LITERATURE REVIEW

Detecting Phishing Websites through Deep Reinforcement Learning

Phishing is the simplest form of cybercrime with the objective of baiting people into giving away delicate information such as individually recognizable data, banking and credit card details, or even credentials and passwords. This type of simple yet most effective cyber-attack is usually launched through emails, phone calls, or instant messages. The credential or private data stolen are then used to get access to critical records of the victims and can result in extensive fraud and monetary loss. Hence, sending malicious messages to victims is a stepping stone of the phishing procedure.

Aphisherusually setups a deceptive website, where the victims are conned into entering credentials and sensitive information. It is therefore important to detect these types of malicious websites before causing any harmful damages to victims. Inspired by the evolving nature of the phishing websites, this paper introduces a novel approach based on deep reinforcement learning to model and detect malicious URLs. The proposed model is capable of adapting to the dynamic behavior of the phishing websites and thus learn the features associated with phishing website detection.

In recent years, the application of various kinds of machine learning algorithms to the classical classification problem and in particular to security and malware detection has received tremendous attention and interest from research community. Furthermore, with the advancement of computational power, deep learning algorithms have created a new chapter in pattern recognition and artificial intelligence. As a result, many classification, decision, and automation problems are now can be formulated through these sophisticated learning algorithms. Deep learning-based approaches are particularly effective when the number of features involved in the computation is large.The proposed approach is robust, dynamic, and self-adaptive since reinforcement learning-based algorithms can estimate a solution (i.e., action) based on the stochastic state conversions and the rewards for choosing an action for that state.This paper presents a deep reinforcement learning-based model for detecting phishing website by analyzing the given URLs. The model itself is self-adaptive to the changes in theURL structure. The problem of detecting phishing websites isan instance of the classical classification problem.

Therefore,we have developed a reinforcement learning model usingdeep neural network, to solve this classification problem.

We have used our model on a balanced and labelled dataset of legitimate and malicious URLs in which 14 lexical features were extracted from the given URLs to train the model. The performance is measured using precision, recall, accuracy and F-measure. The key contributions of this paper are as follows:

1) Model the identification of phishing websites through Reinforcement Learning (RL), where an agent learns the value function from the given input URL in order to perform the classification task.

2) Map the sequential decision making process for classification using a deep neural network based implementation of Reinforcement Learning.

3) Evaluate the performance of the deep reinforcement learning-based phishing URL classifier and compare its performance with the existing phishing URL classifiers.

The reinforcement learning approach has been utilized to gain proficiency for optimal behavior. This adaptive learning paradigm is defined as the problem of an “agent” to perform an action based on a “trial and error” basis through communications with an unknown “environment” which provides feedback in the form of numerical “rewards”.

A typical URL has two principle parts: (1) Protocol: Specifies the protocol to be used for communication between user and web server, (2) Resource identifier: indicating the IP address or the domain space where the resource is located. A colon and two forward slashes separate the protocol from resource identifier.

There are a certain characteristics of websites that helps in distinguishing between phishing sites from the legitimate ones. Examples of such characteristics include: long URLs, IP address in URLs, and request access to additional URLs in which these characteristics are the indications of being phishing websites. The website features in four groups: (1) Anomaly-based, (2) Address bar-based, (3) HTML and JavaScript-based and (4) Domain-based. We followed the proposed work in to build our set of 14 features.

There are some other deep learning based algorithms that should be examined for the problem stated in this paper such as LSTM. Moreover this classifier can be extended for other

binary classification problems like Web spam detection and presence of malicious bots in the network. RL based approach being more adaptive, the classifier can be extended for mitigating various privacy and security concerns in wearable devices.

Effective phishing website detection based on improved BP neural network and dual feature evaluation

Nowadays, phishing poses a big threat to people's daily network environment. By phishing, attackers obtain the network users private information by inducing them to open illegal websites. Due to the active learning ability and preferable classifying ability for many datasets, BP neural network is an important heuristic machine learning method in phishing websites detection and prevention. However, improper selection of initial parameters, such as the initial weight and threshold, will induce the BP neural network into local minimum and slow learning convergence.

Aiming at these problems, this paper proposes DF.GWO-BPNN, an effective phishing website detection model based on the improved BP neural network and dual feature evaluation mechanism. Under this model, the grey wolf algorithm is firstly used to optimize the BP neural network to reasonably select initial parameters.

Then, the dual feature mechanism is used to evaluate the results of the improved BP neural network. By the dual feature evaluation mechanism, the accuracy of phishing website recognition is improved. Meanwhile, the black and white list is used to improve the efficiency of the proposed model. The DF.GWO-BPNN model is compared with some existing phishing website detection models.

Phishing is a kind of cybercrime behavior in which attackers obtain the network users private information by inducing them to open illegal websites. On available of the stolen private information, phishing attackers can get money and other benefits from network users. With the progress of network technology, phishing attackers can use a variety of techniques to makethe phishing websites look legitimate. Phishing websites are becoming more and more capable of avoiding detection.Nowadays, phishing websites are widely flooded in the daily PC and mobile environments. Meanwhile, the number of phishing websites is growing rapidly, which poses a big threat to the people's online life. The distribution and harm of phishing websites are crossing

the national borders and becoming a global problem. It is urgently needed effective techniques to prevent and detect phishing websites.

Meanwhile, for the purpose of efficiency, the black and white list is used to cache websites that have already been processed. Experimental results on testing many commonly used samples have demonstrated thatour proposed DF.GWO-BPNN model is accurate and strong adaptability.

This framework, we propose DF.GWO-BPNN, an effective phishing website detection model based on the improved BP neural network and dual feature evaluation mechanism. In the proposed model, the GWO (Grey Wolf Optimizer) algorithm is used to overcome the shortages of traditional BP neural network.

By optimizing with the GWO, the local minimum and slow coverage problems of BP neural network are avoided. The new the dual feature mechanism is used to evaluate the results of the improved BP neural network. By the dual feature evaluation mechanism, the accuracy of phishing website recognition is improved.

DF.GWO-BPNN is composed of four modules, the Dynamic Hash Library, the Feature Extraction and Classification module, theDF.GWOBPNN Classifier (the BP neural network optimized by GWO) and the ComprehensiveEvaluation module. Specifically, the Dynamic Hash Library caches the black and white lists. For the purpose of improving efficiency, this module is used to cache websites that have already been processed. The Feature Extraction and Classification module extracts and classifies features from the URL composition.

By this module, URL features are divided into two categories, dominant features and the recessive features. URLs with recessive features are feeded to the DF.GWO-BPNN Classifier for further procession. As the core component of the proposed, theDF.GWO-BPNN Classifiers used to classify the phishing websites. Due to the global search ability, the GWO is used to optimize the BP neural network to reasonably select initial parameters.

On constructing this model, we used the GWO algorithm to overcome the shortages of local minimum and slow learning convergence of the BP neural network. In order to elevate the phishing websites recognition rate, we used the dual feature evaluation mechanism to comprehensively evaluate the extracted features from URLs. Meanwhile, the black and white list

was used to improve the efficiency of the proposed model. The DF.GWO-BPNN model was compared with some existing phishing website detection models.

The experimental results demonstrated that our proposed model was accurate and strong adaptability for detecting phishing websites. Although the GWO optimization algorithm in this model can get the global optimal, but the convergence speed is reduced in the later stages. We have already noticed that extracting more explicit features from URLs and enriching the black and white list can alleviate this problem. So, in the later of our work, more features will be extracted from URLs and more reasonable black and white list will be construct.

URL2Vec: URL Modeling with Character Embedding's for Fast and Accurate Phishing Website Detection

A deep learning-based approach to phishing detection is proposed. Specifically, websites' URLs and the characters in these URLs are mapped to documents and words, respectively, in the context of word2vec-based word embedding learning. Consequently, character embedding can be achieved from a corpus of URLs in an unsupervised manner. Furthermore, we combine character embedding with the structures of URLs to obtain the vector representations of the URLs. In particular, an URL is partitioned into the following five sections:

URL protocol, sub-domain name, domain name, domain suffix, and URL path. To identify the phishing URLs, existing classification algorithms can be used smoothly on the vector representations of the URLs, avoiding laborious work on designing effective features manually and empirically. For evaluations, we collect a large-scale dataset, i.e., 1 Million Phishing Detection Dataset (1M-PD), which has been released for public use.

Phishing detection is a challenging task. The traditionaltechnologies for phishing detection need to access webpages, while phishing webpages usually live for quite a short time. Therefore, it is difficult to collect valid webpages, and extract effective features related to their content. In addition, heuristic based anti-phishing approaches rely on laborious analysis of URLs and webpage content to extract hand-extracted features. However, The rules of designing such features, once mastered by phishers, are easily circumvented. Consequently, these techniques may not be able to deal with emerging phishing URLs, which makes anti-phishing researchers quite passive.

To deal with these challenges, anti-phishing researchers have developed quite a number of solutions. Detected malicious webpages by extracting the lexical and host-based features from URLs. However, it relies on the third-party tools such as WHOIS, DNS, etc., which may increase the time cost and be susceptible to the performance of these tools. The success of machine learning in recent years inspires researchers to apply them on phishing detection tasks. Heuristic-based phishing methods have designed a large number of features from several aspects, such as URL, Hypertext Mark-up Language (HTML) codes, and webpage content. Ideally, automatic learning features that are effective for phishing detection are highly expected, which is the aim of our work.

The main contributions in this work are summarized as follows.

1. We propose a novel framework by making use of embedding representation of characters in URLs to detect phishing webpages. The character embedding achieved by the word2vec model does not rely on any external knowledge, hand-crafted information, or network load, overthrowing the conventional methods.

2. We devise a structure-wise strategy to obtain unified representation for the URLs based on the achieved character embedding. In particular, an URL is divided into five parts, including URL protocol, sub-domain name, domain name, domain suffix, URL path, etc., which contributes to the performance significantly.

3. We release a large-scale and real-world phishing detection dataset, i.e.,1M-PD consisting of 1 million URLs, which can hopefully be used to promote the research on phishing detection applications.

Our phishing detection system consists of three modules as shown in bellow Fig. More specifically, the character embedding learning module achieves the vector representation of characters in URLs. Given the character embedding's, the URL vector representation module obtains the vector representation of URLs by concatenating the average embedding's of the characters in the five parts of the URLs. Finally, the detectormodule trains machine learning algorithms on the vector representations of URLs to classify them into phishing ones or legitimate ones. Subsequently, we introduce the modules in each subsections respectively.

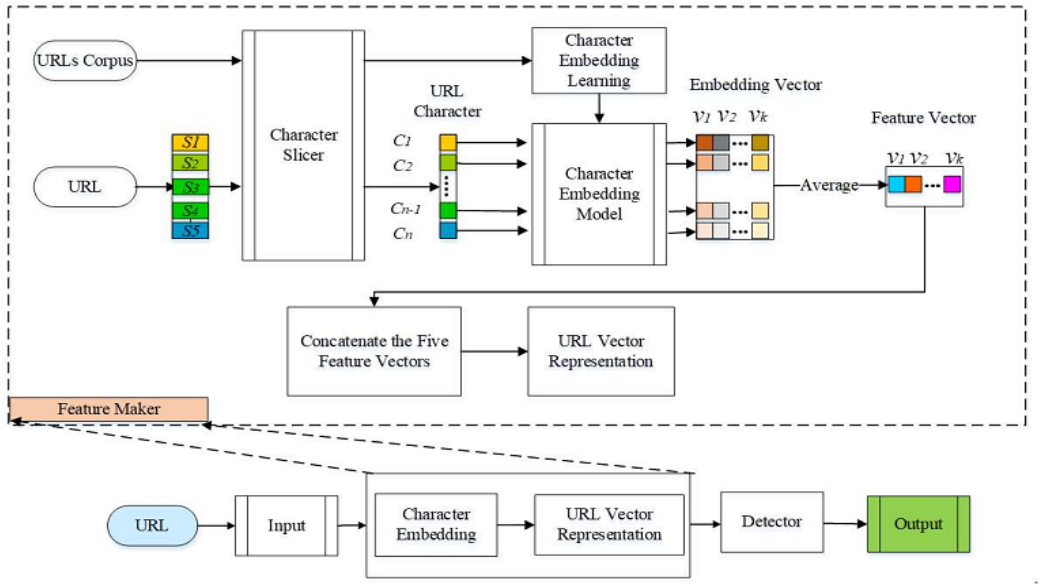


Fig 2.1 Overview of the Phishing framework

Detecting Phishing Websites and Targets Based on URLs and Webpage Links

This work, we propose to extract features from URLs and webpage links to detect phishing websites and their targets. In addition to the basic features of a given URL, such as length, suspicious characters, number of dots, a feature matrix is also constructed from these basic features of the links in the given URL's webpage. Furthermore, certain statistical features are extracted from each column of the feature matrix, such as mean, median, and variance. Lexical features are also extracted from the given URL, the links and content in its webpage, such as title and textual content. A number of machine learning models have been investigated for phishing detection, among which Deep Forest model shows competitive performance, achieving a true positive rate of 98.3% and a false alarm rate of 2.6%. In particular, we design an effective strategy based on search operator via search engines to find the phishing targets, which achieves an accuracy of 93.98%.

Phishing detection is a challenging task, which has attracted a lot of research attention from anti-phishing researchers to seek for effective solutions. The list-based anti-phishing approaches (blacklist or whitelist) store URLs in the database, which is used to match the stored URLs with the URLs entered by users in browsers. These approaches perform quickly but fail to detect newly created phishing URLs as they have not been included in the database. Heuristic-based methods usually extract textual features to detect phishing websites, which can

detect newly created URLs. However, certain textual features extracted from the textual content of webpages cannot be used to detect phishing websites in other languages.

Some researchers propose that similarity-based approaches should be used to compare with the similarity between the given suspicious legitimate webpages under attack, i.e., phishing targets, which should be known in advance. Lately, phishing targets can be detected automatically, unfortunately, its approach is very slow since it needs to obtain and analyze a large number of webpages to form a parasitic community.

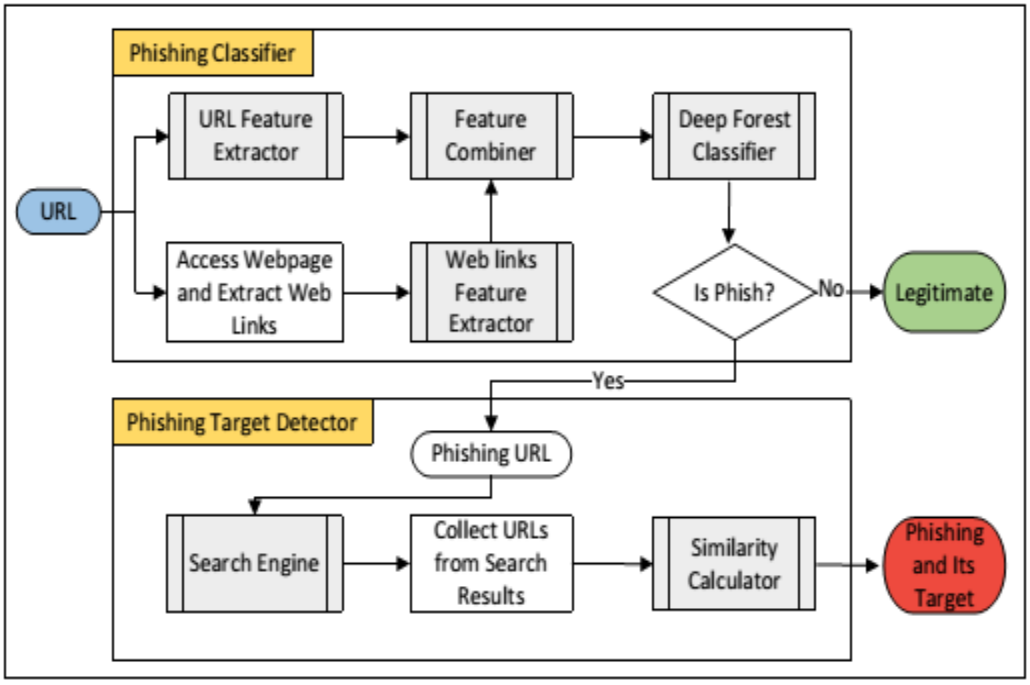
A machine learning based method for phishing detection, which extracts statistical features and lexical features from URLs and the links inside the webpages. Given the URL representation in vector form, Deep Forest and a number of existing machine learning models, such as GBDT and XGBoost, can be applied seamlessly for phishing detection. The proposed approach works regardless of webpages in different languages. The method is efficient and effective, as shown in our experiments. We also propose an effective method based on search operator to detect phishing target, i.e., the legitimate websites under attack.

The main contributions of this paper are summarized as follows:

1. We propose to extract URL features and the statistical features of the links in the webpages to detect phishing webpages, which are effective for phishing detection.

2. We propose a method based on search operator to find the phishing targets of the detected phishing websites, allowing specified matching between query keywords and corresponding sections of the webpages of the phishing target candidates.

3. We investigate and evaluate quite a few machine learning based classification algorithms for phishing detection, among which Deep Forest achieves the highest performance.



Overview of Phishing Websites and Targets Based on URLs and Webpage Links

To combine URL and webpage link features for phishing website detection. The achieved features can be used by various classification algorithms, among which DF shows competitive performance. In particular, we extract features from URLs and the links in the first-level webpages, and do not access the content of the second-level webpages. Therefore, the proposed approach performs quickly in practice and achieves a high accuracy. In addition, we proposed a search operator based method for phishing target detection, which has also achieved a relatively high accuracy.

An Effective Neural Network Phishing Detection Model Based on Optimal Feature Selection

As a common means to obtain user privacy information, phishing poses a big threat to people's daily network environment. The detection and prevention the threats of phishing websites are of importance. Due to the active learning ability from large-scale datasets, neural network is an important heuristic machine learning method in phishing websites detection and prevention. However, during the process of data training, some useless

features may cause the machine learning method to over-fitting which will result in the training model not being able to precisely predict and detect the phishing websites.

Aiming at this problem, based on the optimal feature selection method, this paper proposes an effective neural network detection model (OFS-NN) to detect the pushing websites. Under this model, an optimal feature selection algorithm that adapts to the sensitive features of phishing URLs (Uniform Resource Locators) is firstly proposed. Based on the calculation of the effective value of each feature, this algorithm sets a threshold to eliminate some useless features and selects an optimal feature set suitable for detecting phishing websites. Then, the selected optimal feature set is trained by the neural network to construct an optimal classifier to classify and predict the pushing websites.

Phishing is a kind of cybercrime behavior in which attackers send malicious links through spam and social networks to lure users to obtain private information (user name, account, password, bank card information, etc.). Criminals use the stolen information to get money and other benefits. According to the report of Anti-Phishing Alliance of China, up to the fifth month of 2018, there are 421,070 phishing websites have been identified.

Generally, phishing websites mainly involve in three major industries: the payment transactions, the financial securities and the Ecommerce. Nowadays, phishing websites are widely flooded in the daily PC and mobile environments. Meanwhile, thenumber of phishing websites is growing rapidly, which poses a big threat to the people's online life. Hence, it is urgently needed effective techniques to prevent and detect phishing websites.

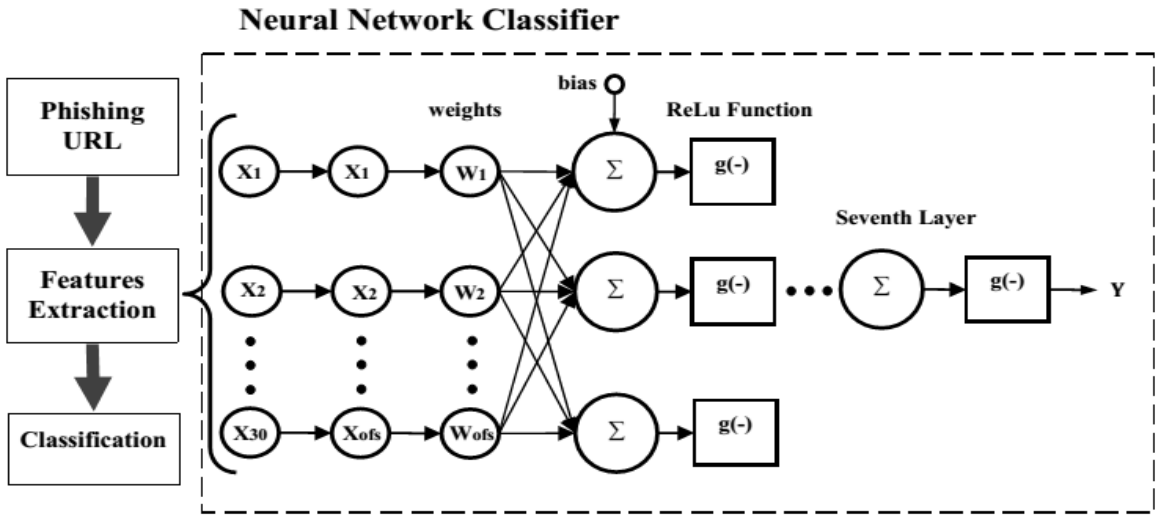
Aiming, this paper proposes OFSNN, an effective neural network phishing website detection model. Under this model, two algorithms, phishing websites feature extraction and optimal feature selection, are firstly employed to obtain the optimal feature set. Then, by feeding the selected optimal feature set, the neural network is trained to obtain the optimal classifier for phishing website prediction and selection.

1) By the feature extraction algorithm, four categories phishing website sensitive features, URL features, script features, security features and statistical features, are extracted from the input URLs. By doing this, we can ensure the high coverage of the feature extraction.

2) By calculating effective values of each feature of the dataset, the optimal feature selection algorithm selects an optimal feature set. During the process of optimal feature

selection, much useless and small impact features are cut off. Since there is no interference from useless features and small impact, the problem of over-fitting is resolved. At the same time, the optimal features selection algorithm brings a certain degree of improvement in performance.

3) OFS-NN takes the neural network as the classifier. By training the selected optimal features, the neural network can effectively predict and select the phishing websites. Due to the powerful learning and fitting abilities of the deep neural network, OFS-NN has the strong ability of generalization.



The detailed structure of the proposed phishing detection model OFS-NN.

Meanwhile, by repeating experimental analysis, we trained the best neural network structure suitable for phishing website detection. Since the OFS-NN model uses the neural network algorithm, it has strong ability of independent learning. The optimal feature selection algorithm can properly deal with problems of big number of phishing sensitivity features and the continuous change of features. This algorithm can reduce the over-fitting problem of the neural network classifier to some extent. Since the sensitivity feature phishing websites are continuous changing, in the future, it is necessary to collect more features to perform optimal feature selection.

Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online

Advancements in information technology often task users with complex and consequential privacy and security decisions. A growing body of research has investigated individuals' choices in the presence of privacy and information security trade-offs, the decision-making hurdles affecting those choices, and ways to mitigate such hurdles.

This article provides a multi-disciplinary assessment of the literature pertaining to privacy and security decision making. It focuses on research on assisting individuals' privacy and security choices with soft paternalistic interventions that nudge users toward more beneficial choices. The article discusses potential benefits of those interventions, highlights their shortcomings, and identifies key ethical, design, and research challenges.

As they go about their online activities, individuals are faced with an increasing number of privacy and security decisions. Those decisions range from configuring visibility on social networking sites to deciding whether to download a mobile app based on the access to sensitive data it requests; from determining whether to trust a website to clicking on—or ignoring—a link in an email.

Such decisions arise in both personal and work contexts: workers who used to rely on workplace system administrators to manage enterprise security often find that they have to configure some security settings on their own, be it in the context of "Bring Your Own Device" workplaces or while interacting with an ever more diverse set of services where system administrators are not available to help.

First, we review research in relevant fields to gain insights into the impact of cognitive and behavioral biases on online security and privacy decision making. Then, we review interventions developed in various fields aimed at helping users make "better" online security and privacy decisions—that is, decisions that minimize adverse outcomes or are less likely to be regretted. We show how this work shares similarities with mechanisms developed to nudge people in a variety of other domains, such as health and retirement planning.

We broadly refer to these efforts as "nudging research," regardless of the originating field of study. We posit that all these efforts can be largely viewed as implementations of soft paternalistic concepts, whereby interventions are intended to gently guide users toward safer practices rather than imposing particular decisions. We suggest that prior work on the design of

user interface technologies for security and privacy can be examined from a nudging perspective: every design decision potentially nudges users in one direction or another.

Our analysis of both the behavioral literature in general and the privacy and security literature in particular has highlighted a vast array of factors that may affect and impair end-users' privacy and security choices. As, the effects of various biases and heuristics on privacy and security choices have already started to be analyzed in a growing body of empirical research. The effect of other biases and heuristics—well known to behavioral researchers, less so to privacy and security specialists—is currently only conjectural, but the hypotheses we present in that section may help drive future research efforts. Furthermore, examples of soft-paternalistic interventions in the field of privacy and security have started to arise both in research and in actual commercial products. As highlighted, we can find growing evidence both of tools aimed at making people reflect on their disclosure or security actions before they take them, and of tools and interface designs that nudge individuals toward more (or more open) disclosures.

In this survey article, our goal has been to document ongoing efforts in this area, discuss some of their limitations, and highlight their potential. We view this new emerging area as one that could lead to the design of a variety of tools and systems that effectively assist humans with online security and privacy decisions without imposing overly prescriptive models of what the "right" decisions might be. The studies we covered in this review have attempted to highlight the human processes that drive privacy and security behaviours, and how those processes can be (and are being) influenced by tools, interfaces, and choice architectures—even when they remain agnostic regarding the appropriateness of such interventions. As noted judging appropriateness is outside the scope of our review—it is, instead, the domain of society's and individuals' autonomous valuations.

In stating that, however, we have also argued that far from seeing nudging interventions as an invasion on individuals' otherwise pristine and untouched autonomy, we should realize that every design decision behind the construction of every online (e.g., software, online social networks, online blogs, mobile devices and applications, etc.) or offline (e.g., conference rooms, vehicles, food menus, etc.) system or tool we use has the potential to influence users' behaviours, regardless of whether the designer, or the user, is fully aware of

those influences and their consequences. In simple terms, there is no such thing as a neutraldesign in privacy, security, or anywhere else. Therefore, we argue for conscious and cautious design of choice architectures and nudges that are inherent to any system, as well as the use of nudging to help users overcome cognitive and behavioral hurdles that may impact their privacy and security decisions.

Detection of Malicious URLs using Machine Learning Techniques

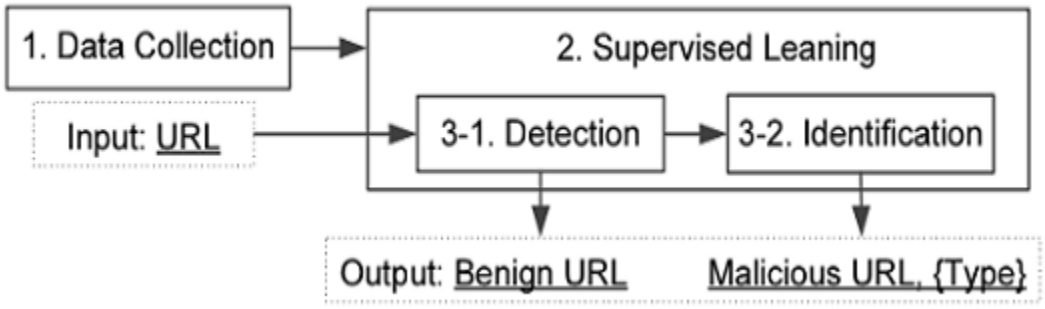
The primitive usage of URL (Uniform Resource Locator) is to use as a Web Address. However, some URLs can also be used to host unsolicited content that can potentially result in cyber attacks. These URLs are called malicious URLs. The inability of the end user system to detect and remove the malicious URLs can put the legitimate user in vulnerable condition. Furthermore, usage of malicious URLs may lead to illegitimate access to the user data by adversary.

The main motive for malicious URL detection is that they provide an attack surface to the adversary. It is vital to counter these activities via some new methodology. In literature, there have been many filtering mechanisms to detect the malicious URLs. Some of them are Black-Listing, Heuristic Classification etc. These traditional mechanisms rely on keyword matching and URL syntax matching. Therefore, these conventional mechanisms cannot effectively deal with the ever evolving technologies and web access techniques. Furthermore, these approaches also fall short in detecting the modern URLs such as short URLs, dark web URLs.

In this work, we used a novel classification method to address the challenges faced by the traditional mechanisms in malicious URL detection. The proposed classification model is built on sophisticated machine learning methods that not only takes care about the syntactical nature of the URL, but also the semantic and lexical meaning of these dynamically changing URLs. The proposed approach is expected to outperform the existing techniques.

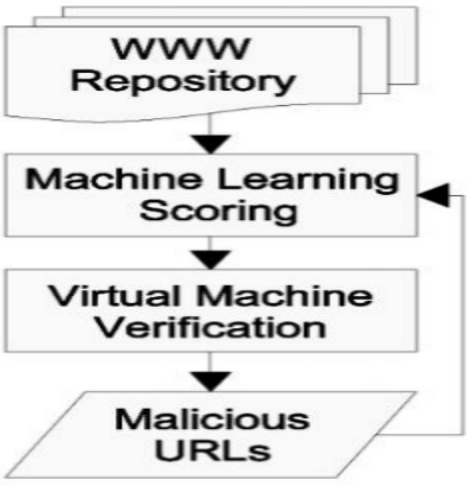
One of the other collaborative work has been initiated by the top tier Internet companies such as Google, Facebook along with many of the startup companies to build a single platform that works all together for one cause of preventing the naive users from the malicious URLs.

To counter these limitations, we propose a novel approach using sophisticatedmachine learning techniques that could be used as a common platform by the Internet users. In this paper, we propose a technique in order to detect the malicious URLs. Various feature sets for the URL detection have also been proposed that can be used with Support Vector Machines (SVM). The feature set is composed of the 18 features, such as token count, average path token, largest path, largest token, etc.



A Framework of Detection of Malicious URLs

The comparison has been made on the various machine learning techniques The detailed view of the results of various techniques has been elaborated in stating that Convolution Neural Networks has shown good performance over the Support Vector Machine algorithm and Logistic Regression algorithm. Compared to the remaining Classification Techniques the Convolution Neural Networks has produced the precision of about 96% over the other two machine learning Techniques.



URL selection and Verification Workflow

Many methods are been proposed to fabricate the Classification Mechanism, Even though we are currently interested in just machine learning techniques, but out of all Convolutional Neural Networks(CNN) provided the better results this is because of the effective learning rate and quite suitable for the feature extraction. To weight the importance of each token, we used the term frequency and inverse document frequency. The term token is the chunk of the URLs. A token can be any part of the URL including the domain and the path.