



# Detecting Phishing, Keystroke and Keylogger Attacks on Computing Resources

<sup>1</sup>Avinash Sonule, <sup>2</sup>Mukesh Kalla

<sup>1</sup>Department of Computer Engineering, A. C. Patil College of Engineering, Kharghar, Navi Mumbai -410210, Maharashtra, India

<sup>2</sup>Sir Padampat Singhania University, Udaipur, Rajasthan, India.

**Abstract:** Cybersecurity attacks are growing both in frequency and sophistication over the years. This increasing sophistication and complexity call for more advancement and continuous innovation in defensive strategies. Traditional methods of intrusion detection and deep packet inspection, while still largely used and recommended, are no longer sufficient to meet the demands of growing security threats. As computing power increases and cost drops, Machine Learning is seen as an alternative method or an additional mechanism to defend against malwares, botnets, and other attacks. This paper explores Machine Learning as a viable solution by examining its capabilities to classify malicious websites as well as detecting keyloggers in a network. First, a strong data analysis is performed resulting in multiple extracted features from the initial datasets. All these features are then compared with one another through a feature selection process. Then, our approach analyzes machine learning algorithms against a dataset containing common characteristics. By leaving your computer unlocked while you are away for seconds can give hackers all the time, they need to obtain your personal information from your computer.

**Keywords**—Machine learning, Phishing attack, Keystroke, Keylogger.

## I. INTRODUCTION

Cybersecurity is evolving and the rate of cybercrime is constantly increasing. Sophisticated attacks are considered as the new normal as they are becoming more frequent and widespread. This constant evolution also calls for innovation in the cybersecurity defense. There are existing solutions and a combination of these methods are still widely used. Network Intrusion Detection and Prevention Systems (IDS/IPS) monitor for malicious activity or policy violations. Signature based IDS relies on known signatures and is effective at detecting malwares that match these signatures. Behavior-based IDS, on the other hand, learns what is normal for a system and reports on any trigger that deviates from it. Both types, though effective, have some weaknesses. Signature-based systems rely on signatures of known threats and are thus ineffective for zero-day attacks or new malware samples.

Traditional behavior-based systems rely on a standard profile which is hard to define with the growing complexity of networks and applications, and thus may be ineffective for anomaly detection. Full data packet analysis is another option; however, it is both computationally expensive and risks exposure of sensitive user information. Machine Learning (ML) has gained a wide interest in many applications and fields of study, particularly in Cybersecurity. With hardware and computing power becoming more accessible, machine learning methods can be used to analyze and classify bad actors from a huge set of available data. There are hundreds of Machine Learning algorithms and approaches, broadly categorized into supervised and unsupervised learning. Supervised learning approaches are done in the context of Classification where input matches to an output, or Regression where input is mapped to a continuous output. Unsupervised learning is mostly accomplished through Clustering and has been applied to exploratory analysis and dimension reduction. Both of these approaches can be applied in Cybersecurity for analyzing malware in near real-time, thus eliminating the weaknesses of traditional detection methods.

Nearly every computer, including desktops, laptops, tablets and Smartphones take input from humans via keyboards. This is possible because there is a specification with every ubiquitous USB standard known as Human Interface Device (HID). Practically, this means that any USB device claiming to be a keyboard HID will be automatically detected and accepted by most modern operating systems including Windows, Mac OS, Linux or Android. Standard USB devices are too simplistic to reliably authenticate. Similarly, secure devices with signed firmware that could permit authentication are rare, leaving it unclear how to defend ourselves against this new attack. One can employ various approaches to penetrate a machine as a hacker or a penetration tester such as social engineering, exploiting vulnerabilities of the system, etc. One of the practical strategies used by the hackers is to plug in a USB stick to a machine. This can be done by using a USB device detected by a victim's computer as a HID (this is called BadUSB) and running the code without the knowledge or consent of the victim. For example, if the user is away for lunch and leaves his or her computer unattended, the hacker can plug in the USB in the victim's machine for malicious purposes.

Day by day cyber-attacks are increasing and the states are sponsoring hackers to do work. Recently there was an attack on power grid and hence in this tech world we are trying to safeguard people's data and other confidential information. If some individual who has no knowledge of the attacks cannot take necessary actions to check if there is an attack on the system therefore we need an application or a set of procedure that may help the individuals with no prior knowledge to detect and prevent the attacks.

The rest of the paper is organized as follows: section II gives related work for detection of system and network attacks. Section III gives in detail of Phishing, Keystroke, Keylogger Attacks. In section IV, proposed methods with all steps discussed. Section V discusses the results. Finally, section VI gives future direction and concludes the work.

## II. RELATED WORK

Many researchers have done work on detection of system and network attacks.

Rege M. et al [1] discusses how machine learning is being used in cyber security in both defense and offense activities, including discussions on cyber-attacks targeted at machine learning models. Specifically, they discuss the applications of machine learning in carrying out cyber-attacks, such as in smart botnets, advanced spear phishing and evasive malwares. They also explain the application of machine learning in cyber security, such as in threat detection and prevention, malware detection and classification, and network risk scoring.

Devakunchar R. et al [2] have done a review of Machine Learning and DL unit methods for network security domain. They have done a comparative survey on the Influence of Machine Learning Techniques on Intrusion Detection System (IDS).

Lakshmanarao A. et al [3] have survey of Machine learning techniques which are applied for solving major challenges in cybersecurity issues like intrusion detection, malware classification and detection, spam detection and phishing detection.

Asiri S. et al have done surveys of HTML and URL Phishing attacks and detection methods. They review the current state-of-art Deep Learning models to detect URL-based and Hybrid -based Phishing attacks in detail. They compare each model based on its data preprocessing, feature extraction, model design and performance.

Ruhani et al [5] provide an overview of keyloggers as a tool for hacking and the potential threats they pose to individuals and organizations. They explore the history of keyloggers, the types of keyloggers, how they work, and the different techniques used to detect and prevent keylogger attacks. They discuss the different types of keyloggers that exist, including hardware and software keyloggers, and explains how they can be used to steal sensitive data. They have given a comprehensive guide to understanding keyloggers as a tool for hacking and the measures that can be taken to prevent and detect keylogger attacks.

Samsoni et al [6] explained how keyloggers work so that prevention can be carried out by carrying out various kinds of solutions and also to maintain data security systems and know supporting and anti-keylogger software.

## III. PHISHING, KEYSTROKE AND KEYLOGGER ATTACKS

**PHISHING ATTACKS:** One common social engineering attack technique is phishing, which is intended to get user information, including login passwords. It happens when a perpetrator tricks a victim into opening an email, text message, or instant message by disguising themselves as a reliable source. The purpose of spoof websites is to deceive people into disclosing sensitive personal and financial data, like credit card numbers, account IDs, and passwords. Phishing emails frequently have a link to click and an urgent request for the recipient to reply right away, all while appearing to be from reliable sources.

### KEYSTROKE AND KEYLOGGER ATTACKS

Keyloggers are executed on the focused machine to record client's keystrokes logging movement, lastly giving over that private information to an outsider. Keyloggers are utilized for both lawful and illicit purposes. Keyloggers are generally utilized by assailants to take private information of an individual or an association. In the past many credit card details have been compromised by attackers with the help of keyloggers. Henceforth, keyloggers are one of the most hazardous sorts of spyware till date. A malicious program having keystroke logging feature using an example of real-time online banking system.

When an attacker gains physical access to your computer devices, they can wiretap the physical hardware like a keyboard to collect the valuable data of the user. This strategy is totally reliant on some actual properties, either the sound transmission created when a client is composing or the electromagnetic spread of a remote console. External keyloggers or hardware keyloggers are small electronic devices which are placed in between keyboard and motherboard, this procedure requires the attackers to have physical access to the system which they are intended to compromise.

Keyloggers, or keystroke loggers, are tools that record what a person types on a device. While there are legitimate and legal uses for keyloggers, many uses for keyloggers are malicious. In a keylogger attack, the keylogger software records every keystroke on the victim's device and sends it to the attacker.

#### IV. PROPOSED METHODS

Our proposed work is divided into URL detection and Keystroke and Keylogger detection as shown in Figure 1.

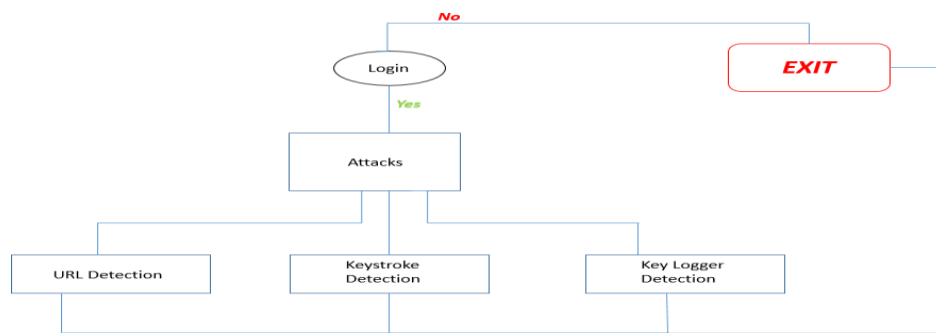


Figure 1: Proposed Work

#### URL Detection

Our first work is related to machine learning web content detection which focuses on HTML files because of their richer structure and higher information content. Since these approaches use orthogonal input information, there is certainly room for HTML- based and URL-based approaches to be combined into an even more effective ensemble system. A body of work including and attempts to detect malicious web content by extracting features from HTML and JavaScript, and feeding them into a machine learning system. This approach extracts a wide variety of features from HTML page and JavaScript static content, and then feeds this information to KNN machine learning algorithms. We use a parser-free tokenization approach to compute a representation of HTML files. A parser-free representation of web content allows us to make a minimal number of assumptions about the syntax and semantics of malicious and benign documents.

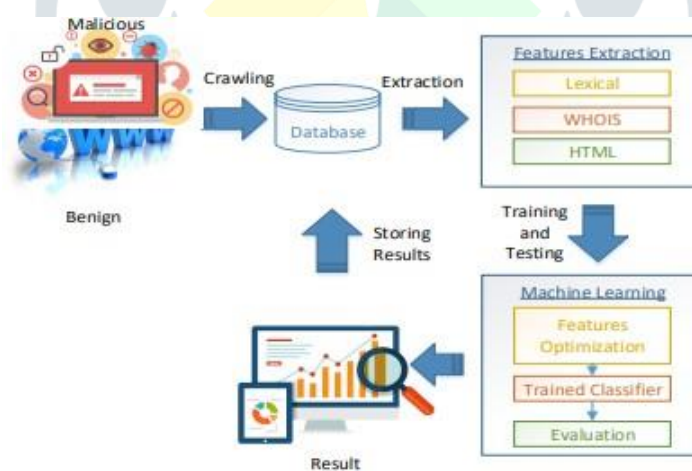


Figure 2: URL Detection System

Figure 2 presents the main components of the malicious URLs detection system. There are three phases in detection including data collection, machine learning, and database. The data collection begins with crawling all the URL including malware and benign website applications. Then the data are processed through a features selection engine to collect relevant features as shown in Table 1 for training purposes.

Table 1: Features of URL

Features	Description
Token Count	The total number count of words in the URLs
Rank Host	The popularity ranking of the hostnames
Rank Country	The popularity ranking of the URLs (websites) among countries
ASNno	Autonomous System Number as the classifier for the IP of each URLs
Sec_sen_word_cnt	The security sensitive word count from the URLs
Avg_token_lenght	The total average number length count of the URLs
No_of_dots	The number of dots in the URLs
Length_of_url	The length of the URLs
Avg_path_token	The average number of the path for URLs

### Keystroke Attack Detection

The keystroke attack is the attack that is done using the USB Rubber ducky it is nothing but a penetration testing device that is capable of passing through the antivirus and remains undetected until you find out the device is not the storage device and works as a keyboard instead. To prevent the attack before it processes further to infect your system we block the entire keyboard and the key cannot be used until we enter a certain combinations that are known to us.

### Anti-Keylogger

Anti keylogger is used to detect and close the software in which keylogger is running in stealth mode in the system. It is software based anti keylogger.

### Scan for Keylogger

Anti keylogger is used to turn off the keylogger working in the computer system. For this software based anti keylogger is developed. Once the start scan button in the anti keylogger software is pressed it will get all running processes from the operating system.

It will compare whether the specified executable file name and the running process executable file name match. If it matches then the software will be closed. So, if any keylogger running in that software in stealth mode will also get closed. Stop scan button will stop the scanning process.

## V. RESULTS & DISCUSSION

Figure 3 shows Admin Panel through which we can detect phishing attack, keylogger and keystroke.

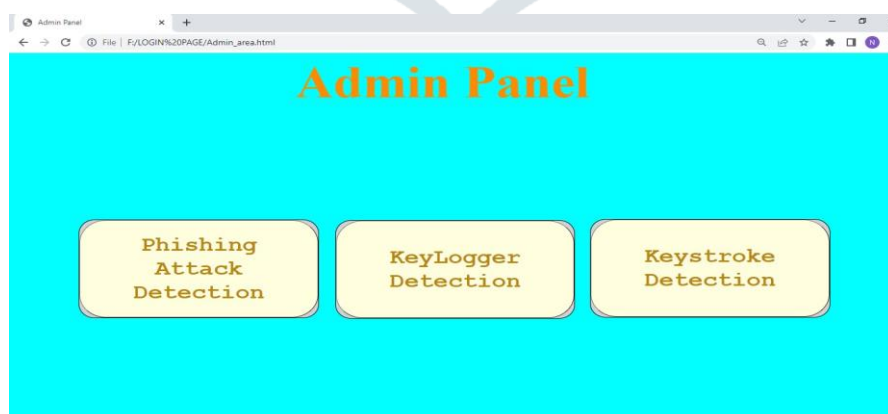


Figure 3: Admin Panel

For Phishing attack detection, the module shown in Figure 4 is used. The URL is given as input.



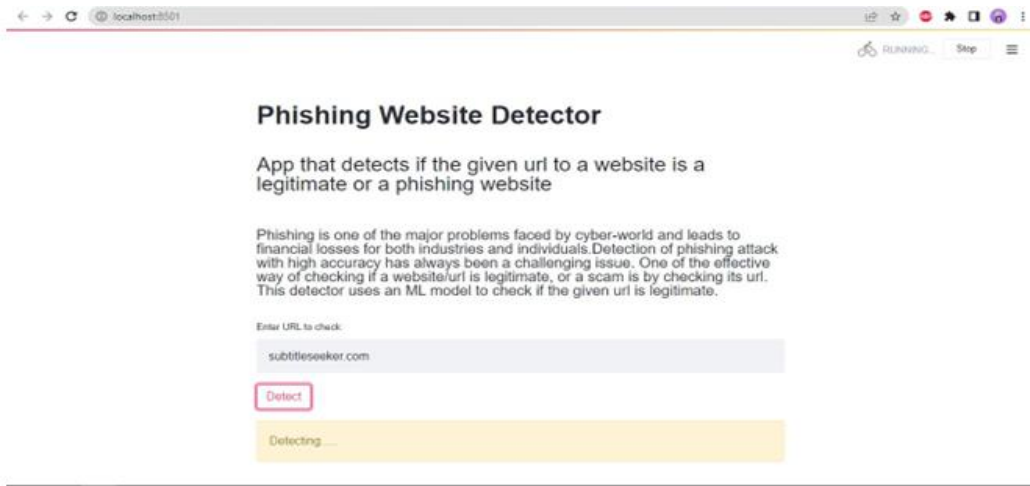


Figure 4: Phishing Website Detector Module

This module checks whether given website is dangerous or safe. The examples of dangerous and safe website are shown in Figure 5 and Figure 6 respectively.

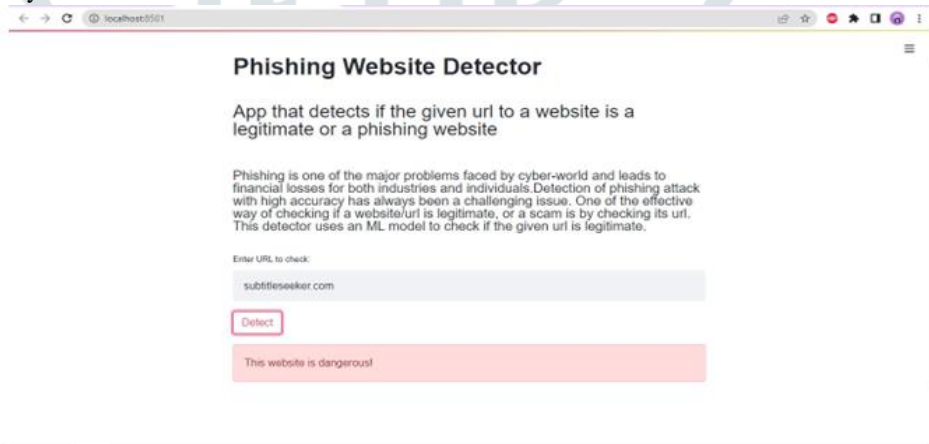


Figure 5: Dangerous Website



Figure 6: Safe Website

Figure 7 shows keylogger is running in stealth mode in the system. User is typing first website name as shown in Figure 8 and then second website as Figure 9.

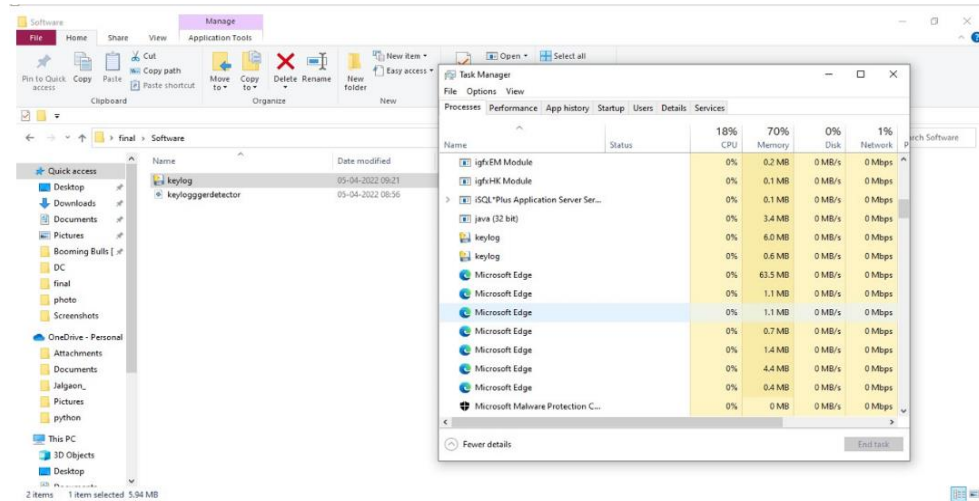


Figure 7: Keylogger Started

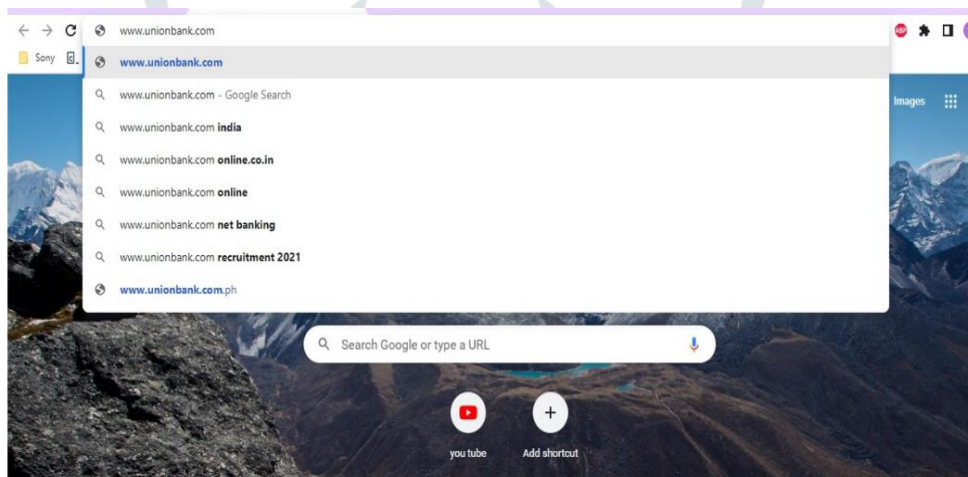


Figure 8: Typing First Website

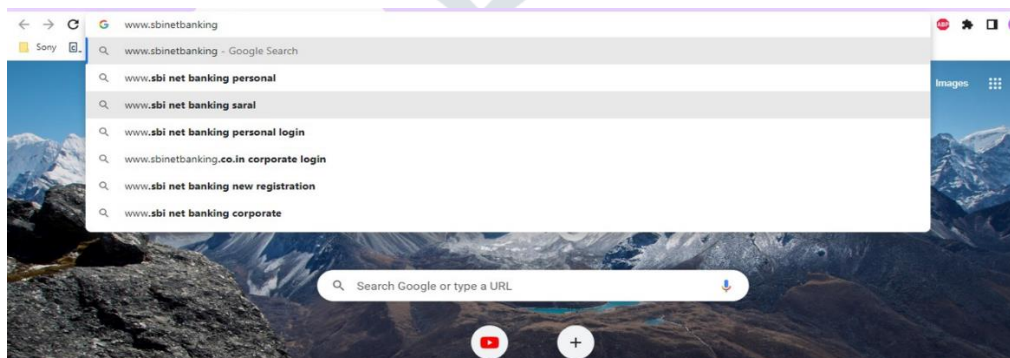


Figure 9: Typing Second Website

Anti keylogger as depicted in Figure 10 is used to create log file to records the keystrokes /Key Logs. Anti keylogger gives alerts/mail to the Admin as shown in Figure 11. If the Keylogger is detected, admin kill the process as depicted in Figure 12. If there is no Keylogger, admin gets the message as shown in Figure 13. Anti Keylogger can also turn off the keylogger working in the computer system.

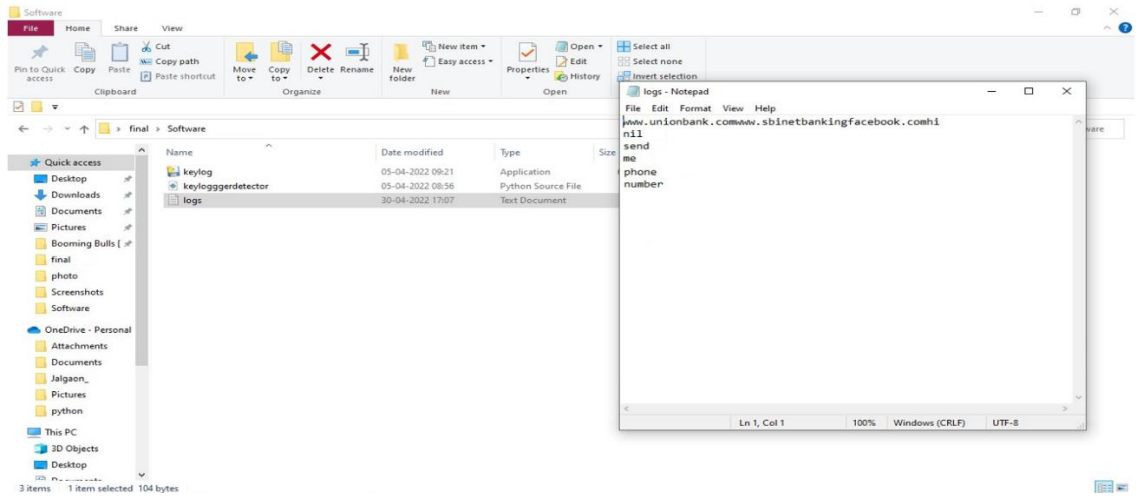


Figure 10: Log File Created

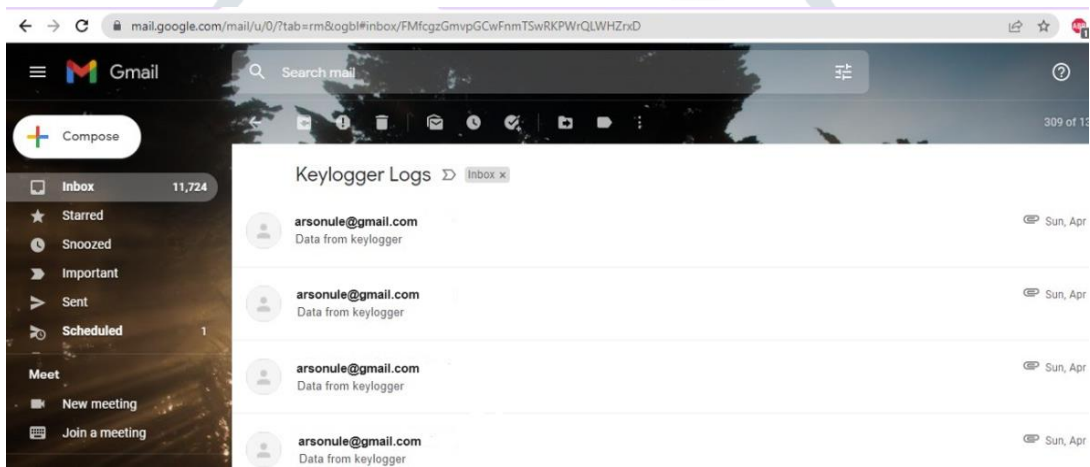


Figure 11: Mailing the Logs to Admin

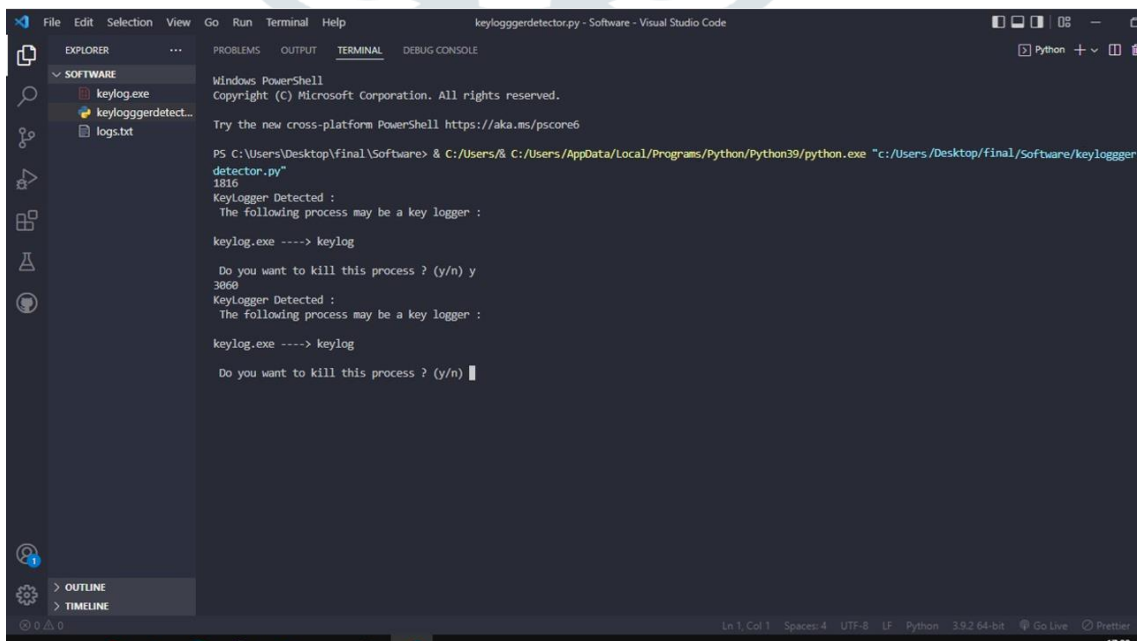
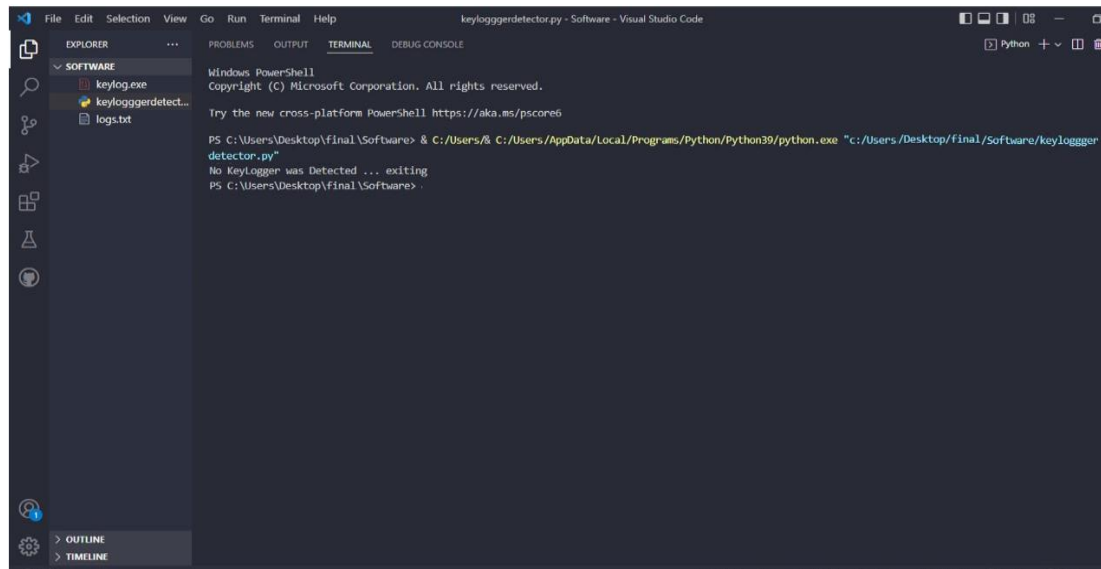


Figure 12: Killing the Keylogger



```

keyloggerdetector.py - Software - Visual Studio Code
EXPLORER
SOFTWARE
keylog.exe
keyloggerdetect...
logs.txt
PROBLEMS
OUTPUT
TERMINAL
DEBUG CONSOLE
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\Desktop\final\Software> & C:/Users/& C:/Users/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/Desktop/final/Software/keyloggerdetector.py"
No Keylogger was Detected ... exiting
PS C:\Users\Desktop\final\Software>

```

Figure 13: No Keylogger

## VI. CONCLUSION AND FUTURE WORK

The Phishing website detector takes the URL as input and based on several factors, detects the URL and classifies it into safe or dangerous. The keylogger does to get all secret data from clients of the framework by getting their keystrokes and mouse clicks without the information on the client. So, the client of the framework is ignorant of things occurring in the foundation. The software is able to monitor data and store the data in a specific folder. The software is also able to hide itself from the owner of the system while it runs in the background. Our keylogger detector can find such keyloggers and kill them as per user requirement.

As future work a bigger and recently updated dataset can be used in order to increase the efficiency of the system. We can use other models in order to validate. More factors that influence the web content can be included. Image processing can be added to detect malicious image contents. More powerful keyloggers can be detected using Heuristic Analysis.

## REFERENCES

- [1] Rege M. and Raymond Blanch K. Mba, "Machine Learning for Cyber Defense and Attack", DATA ANALYTICS 2018: The Seventh International Conference on Data Analytics, ISBN: 978-1-61208-681-1.
- [2] Devakunchari R., Sourabh, Prakhar Mali, "A Study of Cyber Security using Machine Learning Techniques" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-7C2, May 2019.
- [3] Lakshmanarao A., Shashi M. "A Survey On Machine Learning For Cyber Security", International Journal of Scientific & Technology Research Volume 9, Issue 01, January 2020.
- [4] Asiri S. Yang Xiao, , Saleh Alzahrani, Shuhui Li, Tieshan Li, "A Survey of Intelligent Detection Designs of HTML URL Phishing Attacks", IEEE Access, Volume 11, 2023.
- [5] Adi Badiozaman Bin Ruhani, Muhamad Fadli Zolkipli, "Keylogger: The Unsung Hacking Weapon", Borneo International Journal ISSN 2636-9826 (online) VOL 6 NO 1 (2023) March.
- [6] Samsoni, Ditonius Zebua, Basir, Bayu Aji Pamungkas, Hafidsyah Eka Prayogi, Rifaldie Muhammad, Supri Wahyudi, Wira Samudra, "Keylogger Threats in Computer Security Aspects", International Journal of Integrative Sciences, Vol. 2 No. 6 (2023): June 2023.