



ASSIGNMENT COVER SHEET

Student Names & ID:	1. Muhamad Afiq Busari 2000476 2. Jeremy Jevon Chow Zi You 2001558 3. Tse Kin Ping, Matthew 2001568 4. Lee Zhi Yang Aloysius 2001348
Module Code:	ICT2202 Digital Forensics
Tutor's Name:	Goh WeiHan
Assignment Title:	Assignment One - Developing a Solution for a Problem in Digital Forensics
Due Date:	7 November 2021
Date Submitted:	7 November 2021

ASSIGNMENT COVER SHEET

Please complete this declaration and submit with your assignment.





This assignment is entirely our own work and represents our learning in the module.

We abide to the Academic Integrity Policy and are aware of the disciplinary actions that can be taken for plagiarism. Any information sourced from elsewhere has been appropriately acknowledged and referenced.

We have and will continue to maintain the privacy of any person we have referred to in this assignment.

We acknowledge a copy of our work may be used for moderation purposes.

We have kept a copy of the assignment for our records.

S/N	Name	Student ID	Signature	Date
1	Muhamad Afiq Busari	2000476		07/11/2021
2	Jeremy Jevon Chow Zi You	2001558		07/11/2021
3	Tse Kin Ping	2001568		07/11/2021
4	Lee Zhi Yang Aloysius	2001348		07/11/2021

Arcana

Muhamad Afiq Busari
School of Information and Communications Technology
Singapore Institute of Technology
Singapore, Singapore
2000476@sit.singaporetech.edu.sg

Tse Kin Ping
School of Information and Communications Technology
Singapore Institute of Technology
Singapore, Singapore
2001568@sit.singaporetech.edu.sg

Jeremy Jevon Chow Zi You
School of Information and Communications Technology
Singapore Institute of Technology
Singapore, Singapore
2001558@sit.singaporetech.edu.sg

Lee Zhi Yang Aloysius
School of Information and Communications Technology
Singapore Institute of Technology
Singapore, Singapore
2001348@sit.singaporetech.edu.sg

Abstract—The usage of digital forensics has allowed security investigators to study the evidence obtained from a multitude of digital devices to analyze the actions of certain persons. There are lots of forensics tools that are readily available on the market, many of which are open source. This report will give a detailed explanation of Arcana, the product that the team has come up with, regarding its features as well as the research done to develop this product.

I. INTRODUCTION

We are currently living in this age where digital information is easily created, sent, stored, and processed, by billions across the globe. Both individuals and corporations are using digital devices to bring on huge technical and economic advantages. However, when there are pros, there will be cons. The digital era has brought on a new threat that challenges its users, specifically malware. Since the dawn of computers, the usage of malware has been spread across devices, with intentions varying from pranks to criminal activities. Malware which comes in various forms like viruses, worms and trojan horses, have threatened the security and backbone of digital systems. One of the strategies that can be used to combat this growing threat is via the aid of digital forensic tools. Digital forensics tools can allow the investigator to study how the malware might have propagated, in addition to the impact caused by the malware.

II. BACKGROUND RESEARCH

Malware or malicious software can be broadly categorized into various types such as viruses, ransomware, Trojans, worms, and many more. Malware can even overlap multiple categories, depending on its features. The usage of malware by attackers can bring upon financial gain, disruption, or even destruction of the systems that are impacted. For digital forensic investigators, it is essential that they know and understand the motives behind malware attacks so that they can pry deeper into the usage of that malware. The investigator can then investigate the potential damages that have been caused by the malware and provide a remedy to the issue.

A. Ransomware

In recent years, ransomware attacks specifically, have been increasing since 2018, comprising 68.5 percent of malware attacks in 2021 [1]. These attacks have brought organizations to their knees such as in the Colonial Pipeline attack, where it led to major shortages of fuel across the United States [2]. Ransomware is a type of malware that aims to obtain financial incentives from the victim. It does so by infecting data and files stored in the victim's system and encrypting them. Files can be encrypted using either asymmetric, symmetric encryption algorithms, or both. The files that are encrypted with symmetric encryption can be done so via strong algorithms such as Advanced Encryption Standard (AES)-256. AES-256 can encrypt files at a high speed and has 256 bits as its key length which is approved by government-related entities like the National Security Agency (NSA) [3], even for top-secret information, as a cryptographic module. It is computationally infeasible to break the AES-256 encryption with modern personal computers because it would take millions or even billions of years to do so with the current computing power.

In hybrid encryption, RSA is commonly used to generate a public and private key pair due to its strength. The AES keys that were used in the encryption, will go through another round of encryption with the public key. As a result, the public key can only be decrypted with the private key that is held by the cyber-criminal. This is where the ransom fee will come in. There is no need to have internet connectivity for the encryption process and is only required during the decryption, for the hybrid encryption.

The victims of ransomware attacks are often faced with either of the two choices, choose to pay the ransom, or not to do so. In the case that the victim decides to pay the ransom, the price can range from hundreds to millions of dollars. Payment for the ransom is often demanded in the form of cryptocurrencies, for example, Bitcoin [4], which is difficult to trace by the authorities. A blog post by Palo Alto Networks, a cybersecurity company, reported that the average ransomware payment has risen to USD \$570,000 in the first half of 2021 [5]. This is a worrying trend that could influence even more cybercriminals to take part in this kind of ransomware attacks. Even though the victim has paid up the

ransom, there is no guarantee that the files will be decrypted. Therefore, the consensus by most security professionals is not to pay the ransom, to prevent the attackers from gaining any form of monetary benefits, as well as to deter further attacks.

With proper backups, the victims of ransomware attacks have a second lifeline and can decide not to pay the ransom. Most victims will be compelled to pay the ransom when they face the imminent threat that their data can be lost permanently, which can lead to massive financial losses. The victims can thus mitigate this risk by having backups of their systems with the proper procedures taken. These system backups will need to be stored offsite in offshore servers or even private clouds. It is still possible for ransomware to infect backups stored on the same network and render them useless. The victims of ransomware attacks with working backups can hence restore their systems to a previous state and reduce the disruption caused to the systems, without having to pay a hefty ransom to cybercriminals.

WannaCry is a popular ransomware attack that occurred in 2017. More than 200,000 computers across 150 countries fell victim to the attack, which demanded USD \$300 for the decryption of the files [6]. WannaCry took advantage of the CVE-2017-0144 vulnerability found on Microsoft Windows operating systems. WannaCry made use of the EternalBlue exploit to attack the vulnerable implementation of the Server Message Block (SMB) protocol, across the Windows operating system (OS). Although the patch has been released prior to the exploit, however, many of these devices are not patched and were using older versions of Windows systems. Thus, it is essential that digital devices are regularly patched and upgraded to remove such vulnerabilities.

B. Trojan Horse

Trojan horses are broadly categorised as a form of malware that hides its true intentions from users. They are commonly spread via social engineering attacks where the victim unknowingly executes files that do not appear malicious. This can be done by clicking on a fake email attachment from an unknown user. The Trojan can then install backdoors to the infected system and allow attackers access to the host. The attack can then monitor the actions of the victim through the backdoor, without the knowledge of the victim and exfiltrate information from the host.

Furthermore, the system can be turned into a zombie that is part of a botnet. A command & control (C&C) server can then remotely use the infected system to launch distributed denial-of-service (DDoS) attacks on its targets.

Banking Trojan is another widespread form of Trojan that is utilized by cybercriminals. With the rising usage of online banking, attackers can obtain access to the credentials of their victims, in the digital form. Once the victim is infected with the Trojan, the attacker can then easily gain the banking credentials of the victim.

For example, ZeuS is a Trojan that runs on the Windows OS. It carries out the attack through man-in-the-browser (MiTB) keystroke logging as well as form grabbing. A host that is infected with ZeuS will detect when the user visits a banking website. It then records the keystrokes of the victim

in which the attacker can use to log in to the account. Following that, it can take screenshots of the browser when it records a mouse click and can even insert new elements into the web forms to ask for confidential information like bank Personal Identification Numbers (PIN).

C. Computer Worm

A computer worm is a self-replicating malware that spreads itself to other hosts. Once the host is infected, it will scan the network, then attempt to spread itself to systems containing a certain vulnerability. This process will not require any form of user interaction and can spread on its own, causing devastating results to the network. Worms are typically more infectious than viruses because viruses can only propagate with the help of a user. Worms can not only infect local computers but can also affect all servers and clients on the network.

The Stuxnet worm is a malware that was specifically designed to attack nuclear plants. Industrial control systems (ICS) are of massive importance to both corporations and governments. Nuclear plants leverage on such technologies to power and operate their plants. Stuxnet was created to disrupt a Windows-based application that is run on a particular ICS, which was produced by the company Siemens [7]. It was initially spread on an air-gapped network, without an Internet connection, via a thumb drive [8]. It is suspected that the worm was created to target Iranian nuclear plants as most infected systems were found to be in Iran. It manipulated the valves on centrifuges by increasing the pressure and damaging the devices in the process. Stuxnet had been stealthily sabotaging centrifuges found at the Natanz plant in Iran for a year before it was discovered.

D. Windows Event Logs

When a machine has been infected by malware, looking at the Windows event logs might provide a crucial insight as to the actions that were taken by the malware. Windows event logs contain the description of the essential events that have occurred on the machine such as application crashes, failed login attempts and Windows Defender activities. Logs will offer valuable information on the system. It not only provides data on something that is broken on the machine but can also reveal weaknesses that could potentially affect compliance and corporate governance. Log files are essential to a forensic investigation as they can connect a certain point in time to a particular event.

Windows event logs are essential in the post-mortem analysis of the host after it has been infected by malware. They are widely used to identify and detect violations within corporate systems. Event logs are one of the most important features in an OS because they will display the historical and current states of the machine.

An experienced attacker will aim to target the event logs, remove traces of compromise, evade detection, as well as to keep the attack secret. This is to prevent the user from detecting that an attack has happened and have the loophole patched. Hence, in many organizations today, the events log of devices will be piped and stored in a separate location,

commonly in a log server. Even if one of the hosts is infected and has its event logs altered or destroyed, its activities will still be retained on the log server. Proper log management can help to retain confidentiality, integrity, and availability of logs.

E. Malware Detection Methods

There are three main different methods of malware detection that will be covered in this section which are signature-based, behavioural-based and heuristic-based.

- **Signature-Based**

Signature-based malware detection recognises common malware that has been documented in the past by using patterns that have been extracted from the malware as a form of identification. This method is used by commercial antiviruses as it is more efficient and faster than the other methods and has a low rate of error. It is, however, unable to identify new or unknown malware [9].

- **Behavioural Based**

Behavioural-based malware detection tracks a program by its behaviour and observes what they do to identify if the said program contains any malicious malware. Programs with similar behaviour will be identified during this process. It is also able to keep track of the possible sub-processes of the malware as it will use system resources and services in the same way [9].

This behavioural-based detection can identify malware unknown malware variants where signature-based detection would fail without having the relevant documentation. However, it is not entirely accurate as it will produce a high number of false positives and take a long processing time.

- **Heuristic-Based**

Heuristic-based malware detection uses data mining and machine learning techniques to identify possible malware behaviour from a program [9]. It is designed to spot malicious characteristics in new or modified versions of a malware with the help of machine learning. Heuristic-based detection could potentially outperform the signature and behavioural-based detection however the machine learning aspect behind the technology requires the use of in-depth classification tools.

III. CHALLENGES FACED

A. Malware Scanning

For the malware scanning to occur, the team had to first find a database of malware signatures that could be used to compare the files and Uniform Resource Locator (URLs) with. One of the most common strategies is to detect malware in a file via signature-based methods. The file will go through a hashing program to produce a unique fingerprint, that can be used to identify malware. Hashing algorithms such as Message-Digest algorithm 5 (MD5) and Secure Hash Algorithm (SHA-1) can be used to hash the file. Different

security vendors will have their own databases of file signatures that can be used to compare the hashes with, to detect malware.

In addition, malicious URLs can be detected with certain tools as well. Phishing links are popular attacks used on users of the Internet as malware can be easily downloaded by an unsuspecting user of the site. The site can have extremely close domain names to a legitimate website to fool users into believing that it is authentic and can be trusted. Some phishing sites will resemble a banking website for example, where the user will be tricked into entering their banking information which can be stolen by cybercriminals. In other cases, the website might contain nasty viruses that could potentially infect the user. Therefore, it would be crucial for the users to discover potentially malicious websites that they should avoid.

There are many different websites on which users can choose to upload their files or enter a URL to test for malware. Examples of these websites include VirusTotal [10], Kaspersky Threat Intelligence Portal [11], and FortiGuard Labs [12]. These online malware scanner websites are from reputable security companies and are free to use.

Therefore, the team has decided to leverage on the VirusTotal API for the scanning of malware. VirusTotal API is free to use, has lots of different functionalities and the usage of the API is well documented, aiding in the implementation of the project. Depending on the API will also remove the need to maintain a new database of malware signatures which will be almost impossible to do, for a small team and the limitation of time.

B. False Positives

Arcana purely relies on the VirusTotal to scan the files as well as URLs that are uploaded. As such, there is a huge dependency on the results that are generated to be accurate. Even if one of the security vendors flag that the artefact is malicious on VirusTotal, it is not possible to exactly know why it is malicious. The VirusTotal database also leverages on the results from the security vendors that they are partnered with. The methodology to derive the probability of malware in the uploads is not publicly available from the security vendors.

Therefore, to reduce the probability of generating a false negative in the scan results, the team has decided to brand a file or URL, that is flagged by even one of the security vendors as malicious. The user of Arcana can then filter out all the malicious objects that were detected and do a more thorough investigation.

C. Raw Image Processing

One of the functions of Arcana is the ability to allow users to process raw images and identify any malware-infected files on the image. Users will then be able to export the infected file for further analysis.

Initially, the team faced some difficulty in understanding how the E01 file system works. There was a lack of available

documentation for the team to refer to and we did not know where to start. The team had to conduct research on various python libraries that we could use to parse the raw image such as *pytsk3* and *pyewf*.

It takes some time to process the raw image each time we run Arcana and we decided to process and store the file and browser information into separate csv files with the use of the *csv* and *sqlite3* python libraries.

D. File Directory Listing

After processing the raw image with Arcana, the next step was for us to display the information of all existing files in the image. This would provide the forensics investigator with a user-friendly interface to identify and locate the malicious file. Useful information such as the file name, file extension, file path, creation date, modification date and change date will also be provided. However, when we first displayed the file listing, Arcana output the entire list and it was difficult for us to scroll through the list just to locate a single file. The team then came up with the idea of including a paging function as well as a keyword search function that would make it easier for the user to locate specific files. We were able to achieve our listing function with the use of the *os*, *datetime* and *hashsum* python libraries.

IV. TECHNICAL SOLUTION

Arcana aims to act as a digital forensics tool suite where it can perform its various functions on a single program. It leverages on the VirusTotal API to do the scanning of malware on files as well as URLs. The functionalities that Arcana provides include:

- Raw image data pre-processing
- Displaying of files in a raw image
- Keyword search of files in a raw image
- Selecting of files to extract from raw image
- Processing and displaying web history
- Full virus scan of a raw image
- Scanning files by keyword in a raw image
- Scanning URLs by keyword in a raw image
- Scanning of a single uploaded file
- Scanning of a single uploaded URL
- Live event log analysis of user's device
- Displaying of Arcana's user manual
- Displaying of Arcana's GitHub

A. Raw Image Data Pre-processing

The user can upload a raw image in the e01 format to be processed by Arcana. If the user has generated a raw image in another file format, the user will need to rename the extension to the e01 format. This feature is in development to support other raw image formats.

As Windows is currently the most popular OS in the world, most malware attacks will be written to infect the Windows OS. The file structure will vary for different operating systems. Hence, Arcana currently only supports the processing of a Windows raw image.

B. Displaying of Files in a Raw Image

The function will display all existing files in the raw image after the evidence image has been processed. It will contain the file information such as Partition the file is located, File Name, File Extension, File Type, Created Date, Modify Date, Change Date, Size of the File, File Path, and hash value of the file.

	Partition	File	File Ext	File Type	Create Date
0	PARTITION 2	\$AttrDef	NaN	FILE	2008-10-20 14:26:07
1	PARTITION 2	\$Bitmap	NaN	FILE	2008-10-20 14:26:07
2	PARTITION 2	\$Boot	NaN	FILE	2008-10-20 14:26:07
3	PARTITION 2	\$Extend	NaN	DIR	2008-10-20 14:26:07
4	PARTITION 2	\$LogFile	NaN	FILE	2008-10-20 14:26:07
5	PARTITION 2	\$MFT	NaN	FILE	2008-10-20 14:26:07

Figure 1. Displaying all files in the raw image

C. Keyword Search of Files in a Raw Image

This function allows users to locate specific files in the raw image by searching for them using keywords. The function will display all files from the raw image that contains the keyword. For example, if the term "google chrome" is used as the keyword, Arcana will return the results of all files containing those words. This helps to minimize the time needed to find a specific file the forensic investigator wishes to find.

	Partition	File	File Ext	File Type
62	PARTITION 2	Google Chrome.lnk	lnk	FILE
150	PARTITION 2	Google Chrome.lnk	lnk	FILE
1401	PARTITION 2	Google Chrome	NaN	DIR

Figure 2. Searching for files with the term "Google Chrome"

D. Selecting of Files to Extract from Raw Image

This function allows the user to extract a particular file of interest that is found in the raw image to their local directory. The user will first have to locate the file in the image by using the above display or search options and identifying the file path. Once the user enters the file path containing the desired file, the file will then be extracted to the "extracted" folder in Arcana's directory on the user's local drive. The forensic investigator will be able to conduct further analysis on the malware sample that was found after extraction.

```

Choose an option: 3
Enter object file path: /Windows/regopt.log
[+] File ./extracted/regopt.log Extracted Successfully.

```

Figure 3. Exporting the "/Windows/regopt.log" file from the raw image

E. Processing and Displaying Web History

Arcana can process and list out the web history of a raw image to a csv file and reads from it to be displayed in the program. This allows the user to identify any suspicious web activity.

F. Full Virus Scan of a Raw Image

Arcana can conduct a full virus scan of the artefacts that are found in the raw image after it has been processed. This allows the user to identify the malicious files which are deemed to be a virus. However, due to the VirusTotal maximum capacity of 4 files a minute, the program only displays one file being scanned. If the user were to be a premium member of VirusTotal, Arcana could conduct a full virus scan of all the files in the raw image.

G. Scanning Files by Keyword in a Raw Image

Arcana can scan a selected file via VirusTotal. This provides a quick scan instead of conducting a full scan, which will take a lot longer.

```
Choose an option: 2
Please enter the file path, or part thereof: Attr
Scanning Files (Attr):
Scanning Files..

Files:
  Index  File Path Verdict
0       0  /$AttrDef  Safe
```

Figure 4. Performing VirusTotal scan on the file in the raw image

H. Scanning URL by Keywords in a Raw Image

Arcana also provides a quick URL scan of keywords in the raw image. Like the scanning of files, it provides a quick scan of URL using VirusTotal so that the user will be able to identify if the selected URL is malicious.

```
Choose an option: 1
Please enter the URL, or path thereof: http://picasa.google.com/
Scanning Websites (http://picasa.google.com/):
Scanning Websites..

URLs:
  Index  URL Verdict
0       0  http://picasa.google.com/  Safe
```

Figure 5. Performing VirusTotal scan on a URL in the raw image

I. Scanning of a Single Uploaded File

Users can select a file from the user's device to upload to scan. The function will read the user's file directory to ensure that the file exists. If the file does not exist on the user's device, the program will fail to run and will prompt the user to try again. If the file can be read, it will first hash the file using SHA-256, then upload it to VirusTotal. Arcana will fetch the results and output them to the user stating whether the file is malicious or not.

```
Choose an option: 1
Input directory of file you would like to scan: C:\Users\jerem\Desktop\Document.docx
| File | Verdict |
|-----|-----|
| C:\Users\jerem\Desktop\Document.docx | Safe |
```

Figure 6. Performing VirusTotal scan in the user's local directory

J. Scanning of a Single Uploaded URL

The user can choose to upload a URL on Arcana to scan using the VirusTotal API. The program will first check if the URL is a valid one, if it is, the URL upload can then occur. If failed, the program will ask the user to try again. Phishing or malicious URLs will then be flagged out and displayed to the user.

```
Choose an option: 1
Input URL you would like to scan eg. https://www.facebook.com: https://www.facebook.com
| URL | Verdict |
|-----|-----|
| https://www.facebook.com | Safe |
```

Figure 7. Performing VirusTotal scan on any desired URL

K. Live Event Log Analysis of User's Device

The function will parse the .evtx log files to assist the user with the analysis. There is an automated analysis option that analyses the three main logs of concern, the Application, Security, and System logs and identifies all the event IDs and events that may be important for the analysis. With the information, the user can further proceed with the examination or prove that the user's hypothesis is correct.

```
186: Parameters.evtx
187: Security.evtx
188: Setup.evtx
189: State.evtx
190: System.evtx
191: Windows Azure.evtx
192: Windows PowerShell.evtx
193: *ALL LOGS
194: **Automated Scan for suspicious events
Enter numerical value of your choice of log file scan: 194
Starting automated scan on Application log...
```

Figure 8. Prompting the user to select desired log scan

```
The following Event IDs are found:
Event ID: 7045 Log File Name: System.evtx
Event ID: 7040 Log File Name: System.evtx
Event ID: 7023 Log File Name: System.evtx
Event ID: 7034 Log File Name: System.evtx
Event ID: 7024 Log File Name: System.evtx
Scan complete.
```

Figure 9. Automated scan for suspicious events completed

L. Displaying of Arcana's User Manual

The function will open the user manual for Arcana. The user manual contains information on how to set up, install and operate the Arcana program.

When the application prompts the user for input, the user can use the following commands:

Commands	Description
break	To return to the previous menu
exit	To stop the application

Figure 10. Instructions found in the user manual

M. Displaying of Arcana's GitHub

This function will direct the user to Arcana's GitHub repository. The repository will include an installation guide, user manual, source codes as well as other documentation for Arcana.

V. SOLUTION COMPARISON

As mentioned in the previous section, Arcana can perform logical analysis of disk images, display all existing files in the disk image, search for a specific file with the use of keywords, conduct VirusTotal scan on any suspicious files or browsing history URL and export the malicious file for further analysis. It is also able to carry out live analysis of the users' local directory, perform VirusTotal scan on a specific file or URL and view suspicious event logs.

Arcana will aid users in verifying if a disk or disk image has been infected with malware with the use of the VirusTotal API. This helps to prevent harm to the users and protect them from possible malware and ransomware attacks.

Autopsy is an open-source digital forensics tool that has features such as hash filtering, keyword search and web artefacts.

Encase is a paid digital forensics tool that has a comprehensive search function, powerful processing capability and a user-friendly interface.

Both digital forensic tools mentioned above do not provide any form of malware analysis in their default installations. However, Autopsy has a VirusTotal extension that requires an additional installation setup. Furthermore, Arcana has a live event analysis feature that is not included in Autopsy or Encase as well. This live analysis of Windows event logs can inform the user that his machine might potentially be compromised.

Autopsy requires all its modules to be up and running during the ingest stage even for basic functions such as hash lookup. While Arcana does not need to have as many modules on start-up and therefore, is faster when compared to Autopsy.

A raw image file with the E01 extension, around 4.07 gigabytes (GB) in size, was uploaded to both Autopsy and Arcana. This was done to compare the time taken for both programs to process the raw image and to extract artefacts from it. For Autopsy, it took 24 minutes and 30 seconds to process the image while it only took 4 minutes and 50 seconds for Autopsy to do the same task.

Arcana even allows users to conduct a malware scan on files and URLs outside the disk image, unlike the abovementioned digital forensic tools which can only conduct scans and analysis on the items within the disk image.

VI. DETAILED ARCHITECTURE

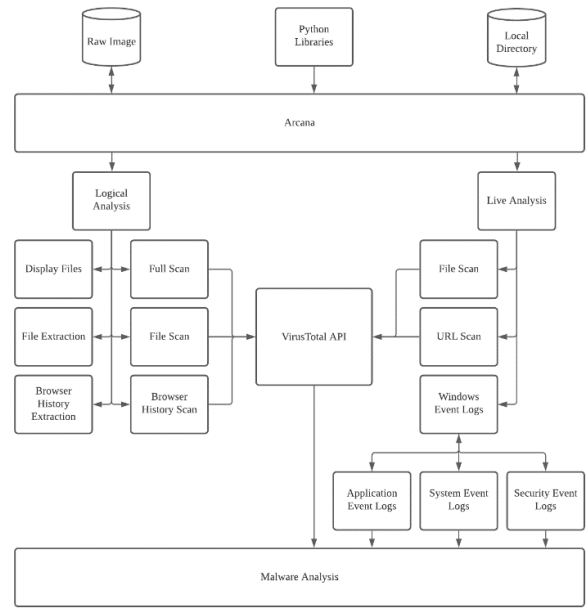


Figure 11. Detailed architecture of Arcana

VII. RESULT VERIFICATION

A. Experiment 1. File Integrity

Maintaining data integrity is essential for a forensic investigator when analyzing evidence. This ensures that the evidence will be permissible in court. As such, to maintain data integrity, the hash checksum is important. For checking the hash integrity, we will be using the file named "Fiesta Bkgrd.jpg". As seen in the figures below, the Arcana Checksum of the evidence file is the same as that of Autopsy which means that the checksum is valid.

Modify Date	Change Date	Size	File Path	SHA256 Hash
8-10-23 21:35:59	2004-08-04 12:00:00	5048	/Program Files/Common Files/Microsoft Shared/S... 610ca8030a4c730a2e2c200133e41497c81f633779...	

Figure 12. Arcana checksum results

Fiesta Bkgrd.jpg - Properties	
Properties	
Name	Fiesta Bkgrd.jpg
Advanced Previewer	C:\Windows\Media
Location	Ang.jpg-2005-dmcsa-cs-EP1-v1_v2\Program Files\Common Files\Microsoft Shared\Stamps\Fiesta Bkgrd.jpg
Created Time	2004-08-04 12:00:00
Change Time	2005-10-20 16:35:59
Access Time	2005-10-20 16:35:59
Created Time	2005-10-20 16:35:59
Size	5048
Flags (DPI)	32-bit color
Flags (Meta)	256 colors
Format	JPEG
MD5 Hash	610ca8030a4c730a2e2c200133e41497c81f633779...
SHA-256 Hash	610ca8030a4c730a2e2c200133e41497c81f633779...
MIME Type	image/jpeg
Extension	.jpg
Modified Time	8-10-23 21:35:59

Figure 13. Autopsy checksum results

Arcana also has an export function so that the forensics investigator can export the file. If he wishes to further investigate using the certUtil in command prompt, we can see that Arcana preserves data integrity as the hash of the file that

was exported, still matches as to when the file was in the raw image.

```
C:\Users\afiqb\PycharmProjects\App\extracted>certUtil -hashfile "Fiesta Bkgrd.jpg" SHA256
SHA256 hash of Fiesta Bkgrd.jpg:
61dca803bba4c726e2e2c2d0133ea414b7c8b1f83337f0a207e9192cdc9f6eb3
CertUtil: -hashfile command completed successfully.
```

Figure 14. certUtil checksum results

B. Experiment 2. File Scan

To ensure that the results of the virus scan done on Arcana are the same as on the VirusTotal website itself, an experiment is essential. The virus scan component is an important feature of Arcana, as it is the main objective of the entire project, to detect traces of malware on files. Arcana leverages on the VirusTotal API, therefore its scan results should mirror that of VirusTotal. The eicar.com file was uploaded on both Arcana as well as VirusTotal. The EICAR sample file is a popular file that is used for malware detection testing. Even though it will be flagged out during the antivirus scan as malicious, however, its contents are not. Another non-malicious file was uploaded to verify a true negative case.

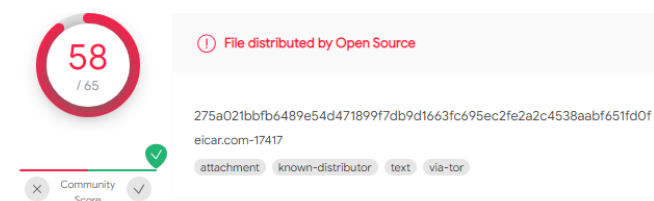


Figure 15. VirusTotal file scan positive result

```
Input directory of file you would like to scan: D:\Downloads\Anti Windows Defender\
File | Verdict
-----+-----
D:\Downloads\Anti Windows Defender\ | Malicious
```

Figure 16. Arcana's VirusTotal file scan positive result

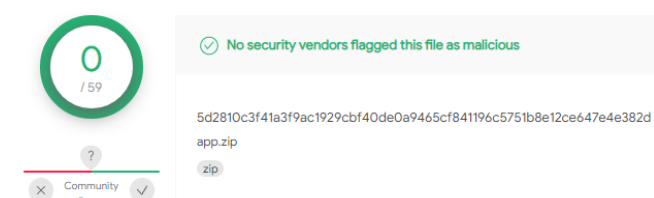


Figure 17. VirusTotal file scan negative result

```
Input directory of file you would like to scan: D:\Downloads\App.zip
File | Verdict
-----+-----
D:\Downloads\App.zip | Safe
```

Figure 18. Arcana's VirusTotal file scan negative result

C. Experiment 3. URL Scan

An experiment to test the validity of a URL is another feature that can be done by Arcana. Phishing sites are common attacks that could be used to steal information, money, or implant malware on an unsuspecting victim. Therefore, it is a must that the results obtained from Arcana are the same as on the VirusTotal website. This is due to Arcana deriving this

functionality from the VirusTotal API. A test was done by scanning a fake malicious website that was provided by wicar.org. The WICAR website provides links to several fake websites, to test the detection capabilities of antivirus software in discovering malicious websites. A separate safe URL was also uploaded to confirm that non-malicious URLs will not be flagged out by Arcana.



Figure 19. VirusTotal URL scan positive result

```
Input URL you would like to scan eg. https://www.facebook.com: http://malware.wicar.org/data/eicar.com
URL | Verdict
-----+-----
http://malware.wicar.org/data/eicar.com | Malicious
```

Figure 20. Arcana's VirusTotal URL scan positive result

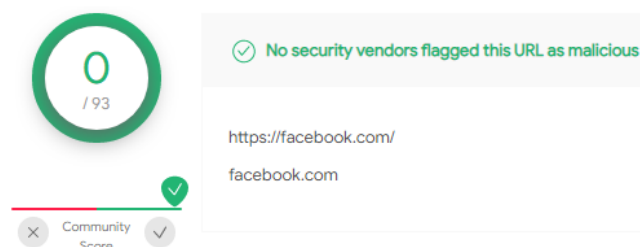


Figure 21. VirusTotal URL negative scan result

```
Input URL you would like to scan eg. https://www.facebook.com: https://www.facebook.com
URL | Verdict
-----+-----
https://www.facebook.com | Safe
```

Figure 22. Arcana's VirusTotal URL negative scan result

D. Experiment 4. Windows Event Log

The existing Event Viewer application developed by Microsoft Windows only allows users to view windows log by sorting them into Application, Security, Setup, System and Forwarded events. Users would then have to manually search for any suspicious event logs for more information. Meanwhile, Arcana can process and display an automated output of suspicious event logs that can be configured to include or omit certain event IDs during the automated scan. The output provided by Arcana only provides the suspicious event IDs detected by the program and thus, it is easier for the forensics investigator to read and identify the event of interest. The user will then be able to further analyse that event or other events with similar event IDs.

The event ID for failed logon event is "4625". When using Windows Event Viewer, users will have to use the find option, and browse through the event logs one event at a time. This process will take too much time and will prove unproductive for a digital forensics investigator.

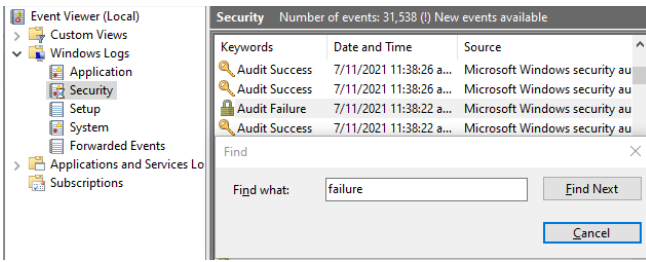


Figure 23. Windows Event Viewer

Arcana allows users to simply key in the relevant event IDs in either the “appEidList.txt”, “secEidList.txt” or “sysEidList.txt” text files when locating for specific Windows events. For this example, users will only need to enter “4625” in the “secEidList.txt” text file and run Arcana along with the “automated scan for suspicious events” option and all event information with the specified event ID will be displayed.

```
Scan complete.
Starting automated scan on Security log...
b'<Event xmlns="http://schemas.microsoft.com/win/2004/08/events/event">
b'<Event xmlns="http://schemas.microsoft.com/win/2004/08/events/event">
b'<Event xmlns="http://schemas.microsoft.com/win/2004/08/events/event">
Scan complete.
```

Figure 24. Arcana scan for suspicious events

VIII. LIMITATIONS

A. Graphic User Interface (GUI)

Currently, Arcana is coded to provide the user with Command Line Interface (CLI) which makes it much more portable and lightweight. To enhance the user-friendliness of our program, Arcana could be coded using python using libraries such as Tkinter which allows its features to be more appealing. Arcana could also be hosted via localhost as a web-based file analysis solution to provide its users with a more pleasant experience as compared to seeing a CLI [13].

B. Evidence File Format

Currently, for the logical file analysis, it supports the E01 file format. To improve the functionality of Arcana, it can be improved to read other raw data images with the file format of .dd or .img. By having more raw data image supporting file type, it will allow Arcana to be more marketable to users.

C. Bruteforcing encrypting files

As Arcana mainly focuses on file analysis to identify malware, there will be situations where the file will be in a folder, and it is encrypted. As such, to break the encryption and gain access to the file, Arcana could have an additional feature where it is able to crack the password hashes. It could be linked to CrackStation which uses a possible wordlist to decrypt the hashes of the file, which will allow the users to have access to the file.

D. Other Anti-Virus APIs

Arcana currently supports the use of VirusTotal API which allows the user to identify malicious files or suspicious links.

Arcana can use more than one API to identify the malicious file or URL, such as the MetaDefender cloud. This allows cross-comparison on the various malware analysis as well as providing Arcana with an alternative form of verification. In the case that one of the APIs is down, it has an alternative solution to conduct its analysis [14].

E. Windows Event Log Analysis

Arcana provides a live raw analysis to identify the various Windows log event. Currently, it can highlight to the user the various suspicious logs it has found, and the user is to investigate the highlighted logs. To enhance this feature, the Arcana can be programmed to not only do live Windows event log analysis but also files where it can identify the files which caused the suspicious Windows Event to appear. Following that, Arcana will need to be launched using administrator privileges in order to run the Windows event log analysis feature.

F. Executable File

Arcana currently requires having all its dependencies installed to be able to run. If the program is to be made executable and can be installed on the user's computer, it will allow the user to run the program without needing to install all the dependencies. The executable file will contain all the dependencies for the computer to run after it is installed by the user's computer. This will allow Arcana to be more user friendly as the initial set-up will not require much effort.

G. Report of Suspicious File, Window Events and Links

Arcana currently does not provide the user with a report summary to its users. This will cause the user to identify the suspicious findings after it has been flagged out by the Arcana features. Arcana has the potential to provide the user with a summary report of its analysis. Instead of prompting the user to input which files or links he wishes to conduct the investigation, the program can provide a summary of suspicious links and files so that the user can get an overview of what to look out for before conducting further analysis using Arcana.

IX. CONCLUSION

Digital forensics is a critical component within the field of cyber security. As humans are getting more and more interconnected with each other via the digital world, the amount of data that is being stored in our devices is growing exponentially. This data must be stored and transmitted securely to deter and prevent attackers. The rapid rise in the usage of malware in modern society has brought politically motivated attacks and huge financial losses to corporations. Hence, with tools such as Arcana, the team aims to provide its users with functionalities such as raw image analysis, URL analysis, and Windows event logs analysis. These features will allow its users to conduct a thorough digital forensic investigation while preserving the integrity of the source. An infected device or a potentially infected host with malware can be identified with the aid of Arcana.

X. ARCANA RESOURCES

GitHub Repository: <https://github.com/afiqbusari7/Arcana>

Demonstration: <https://youtu.be/hvgIOqOFwjE>

REFERENCES

- [1] J. Johnson, "Topic: Ransomware," *statista.com*, Sep. 09, 2021. <https://www.statista.com/topics/4136/ransomware/> (accessed Nov. 01, 2021).
- [2] W. Turton and K. Mehrotra, "Hackers Breached Colonial Pipeline Using Compromised Password," *bloomberg.com*, Jun. 04, 2021. <https://www.bloomberg.com/news/articles/2021-06-04/hackers-breached-colonial-pipeline-using-compromised-password> (accessed Nov. 01, 2021).
- [3] Committee on National Security Systems, "FACT SHEET National Policy on the Use of the Advanced Encryption Standard (AES) to," Jun. 2003. Accessed: Nov. 01, 2021. [Online]. Available: <https://csrc.nist.gov/csrc/media/projects/cryptographic-module-validation-program/documents/cnss15fs.pdf>.
- [4] S. Nakamoto, "Bitcoin: a Peer-to-Peer Electronic Cash System," Oct. 2008. Accessed: Nov. 01, 2021. [Online]. Available: <https://bitcoin.org/bitcoin.pdf>.
- [5] R. Baylor, J. Brown, and J. Martineau, "Extortion Payments Hit New Records as Ransomware Crisis Intensifies," *Palo Alto Networks Blog*, Aug. 09, 2021. <https://www.paloaltonetworks.com/blog/2021/08/ransomware-re-crisis/> (accessed Nov. 01, 2021).
- [6] A. Thomas, T. Grove, and J. Gross, "More Cyberattack Victims Emerge as Agencies Search for Clues," *Wall Street Journal*, May 13, 2017.
- [7] P. Kerr, J. Rollins, and C. Theohary, "CRS Report for Congress The Stuxnet Computer Worm: Harbinger of an Emerging Warfare Capability Analyst in Nonproliferation," Dec. 2010. Accessed: Nov. 01, 2021. [Online]. Available: <https://cyberwar.nl/d/R41524.pdf>.
- [8] D. Kushner, "The Real Story of Stuxnet -IEEE Spectrum <https://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet> 1/6 The Real Story of Stuxnet How Kaspersky Lab tracked down the malware that stymied Iran's nuclear-fuel enrichment program," Feb. 2013. Accessed: Nov. 01, 2021. [Online]. Available: <https://courses.cs.duke.edu/spring20/compsci342/netid/readings/cyber/stuxnet-ieee-spectrum.pdf>.
- [9] Z. Bazrafshan, H. Hashemi, S. M. H. Fard, and A. Hamzeh, "A Survey on Heuristic Malware Detection Techniques," *researchgate.net*, May 2013. https://www.researchgate.net/profile/Zahra-Bazrafshan-2/publication/260729684_A_survey_on_heuristic_malware_detection_techniques/links/54df00e60cf2953c22b0d005/A-survey-on-heuristic-malware-detection-techniques.pdf (accessed Nov. 01, 2021).
- [10] VirusTotal, "VirusTotal," *virustotal.com*, 2019. <https://www.virustotal.com> (accessed Nov. 01, 2021).
- [11] Kaspersky, "Kaspersky Threat Intelligence Portal," *opentip.kaspersky.com*. <https://opentip.kaspersky.com/> (accessed Nov. 01, 2021).
- [12] FortiGuard, "FortiGuard Labs Online Scanner," *fortiguard.com*. <https://www.fortiguard.com/faq/onlineScanner> (accessed Nov. 01, 2021).
- [13] Python Software Foundation, "tkinter — Python interface to Tcl/Tk — Python 3.7.2 documentation," *python.org*, 2019. <https://docs.python.org/3/library/tkinter.html> (accessed Nov. 05, 2021).
- [14] MetaDefender Cloud, "MetaDefender Cloud | Advanced threat prevention and detection," *metadefender.opswat.com*. <https://metadefender.opswat.com/> (accessed Nov. 05, 2021).

END OF DOCUMENT