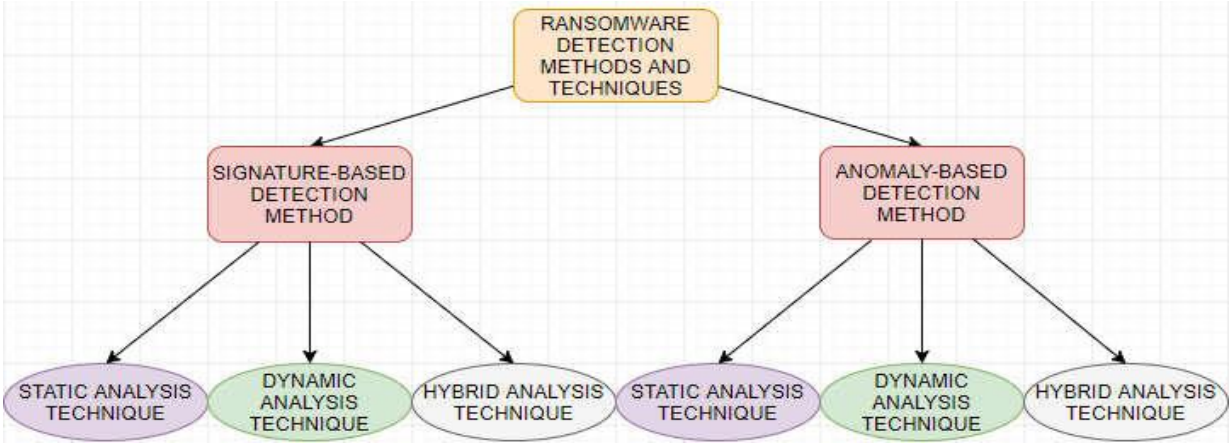

Fig-1 Ransomware Detection Methods [14]

(1) **Signature-Based Detection Method:** In this method, ransomware is detected by such signature information. Storage area such as repository is needed for storing signatures. Current signature based ransomware systems are largely based on syntactic signatures. These systems are equipped with a database of regular expression that are consider malicious.
(2) **Anomaly Based Detection Method**: In this method, it is essential to have the knowledge to decide whether the software is harmful. Rule are sets to determine whether the software is defining its rule sets. But the anomaly based detection system will work steadily as long as the rule sets ae well defined before.

i. Static Analysis

Static Analysis is a method of analysing the structure or code of the malicious program to determine the behaviour or its function. In this method, we do not execute the malware for analysis purposes. This method needs more technical expertise to analyse. If we have any suspicious sample, static analysis can confirm that sample is malicious or not. Static analysis can tell about the signature of the network activity or details about command and control server. Static Analysis includes several processes to analysis the suspicious sample. Some of those processes are enlisted below-

File Fingerprinting: Hashing is a simple method to calculate the hash value of the malware. It helps to identify the malware. This process is also called the file fingerprinting. Secure Hash Algorithm 1 (SHA1), Message Digest 5 (MD5) are most popular hashing method used for file fingerprinting.

Example- MD5 hashing of Saturn Ransomware is bbd4c2d2c72648c8f871b36261be23fd.

File Identification: File identification is a process to determine the type of the file that you are examining. Attackers maybe disguise the malicious files to changing the extension or make it self-extractable archive. Every file type such as document and executable start and end with specific bytes called magic bytes. File Identification tools like TriD use these magic bytes as a signature to determine the type of the file.

Example- All executable file start with MZ in bytes.

Packer Detection: Packer detection is one of the important methods to analysis the malicious sample to identify the sample is packed or not. Basically, packers are used to compress & encrypt executable hide their internals from researchers.

PE Header Analysis: PE Header of an executable provides the information about the operating system it needs to run the program. And it can tell us how the malware interacts with the system. It also tells us about the timeline of the sample like when and where the sample is compiled. The operating system uses this PE header for two reasons. As a container, it contains all of the information needed for the operating system to run the program. And as a director, it directs the operating system where the pieces of the executable go in the memory. PE Header contains File-header, optional-header, file sections etc.

Strings Analysis: String analysis is a process to examine the group of readable characters from a file. Looks for phrases and words in the malware which have means to us. Examination of strings can tell us how the malware behaves when it will be executed. Strings are mainly two types- ASCII based and UNICODE based.

Reverse Engineering: Reverse Engineering is a most advanced and effective method for malware analysis. It needed highly technical knowledge to use it. Basically reverse engineering is a method of understanding how program work and use the information to do something. It is used to find vulnerability in a program, malware analysis, search for sensitive data from the code, and modify the working functionality of a program. In Reverse engineering there are two process researchers are mainly use- Decompilation and Disassembly. Decompilation is a process of converting executable to a readable format. This process is helpful for a security researcher to know about the code of the sample. Disassembly is a process to review the code in machine level language. In this method, researchers can't convert the code into the high-level language.

ii. Dynamic Analysis

Dynamic analysis is a process to examine the behaviour, understanding the functionality of the malware by executing in real time. By the help of dynamic analysis, we can monitor operating system behaviour and malware persistence mechanism. By the registry, process and network level analysis we can detect the indication of compromise. The dynamic analysis also reveals the communication between command and control server.

In the dynamic analysis, we mainly focus on what are files are created, modified and deleted by the malware. Malware can also be hidden by modifying the registry values and changing the security system that is also we are going to detects by dynamic analysis. And we also monitor the processes activity that controlled by the malware and persistence mechanism.

In dynamic analysis, we can also use some pre-configured sandbox such as Hybrid Total, EM-Analyzer Pro, Cuckoo Sandbox, Joe Sandbox etc. These sandboxes can easily analysis the suspicious sample and give the detailed report in a less time.

III. LITERATURE SURVEY

A. A RELAVANCE SURVEY ON RANSOMWARE DETECTION METHODS

In this paper [1] This work will use Big data analysis tools with the application of learning machines. The purpose is to identify behaviour patterns of Ransomware before their attack on public networks. The aim of this research is to design a Ransomware detection and prevention model, by identifying patterns and suspicious behaviours through tools that interpret, learn and process security intelligence.

In this paper [2] the 2entFOX architecture is proposed to detect ransomware. 2entFOX targets the most dangerous and most destructive ransomwares means HSRs. Meanwhile, the detection of the low survivable ransomwares (LSR) is not so important. Although we can provide a system for all ransomwares detection, but this system would have some challenges. The most significant challenge is for detecting all type of ransomwares we should change the top feature weights in conditional probability tables (CPT) that causes some weaknesses.

In this paper [3] they have proposed a new detection method that inspects the consistency between the displayed document contents and the user's document editing operations. By using human file-operating characteristics as a whitelist, the proposed method achieves ransomware detection in the user's real-time environment, and it is difficult for the ransomware to avoid detection. With user's document-editing characteristics, we can detect Encryptors when a non-human Encryptor tries to manipulate a document. If detection occurs while an Encryptor is editing a document, then immediately any subsequent file access to the document file is banned.

In this paper [4] the RKiller system is proposed to detect ransomware. RKiller has 3 components, Implementing core system is an important part of the RKiller as it handles all the major operations of the overall solution, Next important component of the proposed solution is the Threat Intelligence Gathering module, and the last component is Proactive Monitoring System. The PMS is the main Ransomware detection mechanism implemented in the proposed solution R-Killer.

In paper [5] The ransomware detection model based on V-detector negative selection algorithm is proposed to detect ransomware. V-detector negative selection algorithm has three phases. V-detector generation, V-detector optimization and ransomware detection.

B. COMPARISON TABLE

Sr. No	Paper Title	Author	Methodology/Classification	Future Work
1	Large Scale Ransomware Detection by Cognitive Security	Juan A. Herrera Silva, Myriam Hernández-Alvarez	Machine Learning Algorithm, Big data analysis tools with the application of learning machines	The investigator will develop a practical application which will identify behaviour patterns of Ransomware before their attack on public networks
2	2entFOX: A Framework for High Survivable Ransomwares Detection	Mohammad Mehdi Ahmadian , Hamid Reza Shahriari	A framework for high survivable ransomwares detection based on twenty appropriate features.	The detection of LSRs will be added to 2entFOX architecture.
3	Ransomware Detection Considering User's Document Editing	Toshiki Honda, Kohei Mukaiyama, Takeharu Shirai	The detection method is composed of a DLL injector, which injects a DLL into the software, and a DLL file, which monitors API calls and detects Encryptors	The proposed method will not applicable only for text documents but also for various contents (e.g., Image Editor, Excel, PowerPoint, etc.), by using sophisticated image matching instead of OCR.
4	R - Killer: An Email Based Ransomware Protection Tool	Bathiya Lokuketagoda, Medhavi Prathibha Weerakoon, Udara Madushan Kuruppu	Proposed solution can be divided into three main components. Each of these components play a crucial role in successfully isolating a ransomware based email.	It will be fully automate the detection procedure for overall system improvement, The proposed methodology can be further improved to detect not only ransomware, but also all malicious files propagated through any network-based or even removable devices.
5	Ransomware Detection Based on V-detector Negative Selection Algorithm	Tianliang Lu, Lu Zhang, Shunye Wang, Qi Gong	It works artificial immune system, and V-detector	Nil

			generation, V-detector optimization and ransomware detection with the use of ransomware behavioural classification.	

IV. PROBLEM STATEMENT

According to the literature survey there are certain limitations regarding detection of ransomware. All the detection methods and antiviruses can not detect malicious script written in macro file. Most of the ransomware attacks have done through the office files. To avoid the problems we have to detect the ransomware from macro script.

V. PROPOSED MODEL

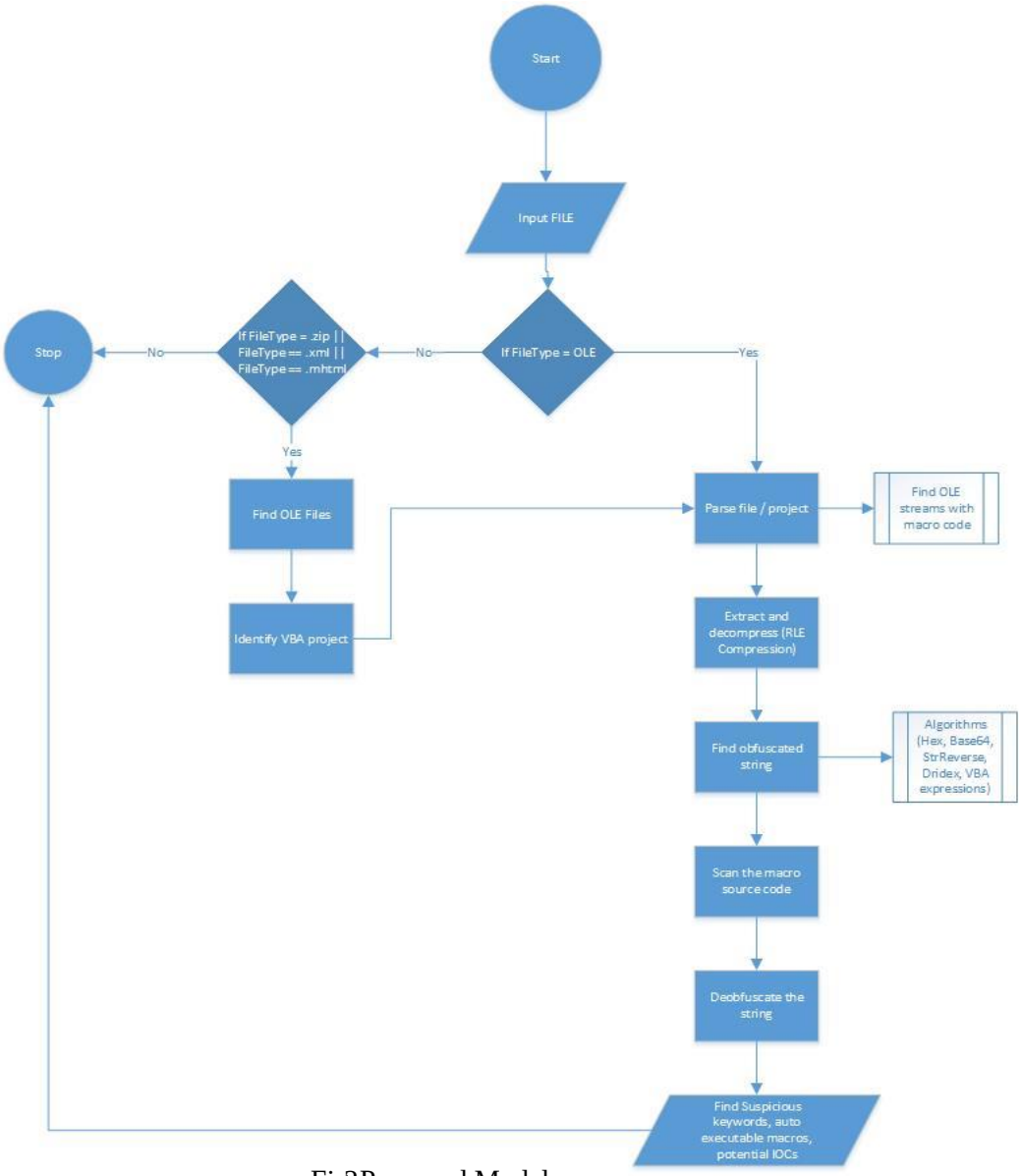

Fig2 Proposed Model

Fig 2 Proposed Model

Step-1: My python script checks the file type: If it is an OLE file (i.e MS Office 97-2003), it is parsed right away.
Step-2: If it is a zip file (i.e. MS Office 2007+), XML or MHTML, My python script looks for all OLE files stored in it (e.g. vbaProject.bin, editdata.mso), and opens them.
Step-3: My python script identifies all the VBA projects stored in the OLE structure.
Step-4: Each VBA project is parsed to find the corresponding OLE streams containing macro code.
Step-5: In each of these OLE streams, the VBA macro source code is extracted and decompressed (RLE compression).
Step-6: My python script looks for specific strings obfuscated with various algorithms (Hex, Base64, StrReverse, Dridex, VBA expressions).
Step-7: My python script scans the macro source code and the deobfuscated strings to find suspicious keywords, auto-executable macros and potential IOCs (URLs, IP addresses, e-mail addresses.

VI. RESULTS

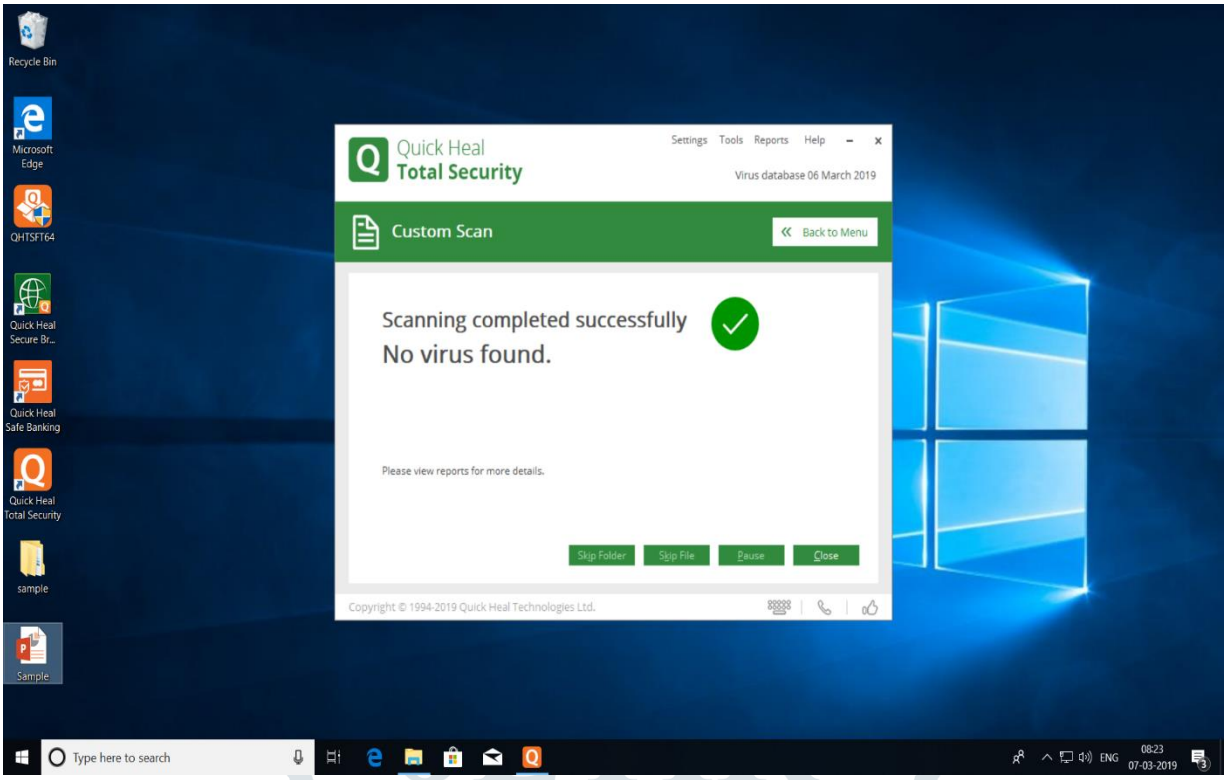

Fig 3 office file scanning with quick heal total security

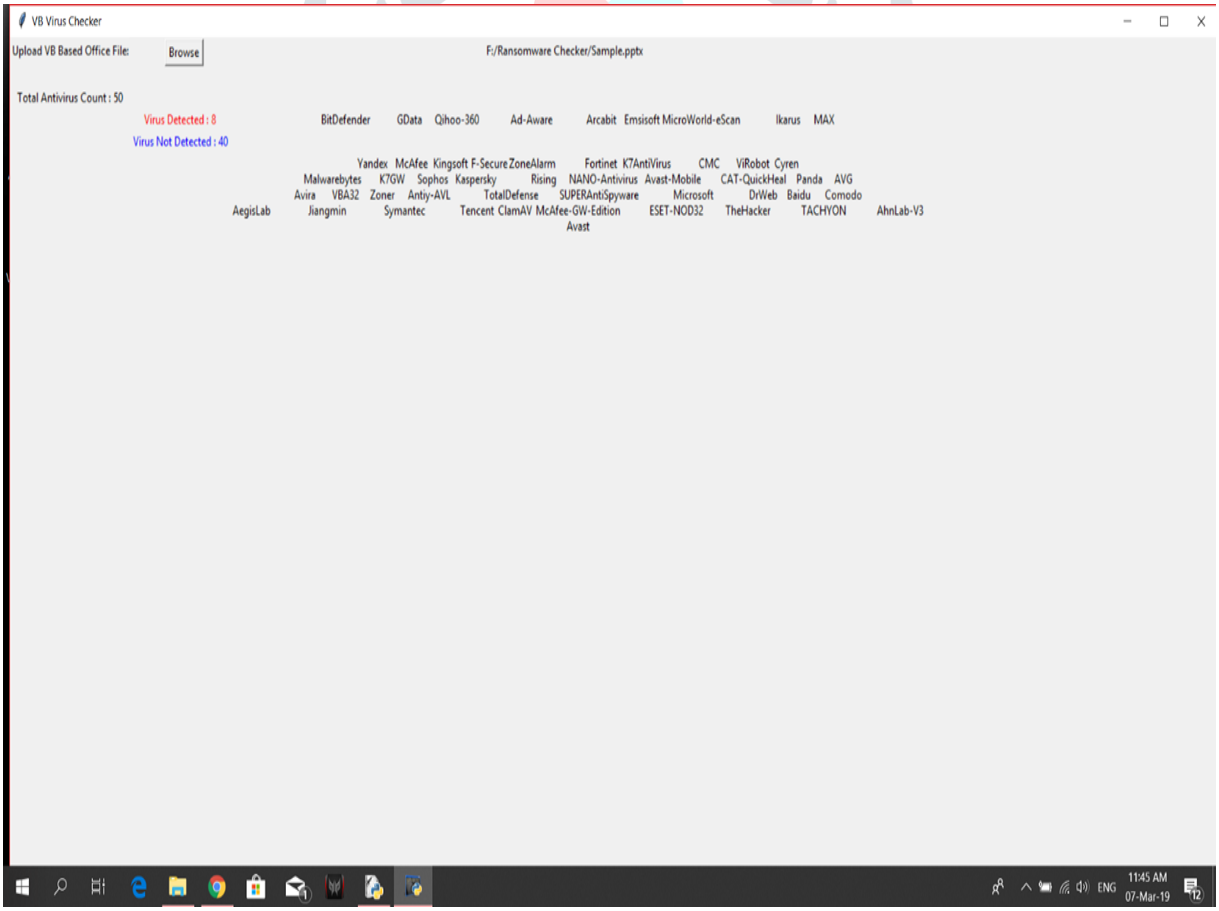

Fig 4 office file scanning with our tool

VII. CONCLUSION

The proposed system will detect the ransomware which is undetectable through the antivirus software. The whole process is manual, we are aim to develop fully automate system to detect ransomware from macro scripts.

REFERENCES

1. Juan A. Herrera Silva, Myriam Hernández-Alvarez (2017) "Large Scale Ransomware Detection by Cognitive". IEEE
2. Mohammad Mehdi Ahmadian , Hamid Reza Shahriari (2016). "2entFOX: A Framework for High Survivable Ransomwares Detection". IEEE
3. Toshiki Honda, Kohei Mukaiyama, Takeharu Shirai (2018). "Ransomware Detection Considering User's Document Editing". IEEE
4. Bathiya Lokuketagoda, Medhavi Prathibha Weerakoon, Udara Madushan Kuruppu (2018). "R - Killer: An Email Based Ransomware Protection Tool". IEEE
5. Tianliang Lu, Lu Zhang, Shunye Wang, Qi Gong (2017). "Ransomware Detection Based on V-detector Negative Selection Algorithm" IEEE
6. Chris Moore (2016). "Detecting Ransomware with Honeypot techniques". Cybersecurity and Cyber forensics Conference
7. May Medhat, Samir Gaber, Nashwa Abdelbaki (2018). "A New Static-based Framework for Ransomware Detection". IEEE
8. Dae-Youb Kim, Geun-Yeong Choi, and Ji-Hoon Lee, Member (2018) "White List-based Ransomware Real-time Detection and Prevention for User Device Protection". IEEE
9. Tianda Yang, Yu Yang, Kai Qian, Dan Chia-Tien Lo, Ying Qian, Lixin Tao (2018) "Automated Detection and Analysis for Android Ransomware" IEEE
10. Vinayakumar R, Soman KP, K.K.Senthil Velany, Shaunak Ganorkar (2017). "Evaluating Shallow and Deep Networks for Ransomware Detection and Classification". IEEE
11. Kul Prasad Subedi, Daya Ram Budhathoki, Bo Chen, Dipankar Dasgupta (2017). "RDS3: Ransomware Defense Strategy by Using Stealthily Spare Space". IEEE
12. Khaled Alrawashdeh, Carla Purdy (2018). "Ransomware Detection Using Limited Precision Deep Learning Structure in FPGA". IEEE
13. Alireza Karimi, Mohammad Hosein Moattar (2017). "Android Ransomware Detection Using Reduced Opcode Sequence And Image Similarity". IEEE
14. https://www.researchgate.net/publication/326191046_Ransomware_Detection_and_Prevention_Techniques_Cyber_Security_Malware_Analysis
15. Ransomware by Timothy Gallo, Allan Liska, EBook