Malicious Uniform Resource Locators Detection using Feature Extraction and Deep Learning

Muhammad Dafa Sirajudin[1], Parman Sukarno[2], and Aulia Arif Wardana[3]

[1] School of Computing, Telkom University, Bandung, Indonesia
dafajudin@student.telkomuniversity.ac.id
[2] School of Computing, Telkom University, Bandung, Indonesia
psukarno@telkomuniversity.ac.id
[3] Wrocław University of Science and Technology, Wrocław, Poland
aulia.wardana@pwr.edu.pl

Abstract. Malicious URLs are a serious challenge in cybersecurity, given the increasing number of threats such as malware, ransomware, spyware, phishing, defacement and trojans. Deep learning has the ability to learn complex patterns in data automatically and effectively, so it can be used to detect anomalies and malicious patterns in URLs. Previous research has proposed various methods to detect malicious URLs, including blacklist-based methods and URL features. However, these methods often lack effectiveness in dealing with evolving attack patterns. In the detection of harmful URLs, according to various studies, applying deep learning has the potential to increase the process's efficiency and accuracy, but there is still an opportunity to further optimize efficiency and accuracy. This paper aims to develop a malicious URL detection system using deep learning based on feature extraction. This method will improve data representation through text analysis and transformation of such data, as well as selection of important features from the dataset. This research utilizes a multiclass dataset that includes categories such as benign, defacement, phishing, and malware. Among the evaluation measures that will be applied in the process of evaluating the model are the following: accuracy, precision, recall, F1-score, and macro average. It is believed that the methodologies used in this study will significantly advance cybersecurity by making harmful URL detection systems more accurate and effective. Among the design tested, the GRU model achieved the highest accuracy at 92.59%.

Keywords: *Malicious URL, Deep Learning, Feature Extraction, Cyber Security, URL Detection*

1 Introduction

Malicious URL is a type of URL designed to perform actions that harm users or computer systems [1]. Cyber criminals usually spread viruses such as malware, ransomeware, phishing, steal personal information, or commit fraud and other cyber attacks [2] [3]. Malicious URLs are often sent to users via email,

text messages, popups, or advertisements with the aim of redirecting them to malicious websites. There has been a meteoric rise in the number of internet sites in recent years, mostly due to the ever-increasing sophistication of computer systems [4] [5]. One of the most common forms of cyberattacks is phishing, a clear example of this attack is the phishing attack in 2022, where more than 255 million phishing attacks occurred in the first six months. In these attacks, attackers send out genuine-looking emails, often posing as well-known institutions or services such as banks, email service providers, or social media companies, to trick users into clicking on malicious links that redirect them to phishing websites [6]. The website is designed to look very similar to the original website so convincingly that unsuspecting users enter their personal information such as usernames, passwords, and other important information. The impact of these attacks is devastating, including identity theft, unauthorized access to personal accounts, loss of sensitive personal data, as well as significant financial losses to both individuals and victimized companies [7]. Given this threat, it is important to understand the mediums often used by attackers. Email is considered the primary means of spreading a wide variety of malicious attacks. As a result, since 2020 until now, malicious URLs and online scams have increased by more than 94% [8].

The use of feature extraction in detecting malicious URLs is essential as it is a key phase in machine learning techniques. Feature extraction involves retrieving important elements from the dataset to effectively train and test the design in detecting malicious URLs [9]. This process helps to identify the most relevant features from the dataset, it can reduce processing time and storage requirements by selecting the most optimal and minimal features. In addition, feature reduction methods are required to assess the suitability of the extracted features and eliminate irrelevant features. Previous research has used various methods to detect malicious URLs, including blacklist-based methods, URL feature analysis, and machine learning. For example, blacklist-based methods use a database of known malicious URLs to block access, while URL feature analysis looks at characteristics such as URL length, specific character usage, and domain structure [10]. However, these methods are often ineffective in the face of evolving and adaptive attack patterns. The main weakness of these methods is their lack of ability to handle new and complex attacks, which are not registered in the database or have different patterns from the previous ones. To overcome this weakness, improvements need to be made that ensure that the classification process runs more efficiently and effectively, thereby improving the design's performance in detecting malicious URLs [11].

One of the main reasons the author conducted this research is to develop a more effective and efficient detection system using deep learning technology. By recognizing the shortcomings that exist in previous methods, this research aims to improve the accuracy and efficiency in detecting malicious URLs. The proposed system will integrate various deep learning techniques to identify new patterns and characteristics of malicious URLs, so as to overcome the limitations of previous methods and provide a better and easier solution to the evolving threats.

The use of deep learning in detecting malicious URLs is highly relevant and effective as it is capable of automatically extracting and learning relevant features from raw data without requiring intensive manual intervention [12]. This technology can handle the growing complexity and variety of cyber threats, enabling rapid adaptation to new threat patterns that are constantly emerging. Various design of deep learning algorithms, include CNN and RNN, and Gated Recurrent Unit (GRU) have their own advantages in recognising complex patterns and handling sequential data [13]. CNN is very effective in extracting features from URL structures, while RNN and GRU are better at recognizing temporal dependencies in data. The combination of CNN and attention mechanism, as used in the Att-BiGRU design, allows the design to utilize sequence information and focus on important features, thus improving performance and accuracy in detecting malicious URLs. In addition, deep learning enables thorough testing and hyperparameter optimization, so that the resulting design is not only accurate but also well-adapted to changes and reliable in the presence of new data [14]. Thus, deep learning offers a modern and efficient approach to detecting malicious URLs by combining the ability to automatically extract features and deal with the complexity and variation in cyberattack data.

2 Related Work

Previous research by Chen et al. on malicious URL detection [15]. This research discusses the various methods used to detect malicious URLs, incorporating approaches that rely on blacklists, URL features, and web page features. Additionally, a novel CNN-based method for detecting harmful URLs in web page images is presented in this study. With a detection accuracy rate of 85.99%, the data demonstrate that the deep learning method produces great results when it comes to harmful URL detection.

Johnson et al. reported in the article titled "Towards Detecting and Classifying Malicious URLs" the efficacy of deep learning design to standard machine learning approaches in this scenario. They tested several design on an ISCX-URL-2016 dataset, including Random Forest, CART, k-NN, and popular deep learning frameworks, such as Fast.ai and KerasTensorFlow [16]. The results showed that Random Forest, KerasTensorFlow, and Fast.ai achieved over 96% accuracy while working with binary and multiclass data. The best design was found to be Random Forest because of its balanced performance, time efficiency, and complexity. This research also shows that combining the top 5-10 features provides optimal results, demonstrating the effectiveness of feature selection in malicious URL detection.

An experimental investigation on malicious URLs was conducted by Roy et al., who utilised artificial neural network (RNN) design, more especially "Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and Gated Recurrent Unit (GRU)". The authors collected a dataset from Kaggle containing harmless and phishing URLs by preprocessing the data and converting the URL