Dynamic Analysis on Crypto-ransomware by using Machine Learning: GandCrab Ransomware

S Usharani¹, P Manju Bala², M Martina Jose Mary³

¹,²Associate Professor, Dept of Computer Science and Engineering, IFET college of Engineering, Villupuram, India.
³PG Scholar, Dept of Computer Science and Engineering, IFET college of Engineering, Villupuram, India.

Abstract. A ransomware is a unique class of malware which has gotten extremely famous in digital crooks to corkscrew cash. It categorizes the client confines by accessing their machines (PCs, cell phones and IoT gadgets) unless the payoff is paid. Consistently, security specialists report numerous types of ransomware assaults, including ransomware families. User’s data will be collected at the time of dynamic process. The collected data will be in crypto ransomware type from that we can extract features like IP address, file length, URL. We will do dynamic analyse of the presently data with the antecedent data. Using machine learning algorithm (by combining Random Forest, Gradient Tree Boosting and Support Vector machine algorithm) we can classify the data as benign or ransomware. The achievement rate of classification using machine learning algorithm is 98.45% with false rate 0.01. The proposed achievement rate will be compared among linear regression, naive Bayes and adaboost algorithm. Gandcrab ransomware-Version, algorithm is to be identified. Keywords: GandCrab, Ransomware, Dynamic analysis, IOT.

1. Introduction

Ransomware is a kind of prevailing malware that is sprawling recently, affecting significant economical losses to wide ranging victim’s files including numerous organizations, medical management’s and personal user’s. Cybercrime is significant nowadays, adding vast security to all the devices and also endangering conventional business activities via jeopardizing security device. The hijackers monitors the users and collects the data including illegal transactions and threatens the users. Ransomware is a malware program conducted to prevent users from accessing their device and data system. Obviously, ransomware looks like malware and infects computers from the vectors (email, botnets, and macros). Usually, Ransomware is silent on a computer device, and makes itself known to the user only after its encryption processes has been completed. According to Cyber security Ventures, victims have paid probably around $11.5 billion in 2019, and a new victim will fall to ransomware for every 14 seconds in 2019, for every 11 seconds by 2021. The highly connected Internet world and cloud simplifies ransomware dissemination including several communication protocols. It enables challengers to launch various attack campaigns- botnets available for hire, Ransomware-as-a-Service creation, and the TOR (Onion routing). Ransomware primarily targets Windows systems even other systems, including Android, ios and Linux servers.

Ransomware can be classified into many forms, in which the most virulent and violent are crypto ransomware, which encrypts the target data, and locker ransomware, which locks the target device, deny accessing the user machine. Crypto ransomware is cerber, wannacry, Gandcrab Ransomware shown in Figure.1 [16]. The characteristics of ransomware are as follows ransomware can either act as stand-alone malicious software or combine to act as a larger unit with other malicious software.
Several programs have added new and additional modules to the programs, and they can still perform decently. This is distributed in vast quantities over the internet, and can quickly infect as many hosts as possible. By introducing them the ability to add more complicated attacks beyond the skill level of the attacker it is simple for hackers. It can neither be easily identified nor detached from the system, and bypasses most of the user's security actions. It can also discharge complex authentication methods. Ransomware touches a wide range of devices. [15]. Ransomware is an integral part of cyber-attack. Now-a-days Ransomware can earn a lot of money to the hackers by performing various criminal activities.

![Gandcrab Ransomware](image)

Fig.1 Gandcrab Ransomware

Machine learner played an important role in the growth of intelligent systems for many years. It obtains a labelled dataset and yield a model as output, which can handle new data. Classifiers learn from the profusion of input and labelled output to create a pattern. As such, it is proven that adopting machine learning classifiers enriches accuracy in detection. Successively, the intention in this work is to evaluate classifiers for machine learning to provide an alternative solution for detecting ransomware using network traffic. In this forthcoming analysis, for an additional advantage of this applications we can use this antivirus solutions which will train the resources of a computer to collect network traffic from the device's exterior. Additionally the network traffic identification process is analysed distinctly. Such an approach reduces consumption of resources and the sophistication of applications.

2. Detection of Ransomware Types

There are currently various active methods for detecting and classifying Ransomware. For a broad variety of antivirus software products the most common approaches used are:

- **Signature-based identification** – Unclear code signature is compared with a database of malicious signatures identified.
- **Heuristic detection** - uncertain code functionality is compared with a recognized database of malicious functionality;
- **Machine Learning** - Using supervised or unsupervised algorithms which is capable of recognizing and classifying new specimens of ransomware based on similar features that are connected to the training set.

The following methods are used by professionals in the antivirus, forensics, and cyber security industries to detect, classify and analyse uncertain code:

- **Static analysis** – the suspect code is evaluated using a disassembler in order to recognize the code structure and the code functionalities.
- **Dynamic analysis** – In an ordered environment the suspect code is executed and its behavior is analyzed using various methods. Types of dynamic analysis are code execution in debugger or in sandbox.
Ransomware acts different from various malware mainly due to its destructive nature. A ransomware's main objective is to efficiently execute the payload that will begin data encryption on the infected device. From a slyness perspective some ransomware utilizes illicit techniques to prevent detection before the encryption process is done, but ransomware usually does not use advanced stealth functionalities because the malware is built to have a short life span. Another reason why ransomware does not use advanced slyness mechanisms is that once the destructive actions of ransomware have been completed, the user will be awake when the system has been infected.

3. Related Works
The author presented [16, 17] the idea of ransomware-related cryptovirology to cryptography. Although some researchers catalogue the ransomware as a cryptovirus. Generally, the ransomware is not a cryptovirus. Certainly it is not self-reproduction of first defined in 1986. In reality, a lot of ransomware such as Princess Locker, Striked and Locky do not duplicate their code into other directories. Ransomware proves to be a virus for GandCrab, and therefore a cryptovirus like TeslaCrypt [18]. GandCrab verified the findings that showed Ransomware is dependent on social engineering and demographics such as personal details to compromise their computers. Really, GandCrab (all its versions) is delivered in many ways through social engineering, using spam and exploit kits the last version v4, released in July 2018, was spread via fake crack pages, for example; In 2010 the first ransomware analysis was published detailing a relative study of four families- Filecode, Kroten, Dirt and Gpcode. Same happened in 2017 i.e the Locky ransomware inspired by the previous work on Princess Locker, Spora and TeslaCrypt [8,9] which discussed the common behaviours of 76 samples of 12 families. Our paper is close to these works, the former work offers only an analysis of crypto-Ransomware but some detection and prevention measures are being proposed in our work. The first attempt was in 2015, when they indicated that a real-world protection against crypto-ransomware can be used to track suspicious file system activities. The outcome of this study is called as Reveal dynamic analysis which can be a motive to detect ransomware. In 2016/2017, for example, further work on ransomware detection was released, including Crypto Drop, a tool focused on file system operations, ShieldFS, Redemption and others. Sources of protection include PayBreak [10,11], a method that hooks the encryption functions that performs a standard library and stores CloudRPS for encryption keys, the ransomware protection cloud analytics software, FlashGuard, CLDsafe and others. For example, there are some existing GandCrab analyses in which they present GandCrab from a different perspective to our analysis. It aims at analysing GandCrab, extracting its behaviours which can be used to detect it or other ransomware detection. It suggests that ransomware analyses for GandCrab Ransomware or other type of ransomware should extricate and recognize a familiar ransomware behaviours to create a wide range of familiar behaviours that have used to detection and prevention [12].

4. Proposed method
In the proposed process, the valuation had been initially done by allowing for payload features in the exploit kit. The experiment was carried out by using windows 7 for malware research in Cuckoo Sandbox2 with Virtual Machine (VM). Some main features and compromise indicators are initially developed in physical selection and analysis. It combined with tracewrangle3 and Wireshark to discover compromising indicators in PCAP data. They are generated via the renowned Wireshark network sniffer, honeypots. Figure 2 shows that the workflow of proposed method, the first stage is data collection, which captures the network traffic of normal and ransomware data and transmits it to the next phase. In the second phase, which is feature selection and extraction, the main features are IP address, file length, url. URL are divided into sub features which is known as indicators to collect data for all probable information about the malicious data samples. Network data is extracted from the high-level features. Each URL captures network traffic, and sample data are tested on virtual machine. It cleans all the noisy data from the dataset which will not be a data anymore. It is used to extract the features from traffic where the network features are extracted, labelled and stored in a database to be applied in the next phase. The machine learning classifier enters the final phase, whereby the information in the database trains the machine learning classifier to produce a detection model.
5. Implementation

The proposed work is implemented in a windows environment. Script writing has been written using Python language. It includes steps like data collection, extraction and classification of features using supervised algorithms.

5.1 Data collection

The collection of data samples from the Dionaea Honeypot. It extract the features of the network data from random forests to extract the important data. It reduce the noisy data from the data collection. The extracted features are URL, IP address, File extension and web page.

5.2 Feature extraction

In this stage, verified malicious samples are checked for further analysis and extraction of the function in experimental setup. At this stage features are observed such as modified extensions files, evaluated privilege, http connections, scanning of open ports, CPU usage, process ID and services. Recognition of payloads has performed major task of this stage. The aim of this step is to identify the payload which contains ransomware. URL, Filetype and Ipaddress were used to draw IOCs as major features. In the main feature interaction with URL, domain names, url parameters and email ids, redirection count are collected as URL functions. It conducts a track/ log into database through which Wireshark capture on retrieving the IP victim file shows that Figure.3. Redirections in a website contains iframe vulnerability of JavaScript in webpages. Redistributions on a website may be implemented using the vulnerability of iframe present in JavaScript to a website. We also noticed that the count of redirection plays a vital role in distinguishing malicious data and benign samples, and also recognized that in rerun of our program, the redirection counts are different in intervals of one week.

Fig.2 Proposed method
MIME types are recognized by tracking data which are transported in network activity. By the aid of mime sort to determine the sort of vulnerability that an exploit kits holds, in our methodology extensions of file are very prudently tested to identify the device concession. Ransomware attack had mostly tracked by extensions of encrypted data. In this instance we have started with 620 samples of numerous ransomwares, two Hundred and eighty samples of data were distinct from six hundred samples. Hacker uses Domain Creation Algorithms in which the domain name continues to shift and it is tough to identify only on a domain-based basis. Then address of IP resides the similar for some time, although IP Spoofing is not an importance in cyber criminals then still can help in blocking a threats by maintaining extensive list of ip addresses. We discovered addresses of IP belonging to the hacker server or to the command and control server using the list of ip addresses obtained in each run of a study. These malicious addresses of IP are used in this work as main features in the protection of ransomware. Investigational tests situation i.e. payload discovery, it is observed that most of the payloads released by exploit kits are ransomware data possessed from September 2016–January 2017. An URL collected from different provenance, are www.malware-traffic analysis.net, total6 virus, don'tneedcoffee8, and twitter7. Any of the samples we checked in the cuckoo sandbox. Of those 10000 samples, 6000 ransomware were released, 1800 samples of data become a drop of Trojan. Although bots gone to 1100 payloads, 200 samples of data fallen spyware whereas the left over 900 are other malwares. These sub-functions are usually called compromise indicators.

5.3 Classification and detection
In this step, classification of the ransomware which classifies the data are ransomware, benign and malware. It contains 40,000 of the ransomware sample and 25,000 samples of benign file shows in figure.4. It can be classified as ransomware and benign by using machine learning algorithm are random forest, gradient tree boosting algorithm and Support vector Machine algorithm. Table.1 shows that the accuracy of the algorithm. It compares the result by using linear regression, Navies Bayes, Adaboost algorithm. In combination of algorithm combined better accuracy. It measures a true positive rate (TPR), true negative rate(TNR), false positive rate(FPR), false negative rate(FNR).

| Algorithm    | TP rate | FP rate | %    |
|--------------|---------|---------|------|
| SVM,RF,GBTA  | 0.987   | 0.01    | 98.45|
| LR           | 0.567   | 0.03    | 58.43|
| NB           | 0.837   | 0.012   | 83.7 |
| Adaboost     | 0.765   | 0.03    | 76.8 |

Table.1 Accuracy
6. Gandcrab ransomware

Gandcrab ransomware have rapidly spreading Ransomware-as-a-Service (RaaS) attack throughout the world meanwhile January, 2018. The online portal GandCrab RaaS was ultimately shut down in June 2019. Over the course of these 15 months, Hackers frequently restructured its code and retailed the malicious code, simplifying hackers to write their own ransomware without understanding. Hackers then distribute ransomware to GandCrab via negotiated websites that are built with WordPress. It started being distributed on underground hacker forums such as Exploit.in via Ransomware-as-an-Affiliate system. GandCrab's novel versions used to Salsa20 stream cipher for encryption of user offline data, with RSA-2048 encryption algorithm that connects to the C2 server. It scans drives in logical from A to Z data then encrypted file by adding with random key and an initialization vector of 8 bytes to the file's contents. The private key uses Salsa20 key to encrypt in the registry, and an RSA-2048 encrypted with a public key enclosed in the malwares. GandCrab is a very powerful Ransomware, and only GandCrab creators can decrypt encrypted files (shows in Figure.6). Gandcrab ransomware includes different version types, such as GandCrab version 1, GandCrab version 2, GandCrab version 3, GandCrab version 4, and GandCrab version 5.
6.1 Gandcrab v1
It is one of the 1st versions of the ransomware discovered in January 2018. It extends email attachments through infected spam and phishing just like its predecessor, and switches via payload once the attachment has been opened. The most widely used files, including .doc,.docx,.txt,.jpg,.gif,.png,.mp3,.mp4,.ppt, were targeted data with the extension of .gdcb files was added into file. These generates a GDCB-DECRYPT.txt (shows that Figure.5) data on the target screen, which involves installing TOR client or routing onions and transmitting 1.54 DASH ransoms. The hacker’s sum grows if the victim is delay in fee for it.

6.2 Gandcrab v2
Gandcrab ransomware2 is the 2nd form of a Ransomware scandalous. With the support of AES-256 in CBC mode with RSA-2048 algorithms the files had encrypted and cannot be retrieved anymore. It was distributed to RIG exploit kit through Seamless malvertising campaign of prominent victims. Unlike its predecessor, however, the HoeflerText Font Update scam may also affect its victims.

6.3 GandCrab v3
GandCrab ransomware uses the .crab extension for the 3rd cyber-attack, then CRAB-DECRYPT.txt ransom note to scare the victims. The main difference for the victims who got GandCrab 3 encrypted is that DASH encrypts the currency and pays the ransom in Bit coins.

6.4 Gandcrab v4
In early July 2018 GandCrab ransomware 4 was observed [19]. A salsa20 encryption algorithm and RSA-2048 is used. The file extension was altered to .KRAB, and one of two different ransom notes given victims files are: CRAB-DECRYPT.txt or KRAB-DECRYPT.txt.It is ransomware which is the most powerful.

![Fig.6 Gandcrab Ransomware file](image)

6.5 Gandcrab v5
GandCrab ransomware group willingly authorized working with the Fallout exploit kit. A crypter service offers malware chaos to evade detection by anti-malware products. However, they decided to unlock all victims in that locale. It is one of the harmful ransomware. It used salsa20 and RSA-2048 cryptographic algorithm.
GandCrab hackers sent an affront to the computer and noted as one of the command & control servers and other organizations identified for ransomware tracking in their executable. To retrieve an IP address, the commands,lookup [insert domain] a.dnspod.com It supported the.bit TLD, with bleepingcomputer.bit, nomoreransom.bit, emisoft.bit, esetnod32.bit, and gandcrab.bit being one of the domains. If the infected machine cannot connect to the server then the computer will not be able to be encrypted by the ransomware. Continuously running in the background process which attempts to get the server IP address and then connect to it. Although it is not known in which time the data is being sent and retrieved, the server is most likely to send the public key to encrypt the files. Gandcrab ransomware will also link in this process to http://ipv4bot.whatismyipaddress.com/ to decide the victim's public IP address. Before GandCrab as showed in Figure.7 encrypts files of the victim, it must first search and terminate those processes. This will process so that any file handles opened by these encrypted ones can be properly closed. Figure.8 shows that the list of processes that are terminated in Gandcrab Ransomware.

**Fig.7 Gandcrab Ransomware**

**Fig.8 Terminated Process**

### 7. Conclusion

In this paper, Detection and prevention of ransomware detects the ransomware by using dynamic analyses of the ransomware with machine learning algorithm. The proposed process of detection of ransomware classifies ransomware from benign files. It presents detection and classification of ransomware by using machine learning algorithm of Random forest, Gradient Tree Boosting Algorithm and support vector which obtained a better accuracy of 98.45% among an adaboost, Navie Bayes and Linear Regression. Gandcrab ransomware version and algorithm used to encrypt their files. In future work, it will automatically classify the ransomware, malware and benign file.
Reference

[1] Abrams, L.: GandCrab ransomware distributed by exploit kits, appends GDCB extension. https://www.bleepingcomputer.com/news/security/gandcrabransomware-distributed-by-exploit-kits-appends-gdcb-extension

[2] Abrams, L.: GandCrab ransomware version 2 released with new .Crab extension and other changes. https://www.bleepingcomputer.com/news/security/gandcrabransomware-version-2-released-with-new-crab-extension-and-other-changes/

[3] Abrams, L.: GandCrab version 3 released with autorun feature and desktop background. https://www.bleepingcomputer.com/news/security/gandcrab-version-3released-with-autorun-feature-and-desktop-background/

[4] Bajpai, Pranshu, Aditya K. Sood, and Richard Enbody. "A key-management-based taxonomy for ransomware." 2018 APWG Symposium on Electronic Crime Research (eCrime). IEEE, 2018.

[5] Biasini, N.: Gandcrab ransomware walks its way onto compromised sites. https://blog.talosintelligence.com/2018/05/gandcrab-compromised-sites.html/

[6] Brewer, Ross. "Ransomware attacks: detection, prevention and cure." Network Security 2016.9 (2016): 5-9.

[7] Cabaj, Krzysztof, Marcin Gregorczyk, and Wojciech Mazurczyk. "Software-defined networking based crypto ransomware detection using HTTP traffic characteristics." Computers & Electrical Engineering 66 (2018): 353-368.

[8] Caivano, Danilo, et al. "Ransomware at x-rays." 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2017.

[9] Checkpoint: The GandCrab ransomware mindset. https://research.checkpoint.com/gandcrab-ransomware-mindset/

[10] Cimpanu, C.: Free decrypter available for GandCrab ransomware victims. https://www.bleepingcomputer.com/news/security/free-decrypter-available-for-gandcrab-ransomware-victims/

[11] Continella, Andrea, et al. "ShieldFS: a self-healing, ransomware-aware filesystem." Proceedings of the 32nd Annual Conference on Computer Security Applications. 2016.

[12] Gallegos-Segovia, Pablo L., et al. "Social engineering as an attack vector for ransomware." 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON). IEEE, 2017.

[13] Gazet, Alexandre. "Comparative analysis of various ransomware virii." Journal in computer virology 6.1 (2010): 77-90.

[14] Gómez-Hernández, José Antonio, L. Álvarez-González, and Pedro García-Teodoro. "R-Locker: Thwarting ransomware action through a honeyfile-based approach." Computers & Security 73 (2018): 389-398.

[15] Young, Adam L., and Moti Yung. "Cryptovirology: The birth, neglect, and explosion of ransomware." Communications of the ACM 60.7 (2017): 24-26.

[16] Yun, Joobeom, et al. "CLDSafe: an efficient file backup system in cloud storage against ransomware." IEICE TRANSACTIONS on Information and Systems 100.9 (2017): 2228-2231.

[17] Nadir, Ibrahim, and Taimur Bakhshi. "Contemporary cybercrime: A taxonomy of ransomware threats & mitigation techniques." 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). IEEE, 2018.

[18] MalwarebytesLabs: GandCrab V4 released with the new .KRAB extension for encrypted files. https://www.bleepingcomputer.com/news/security/ganderab-v4released-with-the-new-krab-extension-for-encrypted-files

[19] Bayer, Ulrich, Christopher Kruegel, and Engin Kirda. TTAnalyze: A tool for analyzing malware. na.,2006.