**1. Introduction**

ThreatScraper was developed as an interface with VirusTotal.com's API to provide malware progression tracking that is not available to users of the website. We will begin with a description of what VirusTotal is, how it is used, and how ThreatScraper works with VirusTotal to provide data on the adoption and detection rates of Anti-Virus programs over time.

**VirusTotal.com**

VirusTotal is a website that hosts over 70 separate types of Anti-Virus programs that will simultaniously scan files submitted to the site for analysis. These programs include well known platforms such as Microsoft Defender, Symantec, McAfee and many more. Users of the platform can upload files to the website, which will be scanned by all programs hosted on the site to determine whether or not a file is malicious. While there are many tools at the disposal of users of the website, it lacks the ability to show the progression of malware detections over time as more Anti-Virus programs update their virus signatures. Therefore, to assist our ongoing research on the speed of malicious file detection and the indication of signature sharing among groups of providers, we developed and utilized a program called ThreatScraper to track detections over time.

A screenshot of a computer

Description automatically generated

Fig 2: VirusTotal.com

**ThreatScraper**

ThreatScraper was developed in Python and utilizes VirusTotal's version 3 of its API interface to submit files, pull reports, and rescan submitted files. The data pulled from VirusTotal's website is saved into an Excel document specified by the user, and each new report is saved on a new line in the document. To help visualize the data, the number of detections over time are displayed on a line graph that will display after pulling a report. Along with this is a pie chart showing the total number of positive detections, contrasted by the percentage of negative detections amongst Anti-Virus programs hosted on VirusTotal. Finally, a table is displayed with the results of the most recent scan, containing names, versions, and dates of each Anti-Virus program hosted on the website.

A screenshot of a computer

Description automatically generated

Fig 1: ThreatScraper

**2. ThreatScraper Walkthrough**

ThreatScraper's Main Window

Upon launching the program, the main window will be displayed with several fields that the user will need to interact with:

A screenshot of a computer

Description automatically generated

Fig 3: Main Window

1. The API Key field requires the user to input their VirusTotal API Key. This key can be found on the VirusTotal website, within the user's main login screen. VirusTotal requires this key in the API request to verify who the request came from, especially with free users as the number of requests made per minute are limited.
2. The Hash Value field is required so that VirusTotal knows which file the user is interested in rescanning or pulling a report on. The specified Hash can be in the form of an MD5, SHA-1, or SHA-256, and the user must specify which one they have used using the dropdown button on the right.
3. The Excel File field is required, as the user must specify which file to save the report information to. This file can be anywhere on the user's computer and is the basis for several other visualizations used in the program.
4. With the previous information provided to the program, the "Check VirusTotal" button is ready to use. The user also has the option to select the "Rescan hash" box, which will submit a rescan request along with the report request to the API. This will add a 10 second delay to the report to allow VirusTotal to rescan the file for premium account holders. Free account holders are still able to rescan the file, but need to be aware that there can be up to a 20-minute delay before the rescan request is performed by VirusTotal. This may cause the updated information from the scan to be unavailable until after the delay period.
5. The Schedule Times field is optional but allows the user to specify times during the day to rescan and pull reports from VirusTotal. Currently, there is no limit set by the program as to how many times can be scheduled, but the user will need to be aware of API request limits imposed on free account users of the website. This feature utilizes the "Start Schedule", and "Stop Schedule" buttons to control the scheduling function.
6. The "Submit File for Analysis" button allows the user to specify a file on their computer to be submitted to VirusTotal. Once selected and submitted, the console output will display the MD5, SHA-1, and SHA-256 hashes of the file to the user for their reference. The program will not automatically perform a report request, as a new Excel file will need to be created and provided to the program by the user.

ThreatScraper's Visualizations

Once a report request is received and processed by VirusTotal, the results are saved into the specified Excel file. ThreatScraper will then pull the cumulative data saved in the file and generate several visualizations for the user to view. The generated visualizations are dynamically updated with new report information as it is received. This would occur if the user had created and started a schedule, or each time the user manually selects the "Check VirusTotal" button.

1. The Detection Results Pie Graph is provided to the user as a quick reference to the current number of positive and negative detections on VirusTotal. This pie chart will always show the results from the most recent scan and is subject to change in response to the most recent report information from the website.

A screenshot of a computer

Description automatically generated

Fig 4: Pie Graph

1. The Detection Timeline is a line graph containing two datapoints from the report data. The Malware Detections line represents the total number of Anti-Virus programs with a positive detection as the number of total scans performed grows. The AV Utilization line represents the total amount of Anti-Virus programs that were available during the scan. This is provided to give context as to why there are sudden drops in detections that recover during subsequent scans. Please note that this graph represents scan numbers and not a timeline such as hours or days.

A screenshot of a computer

Description automatically generated

Fig 5: Detection Timeline

1. Finally, the Scan Results window contains the most recent report information on the individual Anti-Virus programs used to scan the file. The window is split into two categories: Positive results on the left side, and Negative results on the right. Both columns will display the name of the Anti-Virus program, the detection result, the signature version, the type of malware, and the date the Anti-Virus signatures were published.

A screenshot of a computer

Description automatically generated

Fig 6: Scan Results

**3. Conclusion**

In closing, ThreatScraper provides a compelling and highly informative interface that utilizes the powerful antivirus scanning capabilities of VirusTotal. By harnessing the Version 3 of VirusTotal's API, ThreatScraper not only allows users to schedule scans and re-scans of files manually or automatically, but it also captures, stores, and analyzes these scan results over time. This information is beneficial to observe trends in the progression of malware detection capabilities across numerous antivirus platforms, and the sharing of virus signatures between platforms.