**EVENT-DRIVEN MALICIOUS URL EXTRACTOR**

2021-085

Final Report - Draft

Renu Harshatha A.

(IT8034400)

B.Sc. (Hons) Degree in Information Technology Specialized in Cyber Security

Department of Computer Science and Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

February 2021

**EVENT-DRIVEN MALICIOUS URL EXTRACTOR**

2021-085

Final Report - Draft

B.Sc. (Hons) Degree in Information Technology Specialized in Cyber Security

Department of Computer Science and Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

February 2021

# **DECLARATION**

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

|  |  |  |
| --- | --- | --- |
| Name | Student ID | Signature |
| Renu Harshatha A. | IT18034400 | Text, letter  Description automatically generated |

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

..…………………….……. ………………………… Signature of the Supervisor Date

(Mr. Amila Seneviratne)

# **ABSTRACT**

The evolution of technology has been growing rapidly in all industries, along with the cybercrimes. Cybercrime is defined as criminal offenses carried out using the means of electronic devices connected with internet. This is an area which has been increasing in the past decade. Along with the evolution of technology the evolution of cybercrimes has being continued to grow. Cyber attackers use different mechanisms and techniques to conduct malicious exploitations on their victim. As of last decade, the damages caused by cybercrimes has increased from $485 million to over to $4 billion where 60% of small medium enterprises are prone to cyber-attacks.

Humans are considered to be the weakest and vulnerable link in order to carry out the malicious activity. Reports and research states that 98% attacks use social engineering as their main attack vector by cybercriminals, which are extremely effective. As people do not have a proper guidance and identification method to detect malicious URLs, is one of the main causes in the rapid increase of cybercrimes along with the technology advancement and transformation.

This research study proposes to identify malicious URLs based on events which happens both locally and globally using ensemble model which notifies the user real time with as a solution to this problem. In this research, the major components, the web application, and feature extraction will be used to identify the malicious URLs. Training the model with keywords, identifying the malicious URLs based on the trending events and incorporating scalability and ease of use will be used as components to identify, detect, and proactively block malicious URLs based on an event. For extraction process this research uses a well-trained ensemble model using Machine Learning and Deep Learning algorithm, which results accurate detection of malicious URLs. Creation of malicious URLs proportionally increases with a given occurred event. This results a delay in identification process.

To overcome this challenge, in this research we will be incorporating a scalable model using Nomad. This results to achieve the main output of the entire study, which is extracting malicious URLs with better accuracy and efficiency. During the course of identifying the malicious URLs, our research study also focuses on the protecting the users’ privacy as it is one major concerns in terms of cybersecurity.

Key words: - machine learning and deep learning, ensemble model, scalability, Nomad, malicious URLs

# **ACKNOWLEDGEMENT**

My sincere gratitude goes to my supervisor Mr. Amila Seneviratne for his guidance and support. His positive influence, dedication and patience has helped in all means for the success of this research project.

Mr. Kavishka Yapa, the co-supervisor of this research project has always been a positive influence and inspiration. It would be remiss of me not to express my gratitude to Dr. Lakmal Rupasinghe for willingly sharing his knowledge in Machine Learning and Federated Learning whenever required.

My sincere thanks to my group members S W Sajeth Jonathan, B L D V Wishvajith and A M Ramanayake for supporting me despite the challenges that I encountered during the course of the research.

Finally, to my biggest source of blessing and strength my beloved grandmother, my family, and my friend whose emotional support was much needed during the research.

# **TABLE OF CONTENTS**

[**DECLARATION** i](#_Toc85061499)

[**ABSTRACT** ii](#_Toc85061500)

[**ACKNOWLEDGEMENT** iii](#_Toc85061501)

[**TABLE OF CONTENTS** iv](#_Toc85061502)

[**List of figures** v](#_Toc85061503)

[**List of tables** vi](#_Toc85061504)

[**List of abbreviations** vii](#_Toc85061505)

[**List of appendices** viii](#_Toc85061506)

[**1.** **INTRODUCTION** 1](#_Toc85061507)

[**1.1** **Introduction** 1](#_Toc85061508)

[**1.2 Background Literature** 6](#_Toc85061509)

[**1.3 Research Gap** 8](#_Toc85061510)

[**1.4 Research Problem** 9](#_Toc85061511)

[**1.5 OBJECTIVES** 11](#_Toc85061512)

[**Main Objective** 11](#_Toc85061513)

[**Specific objectives** 12](#_Toc85061514)

[**2.** **METHODOLOGY** 13](#_Toc85061515)

[**3.1 System Architecture** 24](#_Toc85061516)

[**3.1.2 Commercialization** 26](#_Toc85061517)

[**3.** **BUDGET AND BUDGET JUSTIFICATION** 28](#_Toc85061518)

[**REFERENCES** 29](#_Toc85061519)

# **List of figures**

Page

[Figure 1.1: Malicious Activity increasing over the given period iv](file:///C:/Users/Acer/Desktop/IT18034400-2021-085.docx#_Toc67492726)

[Figure 1.2: Malicious Activity increasing over the given period 3](file:///C:/Users/Acer/Desktop/IT18034400-2021-085.docx#_Toc67492727)

[Figure 1.3: Percentage of Phishing Attacks hosted on HTTPS 4](#_Toc67492728)

[Figure 1.4: Sudden spike after an event has occurred 6](file:///C:/Users/Acer/Desktop/IT18034400-2021-085.docx#_Toc67492729)

[Figure 3.1: Activity diagram of pre-trained plugin 13](file:///C:/Users/Acer/Desktop/IT18034400-2021-085.docx#_Toc67492730)

[Figure 3.2: High Level Diagram of the system 14](file:///C:/Users/Acer/Desktop/IT18034400-2021-085.docx#_Toc67492731)

[Figure 3.3: High Level Diagram of the pre-trained plugin 15](file:///C:/Users/Acer/Desktop/IT18034400-2021-085.docx#_Toc67492732)

[Figure 4.1: Watcher web application wireframe 21](file:///C:/Users/Acer/Desktop/IT18034400-2021-085.docx#_Toc67492733)

[Figure 4.2: Plugin wireframe: generate report 21](file:///C:/Users/Acer/Desktop/IT18034400-2021-085.docx#_Toc67492733)

# **List of tables**

[**Table 1.1: Comparison of previous work 5**](#_Toc67493390)

**Table 1.2: Comparison of Nomad vs Kubernetes 11**

**Table 1.3:Required technologies, techniques for research study 12**

# **List of abbreviations**

Abbreviation Description

CNN Convolutional Neural Network

CoVID Corona Virus Disease

DL Deep Learning

HTTPS Hyper Text Transfer Protocol Secured

ML Machine Learning

NLP Natural Language Processing

RNN Recurrent Neural Network

URL Uniform Resource Locator

WHO World Health Organisation

WWW Word Wide Web

# **List of appendices**

Appendix Description Page

[**Appendix - A : Plagiarism Report** 27](https://mysliit.sharepoint.com/sites/CDAPSubmissionCloud/2021RegCloud/2021-041-Students/1.%20Project%20Proposal/Report/IT18002102.docx#_Toc67340062)

# **INTRODUCTION**

## **Introduction**

As the technology environment grows, we become more dependable on Information Technology and its services. One such service that we depend on so much is the World Wide Web, which is also known as the Internet. At present a world without the internet cannot be imagined. Almost 4.66 billion people were active internet users as of October 2020 [1]. This technology advancement has caught the eyes of cybercriminals to lure internet users and propagate cyber-attacks.

Cyber-attacks and cybercrime are defined as offensive acts carried out by means of electronic devices with the use of the Internet to steal sensitive information by infiltrating the data systems. Cybercrimes is one of the criminal acts which has been growing rapidly in this century. Along with the evolution of technology cybercrimes have been increasing sophisticatedly. Every recent technology that has be born brings new crimes along with it. Viruses, Malware, Spyware, DDoS attack, Ransomware, Man in the Middle Attack, Phishing are few of the cybercrimes that humankind has faced since 1962.

Cyber attackers use different mechanisms and techniques to conduct malicious exploitations on their victim. As of last decade, the damages caused by cybercrimes has increased from $485 million to over to $4 billion where 60% of small medium enterprises are prone to cyber-attacks [2] [3]Cyber-attackers are opportunistic who lure and exploit both the business as well as individuals with ongoing trending events around the globe.

A campaign targeting businesses could see cybercriminals sending out documents that are made to look like government advice on dealing with Brexit which in fact download malware [4]. From January to April 2020, some 907,000 spam messages, 737 incidents related to malware and 48,000 malicious URLs - all related to COVID-19 were detected [5]​. As of April 2020, roughly 60,000 include COVID-19 related malicious attachments or malicious URLs [6].​

During the COVID-19 pandemic, the entire globe shifted to online platforms from usual traditional platforms. While the number of internet users increased, cybercriminals took this opportunity to lure them for their personal and/or business sensitive information by sending malicious URLs with attracting titles. Internet users or the target victim tend to click on these malicious URL as they look legitimate.

Although there are various other attack mechanisms available to lure the victims, most cybercriminals propagate the attack through malicious URLs using social engineering. Reports and research states that 98% [2] attacks use social engineering as their main attack vector by cybercriminals, which are extremely effective. Thus, humans are considered to be the weakest and vulnerable link in order to carry out a malicious activity.

Graphical user interface, table

Description automatically generated

Uniform Resource Locator (URL) is an address of a resource in World Wide Web (WWW) which helps a user to access websites [7]. An URL consists of three main parts protocol, hostname, and path namely. A cyber attacker modifies this URL for attack purposes. Thus, these modified URLs are called malicious URLs. In order to identify these malicious URLs, researchers and developers started building models and applications with the help of Machine Learning (ML) and Deep Learning (DL). Using Machine Learning & Deep Learning researchers were able to produce various models and detection system for malware detection, network intrusion detection, spam detection and malicious URL detection [8].

We witness that machine learning is explicitly growing in the technological industry in terms of detection systems. Detection systems which use pronunciations, characters, and shapes stay constant over the period [8] , but applications used to detect malicious content changes and evolve over time which makes the detectors unusable. This has become one of the greatest challenges for malicious URL detectors.

Background literature states that research works have been explored in the past to identify and differentiate malicious URLs. In addition, they used machine learning approaches to detect the malicious URLs and give better solutions. These research works were focusing on increasing accuracy and blocking the detected URLs.

However, these models do not incorporate scalability and privacy dominantly. This prevents a system to scale according to the processing power given to it. A malicious URL is identified with a delay due to lack of scalability. Additionally, the existing solution for detecting malicious URLs takes extra time which reduces the speed and efficiency of identifying them [9].

The proposed system for this problem is a web application that takes URL as input from the user and detects, extracts malicious URLs related to the keyword in real-time. To improve the efficiency and accuracy of the results, this research uses ensemble model which incorporates ML & DL. With excess creation of malicious URLs depending on the event that has occurred, the accuracy level is questionable, and time taken to provide the result is high. During this stage, the trained model should be able to detect the number of malicious even if there is an increase in number of links. In order to achieve this, the ensemble model should be able to scale in and out depending on the event given. For this process, TensorFlow, Azure Machine Learning and ML and DL algorithm will be used.

According to APWG 4th Quarter report, throughout the COVID-19 outbreak malicious attacks have become more pervasive and damaging than ever [10]. During the course of COVID-19 pandemic there was a rapid increase in phishing activity, by end of December 2020 the number of phishing sites reached 200,000. (Figure 1.1).



Figure 1.1: Malicious Activity increasing over the given period

Several years ago, to protect humans from malicious websites and URLs, users were requested to check for “HTTPS” and a padlock sign. Recently, the Anti-Phishing Working Group discovered that attackers use HTTPS encryption protocol to fool humans and lure them to gather sensitive information [10].

The figure below represents the percentage of phishing attacks which were hosted using HTTPS over the last couple of years starting from 2017. (Figure 1.2). Most of the attacks have been held in the last quarter of 2020 and it is depicted that over the years, attackers using encryption protocol have increased although there has been decreased values.

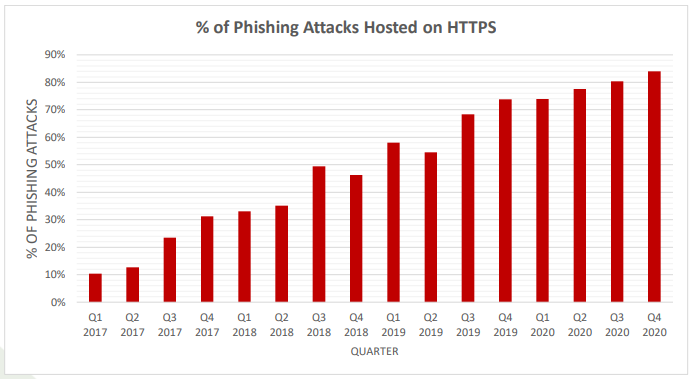


Figure 1.3: Percentage of Phishing Attacks hosted on HTTPS

In the past, various research studies were performed to identify and detect malicious websites and URLs. These research studies have used various detecting methods using ML, DL algorithms like lexical analysis, NLP, CNN, RNN etc. [11] [12] [13].

Most of the research studies either showed ML, DL algorithm used for detecting the malicious URLs. One research study introduced ensemble models to detect malicious URLs [14]. Ensemble model is a process where multiple diverse models are created to predict an outcome, either by using different algorithms or using different training data set.

## **1.2 Background Literature**

There are several research studies and technological methodologies to identify malicious URLs and to improve the output’s degree of accuracy level.

According to Research A [12], malicious URLs were detected using ML lexical analysis methodology. This research study focused on detecting the malicious URLs effectively and efficiently by using a lightweight approach which is lexical analysis.

Secondly, Research B [13], uses seven different classification algorithm and NLP. This research focuses on real-time systems. Using these two-algorithm methodologies the detection system was able to give its best performance and better accuracy level. Thus, it is understandable that this research study has focused on performance and accuracy to detect the malicious URLs and blacklist them.

Research C [14] has used ensemble model to detect malicious, benign URLs for detection process, the proposed system in this research study replaces the simple traditional approach of blacklisting and reduces the false positive rate. This improves the efficiency of the model and helps to detect malicious URLs from benign ones.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Efficiency** | **URLs based on events** | **scalable system** | **Privacy** |
| Research A [12] | **û** | **ü** | **û** | **û** | **û** |
| Research B [13] | **ü** | **û** | **û** | **û** | **û** |
| Research C [14] | **ü** | **û** | **û** | **û** | **û** |
| Proposed system | **ü** | **ü** | **ü** | **ü** | **ü** |

**Table 1.1: Comparison of previous work**

Previous research studies indicate that used only accuracy, efficiency and performance are dominantly focused. This research study proposes to classify the URLs based on a specific event and scale accordingly by maintaining accuracy, efficiency, and the system’s performance in real-time and user’s privacy.

## **1.3 Research Gap**

With the research study conducted in the background and the conclusions arrived in Figure 1.1 it is clearly understandable that the existing applications are not scalable and did not take users’ privacy into consideration. Thus, a new modern application which uses Machine Learning with adaptation of user privacy and system scalability is needed.

Addressing this issue, the system proposed in this research study helps users not only to identify the malicious URLs with feature extraction but also to safeguard user’s privacy on the data collected during the process and the system’s scalability.

From the literature survey on the research studies given in figure 1.1 we can understand a few research studies had common features that could be identified. Considering this is the starting point of this research, a solution was needed to fill the gaps. Although the main idea behind this research study is feature extraction using keywords to extract malicious URLs, it was identified scalability and user’s privacy is also equally important to fill the gap.

Unlike the existing research study solutions, the final deliverable of this system would be a system which protects users from malicious URLs even when there are numerous links being created and data collected from users are protected.

## **1.4 Research Problem**

Cyber-attacks have been leveraging as the world digitally transforms. Numerous malicious URLs are spawned at an instant. Cybercriminals produce diverse ways to make the site look legit as possible. During the pandemic period, the COVID-19 malicious URLs had keywords like covid-19, WHO, vaccine etc. These keywords change with the trending words related to COVID-19 as the global environment changes.

The number of malicious URLs is higher than the registered legitimate URLs. [15] This proves that if any trending event occurs around the globe, there is a high chance that the number of malicious links will increase. In order to detect these numerous links, there should be a scalable model which will scale in and scale out according to the weightage of the event. If a malicious URL is identified with a delay due to lack of scalability, the internet users are prone to cyber-attacks easily.

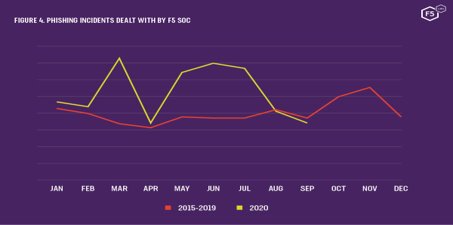
Figure 1.3 depicts a sudden increase in the malicious URL incident for the year 2020. The sudden spike in trend lines was witnessed after an event occurred. A solution for timely detecting malicious URLs after an event is still not proposed. A rapid increase in the number of malicious URLs might slow down the detecting system as they are mostly focused on efficiency and accuracy dominantly.

Figure 1.4: Sudden spike after an event has occurred

There is a need of a system which could detect malicious URLs depending on the event that has occurred and being trending locally and globally. This system model should be able to scale in and out depending on the number of keywords and malicious URLs that could be born. Thus, the research problem for this study would be “A feasible way to deploy scalable model which is comaptible with the ensemble model selected to detect malicious URLs.”

Through the background and literature study it was identified that systems which use to extract malicious URLs using Machine Learning did not focus on the user’s privacy. Data collected from the user’s end should be secured, protected without breaching data privacy laws.

## **1.5 OBJECTIVES**

### **Main Objective**

The main objective of this research study is to identify a suitable ensemble model using deep learning and machine learning algorithm for accurate analysis and faster detection. This ensemble model should be able to scale in and scale out according to the relevant hot events that is happening around the globe. This will include both local and global events.

The research component scalability will be tested on other three components in the research. The three components are,

* Web Application with suitable system architecture
* Keyword token to corelation with malicious URLs database
* Feature reduction

The web application should be able to be deployed in both mobile devices and desktop devices. Our research study is focused on that the web application which should be highly performable even when there are more inputs than usual. This shall make the application to work efficiently and effectively.

When training the Machine learning, Deep learning system, which is known as feeding the data, the system should be able to leverage horizontally. Thus, the time, cost and performance of the system will improve.

Scaling the feature reduction is one of the challengeable tasks in the research. The study will take effort to scale this, however this shall be kept for future research work.

Another main objective in the context of research study is the model or service which will be chosen should be compatible with the ensemble model. Therefore, the scalability algorithm needs to be studied and understand.

### **Specific objectives**

In order to achieve the main objective, the specific objectives that need to be attended are as follow,

* Identify a solution for scalability which can be compatible with Ensemble Model. Compare available solutions for scalability, in order to do this a deep analysis on how each model works need to be studied and their algorithms needs to be compared.
* Integrate the chosen scalable solution (Nomad) to scale the proposed system in this research.
* Contribute towards data privacy laws.
* Improving the scalability of the system using both machine learning and deep learning techniques to enhance the efficiency and reliability of the system.
* Develop a browser extension for the proposed solution, which analyse, detects, and extracts malicious URLs real time.

# **METHODOLOGY**

As the beginning stage of this research study, a comprehensive background study was conducted to understand and confirm the impact that this research study has to the users. By doing so, it was identified the need of a system which could detect malicious URLs related to COVID-19 in real time. This helped to lead the project by setting its goal on identifying malicious URLs based on events that are in trend.

The end product of this system is expected to extract the URLs based on the keywords of trending events and expected to create an application which integrates it. It is also determined that this research project will focus on the scalability and privacy of the proposed solution. This end product will be beneficial for large scale as well as small-medium scale enterprises to protect their organization from data breaches.

As stated in the research gap, the main aim of this proposed model is to enhance the degree of accuracy of the result. With the advancement made in artificial intelligence, the need for an accurate model is justified. The domain adopts innovative approaches every year and previous URL classifiers become less efficient as time goes on. It has been proven that the ensemble model provides high classification accuracy with low false positive.

Figure 3.2 depicts the overall system diagram to identify the malicious URLs. As the system uses ML and DL algorithm methodologies, deploying the ensemble model system in a less complex method. To solve the research problem mentioned above, this research study will containerise the created application and deploy it in a workload orchestrator.

Following are the technologies and techniques that will be used to accomplish the aforementioned process (Table 1.3).

**Table 1.3 Required technologies, techniques for the research study**

|  |  |
| --- | --- |
| Technologies | **Azure Machine Learning**  **Python**  **TensorFlow**  **Nomad**  **HTML**  **CSS**  **JavaScript** |
| Techniques | **Containerisation**  **Orchestration** |

Along with the system created in this research study, a pre-trained plugin will be created for users. Considering the prevailing pandemic situation around the globe, remote working has become the new norm of working in the industry. For ease of use and access to the users this pre-trained plugin will be in handy.

Figure 3.1, The activity diagram gives a basic idea on how the plugin works. The pre-trained plugin should be installed in the users’ web browser. Once the user visits the URL page, the pre-trained plugin sends the user visited URL to the ensemble model which detects whether the URL is malicious or benign.

If the URL is considered to be malicious the plugin, parallelly performs the following tasks:

* Blocks the user from accessing the URL
* Records the event in the database
* Notifies system administrator as an Alert

The research component of this research study focuses on scaling the system and up and down accordingly based on the number of malicious URLs that has been detected, and focuses on the user’s privacy. This component is divided into the following sub-tasks:

**Problem Identification**

**Solution design**

**Background study on Scalability**

**Background analysis on User Privacy**

**Data gathering for malicious URLs**

**Analysing scalability models**

**Design extension**

**Problem Identification**

**Solution Design**

**Background study about scalability**

**Background Analysis on User Privacy**

**Data Gathering**

**Analysing scalability Models**

**Design Extension**

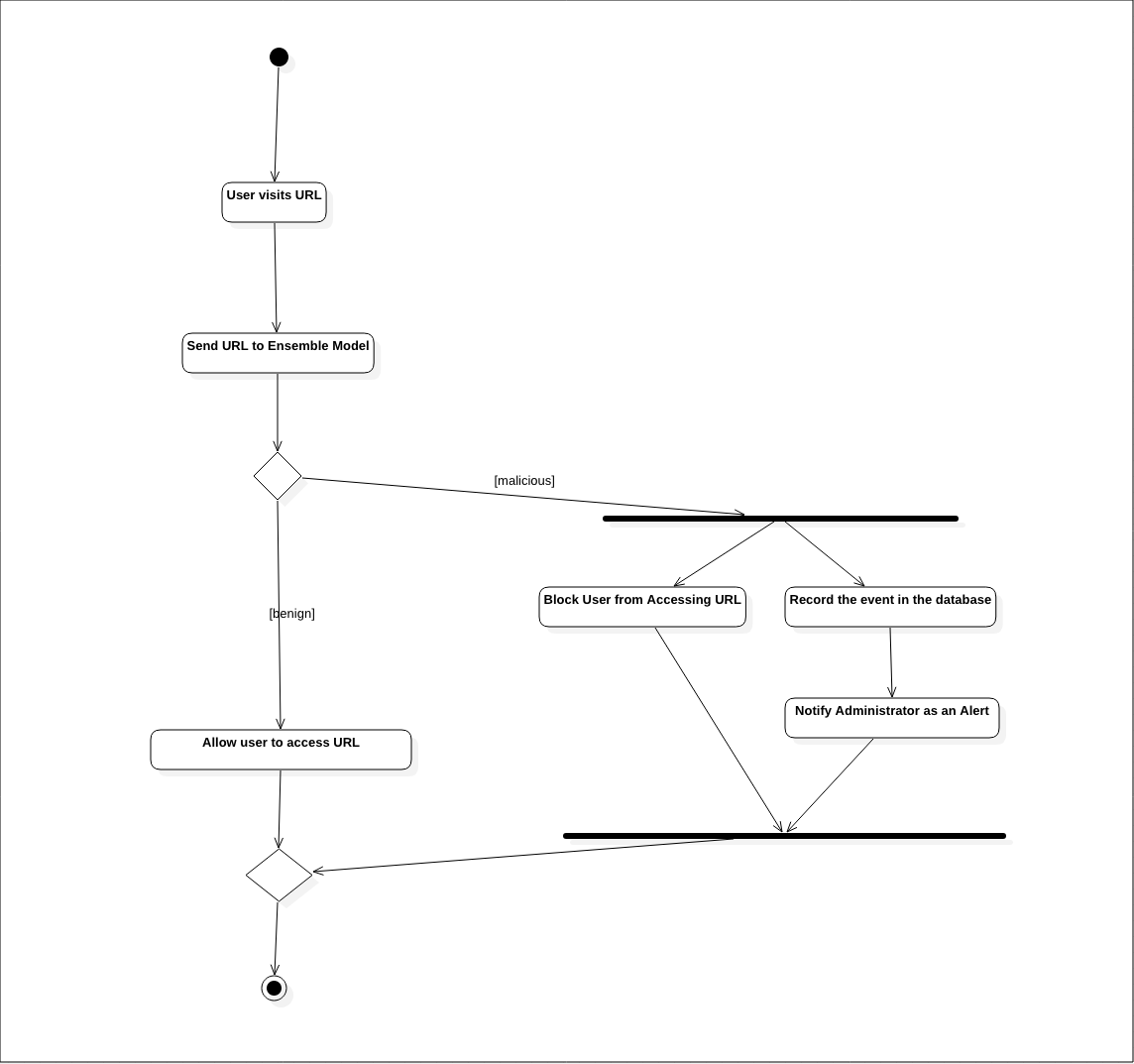


Figure 3.1: Activity diagram of pre-trained plugin

## **3.1 System Architecture**

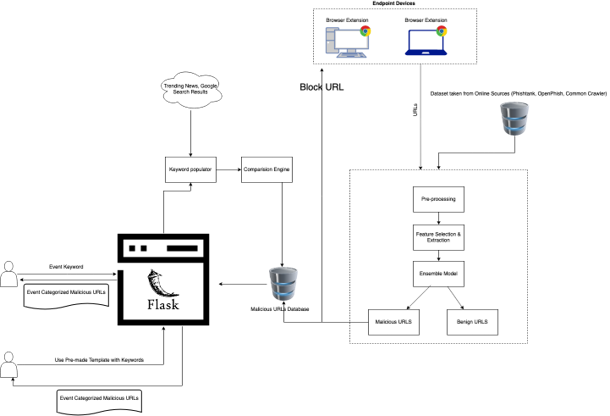


Figure 3.2: High Level Diagram of the system

System overview diagram for the pre-trained plugin

Graphical user interface, application

Description automatically generated

Figure 3.3: High Level Diagram of the pre-trained plugin

### **3.1.2 Commercialization**

Since a part of the userbase for this system would be small-medium enterprises (SME), this system can be commercialized with the note of providing basic security. Even though, at this stage the system cannot compete with state-of-the-art endpoint management system, this will be suitable for SMEs due to the lower cost and higher usability.

Two versions of this system can be implemented.

* A free version that SMEs and Researchers can use to collect URL lists based on events with limited export capability.
* A paid version that will provide seamless export capability in addition to the basic endpoint protector using the browser plugin.

|  |  |
| --- | --- |
| **Free Version** | Rate Limit on event-based malicious URL list and restricted export capabilities. |
| **Paid Version** | Browser plugin to protect users from malicious pages with basic reporting to the administrator  No limit on export and event-based malicious URLs |

The low-cost barrier should be a key aspect to attract SMEs. Researchers/ Investigators investigating certain events can make use of the data to correlate with the incident.

# **BUDGET AND BUDGET JUSTIFICATION**

|  |  |  |
| --- | --- | --- |
| **Task** | **Cost ($)** | **Cost (Rs.)** |
| Azure Machine Learning Studio (Implementation and Cloud Deployment) | 100 (Monthly Fee) | **19,900** |
| Web Application Hosting | 72 | **14300** |
| Database – Mongo DB | 57 | **11400** |
| Plugin upload to Chrome Web store | 5 (One-time fee) | **995** |
|  | Total Cost | **46,595** |

To cover these costs, we can use the pricing models mentioned in the commercialization section.

# **REFERENCES**

|  |  |
| --- | --- |
| [1] | 27 January 2021. [Online]. Available: https://www.statista.com/statistics/617136/digital-population-worldwide/. [Accessed 15 February 2021]. |
| [2] | S. Brathwaite, “Growth of Cybercrime in the last decade,” oppos, 2021. |
| [3] | “Artic Wolf,” [Online]. Available: https://arcticwolf.com/resources/blog/decade-of-cybercrime. [Accessed 8 April 2020]. |
| [4] | “ expected once withdrawal agreement is finalised:,” [Online]. Available: https://www.itpro.co.uk/security/32560/spike-in-brexit-themed-phishing-attacks-expected-once-withdrawal-agreement-is​,. |
| [5] | RTE, “ 'Alarming' cybercrime increase during Covid-19 pandemic, says Interpol.,” [Online]. Available: https://www.rte.ie/news/newslens/2020/0804/1157249-cybercrime/. [Accessed Tuesday August 2020]. |
| [6] | R. Lefferts, “ Microsoft shares new threat intelligence, security guidance during global crisis.,” Microsoft, [Online]. Available: https://www.microsoft.com/security/blog/2020/04/08/microsoft-shares-new-threat-intelligence-security-guidance-durin. [Accessed 8 April 2020]. |
| [7] | M. SIDDHARTHA, “Wisdom ML,” 19 August 2020. [Online]. Available: https://wisdomml.com/detecting-malicious-urls-using-machine-learning-approach/. [Accessed 14th February 2021]. |
| [8] | B. A. Miller, “Scalable Platform for Malicious Content Detection Integrating Machine Learning and Manual Review,” p. 1, 2015. |
| [9] | Z. M. X. L. J. A. J. Z. K. K. .. Basit A., “ A comprehensive survey of AI-enabled phishing attacks detection techniques,” 2020 . |
| [10] | “APWG Trend Report,” 2020. [Online]. Available: https://docs.apwg.org/reports/apwg\_trends\_report\_q4\_2020.pdf. . [Accessed 25 February 2020 ]. |
| [11] | C. G. M. V. P. D. S. Kurt Thomas, “Design and Evaluation of a Real-Time URL Spam Filtering Service,” *IEEE Symposium on Security and Privacy,* 2011. |
| [12] | M. A. R. H. L. N. S. A. A. G. Mohammad Saiful Islam Mamun, “Detecting Malicious URLs Using Lexical Analysis,” *Springer International Publishing,* 2016. |
| [13] | E. B. O. D. B. D. Ozgur Koray Sahingoz, “Machine learning based phishing detection from URLs,” *Expert Systems With Applications,* 2018. |
| [14] | S. T. Lekshmi A R, “The Kozinec-SVM Model for Detecting Malicious URLs,” *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT),* vol. 08, no. 07, 2019. |
| [15] | “APWG,” 2020. [Online]. Available: https://docs.apwg.org/reports/apwg\_trends\_report\_q3\_2020.pdf. |
| [16] |  |