Journal of Information Security and Applications 40 (2018) 44–51

Contents lists available at ScienceDirect

Journal of Information Security and Applications

journal homepage: www.elsevier.com/locate/jisa

Ransomware behavioural analysis on windows platforms

Nikolai Hampton[a], Zubair Baig[b,*], Sherali Zeadally[c]

[a] Impression Research Pty Ltd, Australia
[b] Edith Cowan University, Australia
[c] University of Kentucky, USA

ARTICLE INFO

Article history:

Keywords:
Cryptovirology
Cybersecurity
Intrusion detection
Malware
Ransomware
Win/32

ABSTRACT

Ransomware infections have grown exponentially during the recent past to cause major disruption in operations across a range of industries including the government. Through this research, we present an analysis of 14 strains of ransomware that infect Windows platforms, and we do a comparison of Windows Application Programming Interface (API) calls made through ransomware processes with baselines of normal operating system behaviour. The study identifies and reports salient features of ransomware as referred through the frequencies of API calls.

1. Introduction

Malware or malicious software is defined as any program or process that is crafted by the adversary to either affect routine operations of a computer, its operating system and hosted software, or to steal sensitive data. When such malware is crafted with the intent of extorting user data and holding it for ransom, then it is categorized as ransomware. While malware has been persistent for decades, the emergence of ransomware as the next big threat adopts a new business model by threat actors. The evolution of malware capabilities over the past 30 years is attributed to the rapid advances in computing power, memory, and communication bandwidth. Extortion of user data through malware dates back to 1989, when the PC CYBORG (AIDS) Trojan was released on floppy disks. Infected floppy disks when inserted by naïve users into their workstations would cause a Trojan infection, locking user files using basic cryptographic techniques, presenting a message stating the user's 'breach of software license', and demanding an amount of approximately US $200 for release of the extorted data. The Trojan was not very successful because the payment procedure adopted by the adversary was through bank cheques and the proliferation of malware through the crude floppy-disk medium was excruciatingly slow.

Strong data encryption techniques, attributed to advances in computing power and memory technology/affordability, alongside advances in payment techniques and cryptocurrency [1] have led to rapid evolution of ransomware during the period 2007–2016. The ability of the adversary to conceal his/her identity and reaps profit through ransomware infections proliferating across billions of Internet-connected devices, is thus easily achievable in today's highly connected landscape. CTB-Locker (Curve, TOR, Bitcoin), is considered to be the first variant of ransomware to effectively combine three key characteristics required to achieve a high degree of success in infection, namely, the *anonymity capabilities* of the TOR routing protocol to conceal adversary location, the *anonymous payment capabilities* of Bitcoin to keep payment path untraceable, and *strong encryption* based on Elliptic Curve Cryptography with sufficient key lengths to resist attempts to crack the key including those involving brute-force [2].

In 2013, a 500% growth in ransomware variants and capabilities was reported [2]. This can be attributed to the three technological advances enumerated above. The common families of ransomware alongside their respective dates of emergence are listed as follows [2]: PC CYBORG Trojan (12/19/1989), One Half Virus (>1994), GPCode family (~2004), Reveton (~2012), CryptoLocker (~2013), CryptoWall (~2014), CryptoDefense (~2014), PoshCoder (~2014), Virlock (~2014), TeslaCrypt (~2015), CryptoFortress (~2015), CryptoTorLocker2015 (~2015), CTB-Locker (~2015), CryptoWall (~2016), Xorist (~2016), Filecoder (~2017) along with variants such as Petya (~2017), JAFF (~2017), and Wannacrypt (~2017).

Ransomware evolution witnessed the first brief increase in 2006–07 [3], mainly through the emergence of the GPCode variants. The GPCode.ak variant in particular was known to write the encrypted file contents to a new location in the user's disk, deleting the unencrypted user files. Through application of the '*undeletion utility*', partial recovery of user data was possible without having to pay the ransom to the adversary. Newer variants of GPCode used stronger encryption techniques with longer encryption keys

[*] Corresponding author.
E-mail addresses: nikolai@pirionsystems.com.au (N. Hampton), z.baig@ecu.edu.au (Z. Baig), szeadally@uky.edu (S. Zeadally).

https://doi.org/10.1016/j.jisa.2018.02.008
2214-2126/© 2018 Elsevier Ltd. All rights reserved.

Table 1
Virtual machine victim user's file structure.

Location	File count and size
Desktop	1.07GB, 442 files, 90 folders
Documents	524 MB, 66 files, 22 folders
Pictures	417 MB, 1344 files, 9 folders
Videos	661 MB, 16 Files, 0 Folders

(1024 or 2048 bits), thus encumbering the user data recovery attempts at the victim's machine.

A close look at the evolution of several versions of ransomware releases revealed that they were mostly copy-paste code from previous versions. Therefore, many of the limitations of one version were carried over to the next. In addition, several ransomware variants operated in unconventional ways. For instance, the Reveton ransomware [2], released in 2015, was found to merely lock the operation system's boot process without encrypting user data. Consequently, the ransomware activity was limited to disruption of operations and recovering user data without having to pay the ransom amount, was found to be easily achievable.

Another observed characteristic of recent ransomware traits is the ransomware procedural requirement to contact a centralized Command-and-Control (C2C) Server, once the victims' machine is infected, prior to encrypting the data. The C2C Server typically holds the cryptographic key required to decrypt the victim's data which has been held for ransom. In summary, the four stages of a ransomware-based attack can be described as follows:

- *Infection*: The ransomware software infects a victim's machine when the naïve victim opens an attachment that accompanies a spam message. Alternately, the victim's machine can also be infected when a compromised website is accessed.
- *Data encryption*: Once the victim's machine is infected with ransomware, cryptographic keys utilizing the Public Key Infrastructure (PKI) are generated either on the infected PC or the C2C server. The ransomware then proceeds to lock down the user's files or device. Ransomware specific definitions commonly result in one of two actions being undertaken: either the data/files on the victim's machine are attacked on a file-by-file basis, or critical file system structures such as the Windows Master File Table are altered. In both cases, the original files or data are encrypted with the host specific cryptographic keys, and the original files or metadata are then deleted.
- *Demand*: The ransomware software displays a message to the victim demanding that a certain amount be paid so as to release the locked data/files.
- *Outcome*: Based on the action taken by the victim, the following are possible outcomes: a) the data is recovered through elimination of ransomware trait from the victim's machine without paying the ransom amount, b) payments are made through anonymous channels such as BitCoin/MoneyPak or DarkCoin, or c) payments are not made and the ransomware trait is not eliminated, upon which the data/files are destroyed; with no backup in place, permanent loss of victim's data/files thus occurs.

It can be seen from the above examples that ransomware activity must by nature follow specific patterns of behaviour. These patterns include the file identification process, encryption of files, network command and control communications, and use of anonymous networks. Quite simply, there is no optimal way to scan files and encrypt their content without making system level calls facilitated through the Windows Application Programming Interface (API). The Window API [4] provides a set of programming interfaces that simplify the process of developing software. For example, while a developer makes the system call "FileOpen", the operating system executes a series of instructions to locate the file in the file system, checks file access rights and permissions, and locates the file on the hard disk before returning the handle or reference back to the developer. By using the Windows API, developers are free to focus on the logic of their program (or malware) code and use the pre-defined procedures to accomplish their tasks [5].

Windows API sequence of calls has been an area of research during the recent past. In [6], the authors have presented a ransomware detection scheme that operates on Windows platforms and identifies modifications to various application types. Thirty most common Windows applications were evaluated and attempts by ransomware to access these file types, were analysed and reported.

In [7], the authors present a call tracer approach for identifying the sequence of Windows API sequence of calls, by comparing the patterns of calls with known databases of malware, and by applying machine learning techniques for data analysis. Malware samples obtained from popular repositories were analysed and the results of the machine learning based classification of these samples were reported. In [8], the authors proposed an approach for identifying API sequence calls for malware samples. The lack of accuracy in anti-virus tools was highlighted as one of the motivations for the research conducted. Malware behaviour was generalised across 23,080 popular samples of malware.

As the number of Windows API calls is limited, and generally lower level file, network and cryptographic operations are exposed through a limited set of instructions, it may be possible to detect ransomware specific activities by analyzing their usage (or calls) to certain Windows API functions. We analyse and report ransomware activity based on the executing payload that has been transferred to a victim's machine beforehand. API call patterns and frequency analysis are used to help determine the behaviour of ransomware in a real-world environment. By identifying the programming patterns used by ransomware programmers, we can improve operating system or Kernel level protection mechanisms. TBased on the results reported in this paper, we provide a fundamental platform for researchers to examine methods of ransomware detection based on behavioural analysis and/or entropy-based analysis, for future research.

Table 2
Shared folders (network) file counts.

File type	Count of files in shared (network) folder
jpg and png image files	1337
ppt (and pptx)	2
pdf	55
doc (and docx)	34
xls (and xlsx)	17
mp3/mp4 (audio and video media)	20
other filetypes	27
directory and subdirectory entries (maxdepth = 5)	31