Exploiting Feature Interactions for Malicious Website Detection with Overhead-accuracy Tradeoff

Shuaiqi Shen[1], Chong Yu[1], Kuan Zhang[1], Song Ci[2]
1.Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Omaha, NE, USA, 68182
2. State Key Laboratory for Safety Control and Simulation of Power Systems and Large Power Generation Equipment, Tsinghua University, Beijing, P. R. China, 100084
Email: sshen@huskers.unl.edu, cyu6@huskers.unl.edu, kuan.zhang@unl.edu, sci@tsinghua.edu.cn

ICC 2021 - IEEE International Conference on Communications | 978-1-7281-7122-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC42927.2021.9500731

***Abstract*—Malicious websites attempt to install malware on user's devices without permission, which can disrupt device operation, steal personal information, and even acquire access to the device for future attacks. Accurate detection of malicious website behaviors is crucial for network security but still faces challenges. Firstly, various types and semantics of website features are required to identity the wide range of malicious characteristics, leading to massive training data and computational overhead. Secondly, to reduce model dimensionality, a proper selection of website features is essential but difficult due to the complex relations among features that can affect each other's contribution to detection outcomes. In this paper, we propose a lightweight feature-based detection scheme against malicious websites considering the interaction measures among features and the overhead-accuracy tradeoff. Specifically, we systematically characterize the interactions among website features in a non-additive manner to indicate the aggregated impacts of feature subsets. Then we propose a quantification method to measure the feature interactions based on multivariate regression. With this method, important features are selected to substantially reduce the model dimension and computational complexity while maintaining desirable accuracy. Meanwhile, the proposed scheme provides an interpretable model that preserves the physical meanings of original features. It allows users to balance the overhead-accuracy tradeoff for detection model training through feature subset selection to fit the requirements and constraints of real applications.**

***Index Terms*—network security, classification, non-additive measure theory, feature selection**

I. Introduction

Internet users are likely to encounter attacks when accidentally connecting to or being redirected to malicious websites, which respond to users' requests while transmitting malware to their devices at the same time. The malware attempts to install and execute programs on users' devices without permission, launching a drive-by download attack. Then the attackers are able to disrupt users' operation, steal personal information, send spams to users and even obtain total access to the devices for future attacks. In addition, malicious websites create phishing pages to collect victim's credential information and imitate victim's identify on other legitimate websites. For example, users of e-healthcare services may encounter phishing attacks that intend to gain access to users' personal health data. According to Verizon [1], in USA around 43% of data breaches on the Internet are based on malicious websites in 2020, and malware. Cisco report [2] pointed out the rapid increment in the number of malicious websites, in which about 20% of malicious domains are relatively new and used around 1 week after they are registered.

As malicious websites play significant roles in cyber crimes, accurate detection becomes crucial to both service providers and users. Widely-used techniques for malicious website detection include blacklists, honeypot and machine learning. The blacklists of malicious URLs are maintained by most commercial browsers and utilized in many applications, such as e-healthcare services [3]. The drawback of blacklist is obvious, as it requires manual maintenance and periodic update. Honeypot [4] is another common security facility deliberately established to be attacked and compromised for detecting malicious activities. Nevertheless, honeypot cannot be adopted to implement the overall security architecture along, since it is a passive security scheme that only detects malicious activities directed against it. To address the drawbacks of blacklists and honeypot, machine learning techniques are applied to develop feature-based detection schemes by collecting website features to train classification models. Features can be extracted from the URL of website, such as lexical features based on the properties of URL string [5], and host-based features including host names, different layers of domain names and path to locate the host server [6]. Content-based features also derive information from the entire web page [7]. To the best of our knowledge, most of the state-of-art malicious websites detection schemes are feature-based.

However, challenges are raised when further improving the effectiveness and efficiency of existing feature-based detection schemes. Firstly, to handle a large number of website features is challenging for capturing comprehensive snapshots of legit and malicious web pages. Various features need to be collected in both static and dynamic types, including URL structure, text content, source code, host-based information, and manifested behaviors when the page is rendered. For devices with limited computational power, such as sensor devices in e-healthcare system, quick responses are expected to determine whether the connections are secure. In this case, a lightweight detection scheme is necessary to detect malicious websites with reduced complexity and computational overhead. Moreover, a large number of features requires massive training data to avoid overfitting when training the classification model, causing greater overhead to data collection. Secondly, to decrease the

978-1-7281-7122-7/21/$31.00 ©2021 IEEE
Authorized licensed use limited to: University of Cincinnati. Downloaded on January 23,2025 at 01:48:42 UTC from IEEE Xplore. Restrictions apply.

complexity of generated model, dimension reduction methods, such as Principle Component Analysis (PCA) [8] and Factor Analysis (FA) [9] are usually applied. Despite the capability of dimension reduction on training data, PCA and FA cannot interpret variables that carry physical meanings in the domain of network security. In addition, the information loss caused by PCA and FA can hardly be quantified after removing partial variance from the original dataset, so that users have difficulties in deciding the proper number of dimensions. Therefore, a lightweight and quantitative scheme is important for malicious website detection to measure the aggregated contributions of features and balance the overhead-accuracy tradeoff.

In this paper, we propose a malicious websites detection scheme that investigates the complex interactions among website features and achieves overhead-accuracy tradeoff for different application requirements. Specifically, the proposed scheme systematically quantifies the aggregated contributions of features on detecting malicious website, so that proper feature selection can be performed to reduce computational overhead while maintaining detection accuracy. The major contributions of this paper are threefold.

- We quantitatively characterize the interactions among website features with non-additive measures to determine their aggregated impacts on detection outcomes. The significances of website features are considered in composite form instead of individual ones, so that the contribution of every possible feature combination is estimated with supervised learning to achieve proper feature selection. Compared to existing dimension reduction methods, the proposed scheme provides a detection model whose parameters can be interpreted with the physical meanings in respect of network security domain. As the original features are preserved as input, the proposed scheme requires less effort in data collection and provides deeper insight on the patterns of malicious websites besides detection outcomes to facilitate further study.
- The proposed scheme balances the tradeoff between detection accuracy and computational overhead by selecting proper subset of website features. After omitting the less significant features and their interference, the classification model is generated based on multivariate regression with composite variables representing all combinations of selected features. The computational overhead for modeling can be substantially reduced while maintaining reliable detection accuracy to fulfill the demands and constraints of different devices and application scenarios.
- We conduct extensive simulations to validate the proposed scheme. Applying different numbers of selected features, the proposed scheme maintains higher detection accuracy compared with other existing feature-based detection schemes. By achieving the same desired accuracy, the proposed scheme requires much less computational overhead. Since new malicious websites keep being created on the Internet, iterative update of the detection model is necessary. The lightweight proposed scheme takes advantage in fast training and providing quick response in real-time network security applications.

The remainder of the paper is organized as follows. We review the related works in Section II. The detection problem is formulated in Section III. In Section IV, the proposed detection scheme is developed based on feature interaction measures. In Section V, performance evaluation is conducted. Finally, conclusions are drawn in Section VI.

II. Related Works

Among the large number of proposed detection schemes, the major variations reside in the different choices of feature representation and machine learning algorithms. Marchal *et al.* [10] defined the term of intra-URL relatedness for detecting phishing URLs and utilized Random Forest to process the collected features. Wang *et al.* [11] extracted fingerprint features from web page texture to measure the similarity of URL binary file contents. The URL word vector and vocabulary features are also merged into the classification model to classify malicious and legit websites based on convolution and recurrent neural networks. Except URL-based features, Altay *et al.* [12] retrieved features from web page context in terms of keyword existence and frequencies, then train the classification model with support vector machine and maximum entropy methods. Decision Tree is also applied to process the features of forwarding-based, URL-related and graph-based features collected from original messages on social media [13]. Other schemes proposed in recent years include applying Extreme Learning Machines for classifying the malicious websites in [14], and the spherical classification scheme that enables batch learning models to handle a large number of instances [15].

Most of the existing machine learning-based schemes perform detection tasks by focusing on deducing the boundaries between malicious and legit websites. However, to obtain high accuracy, most detection schemes require a large number of website features to be collected, increasing the model complexity and the computational overhead exponentially. Furthermore, dimensionality reduction can be quite difficult without the capability to compare the contribution of features in recognizing malicious characteristics. Directly conducting sensitivity analysis on individual features towards the classification result is not applicable either. Similar to cross-layer design in communication systems, different layers of website features, such as URL-based, lexical based and content-based features, are also likely to have complex interrelationships [5]. Changing one feature value may affect the impact of another feature towards the result. Therefore, instead of generating classification model based on individual features, the proposed scheme deduces the model based on feature subsets to achieve optimal feature selection and favorable detection accuracy.

III. Problem Formulation

In this section, we formulate the communication model between client and web server, and identify the typical attacks from malicious websites as well as the collected features

Authorized licensed use limited to: University of Cincinnati. Downloaded on January 23,2025 at 01:48:42 UTC from IEEE Xplore. Restrictions apply.

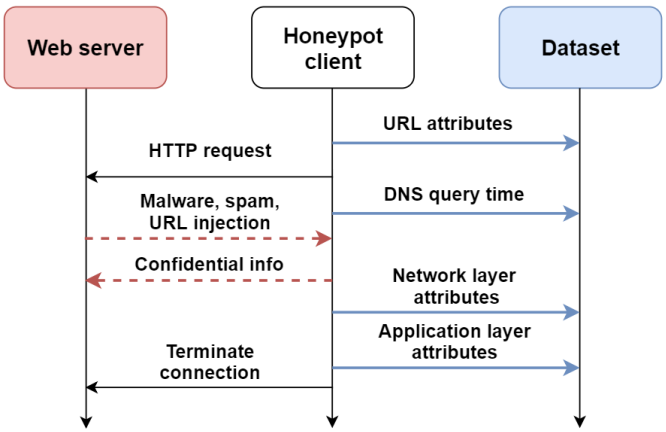

Fig. 1: Communication between web server and honeypot client

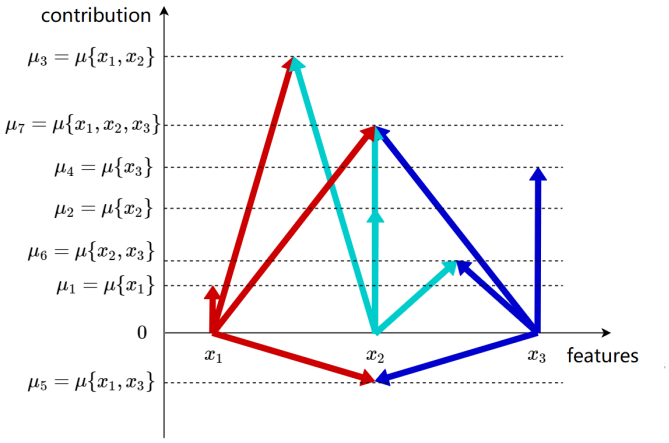

Fig. 2: Illustration of interaction measure among three website features

from the connection. Given the verified sources of legit and malicious website URLs, we apply a low interactive client honeypot to isolate other network traffic. The procedures of communication with web server are illustrated in Fig. 1. Four types of features can be monitored by the client during the communication and recorded in the dataset as a single sample. The retrieved features are effective in distinguish the patterns of corresponding attacks if the visited web server is malicious. For example, *URL injection attack* tends to embed executable code in URL string used on client's browser to redirect the client to other malicious or phishing websites. URL features are able to recognize URL injection by retrieving lexical features within URL string. Malicious websites also tend to install malware on victim's device through drive-by-downloads attack, which results in abnormal increment in the network traffic. Network layer features are able to distinguish drive-by-downloads from normal packet transmission. Other malicious activities, such as spam messages and information theft can also be detected.

Let $X = \{x_1, x_2, \ldots, x_N\}$ be a set of of features collected from a website and y be the label of website as legit or malicious. Let $s : X \rightarrow R$ be the system function of classification, which takes a specific set of features as input and return the malicious probability from 0 to 1 as output. The classification problem can be formulated

$$\begin{aligned} &\min \sum_{i=1}^{Q} (s(x_1, x_2, \ldots, x_N) - y)^2 \\ &\text{subject to} : h_m \geq 0, \ \ where \ m \in \{1, 2, \ldots, M\}. \end{aligned} \tag{1}$$

Here, h_1 to h_M are M constraints that may be posed by network and website limitation, and Q are the number of samples. The objective function computes the square summation of classification error between the predicted malicious probability and real label, so that the optimal solution provides the classification model.

IV. Proposed Malicious Website Detection Scheme

In this section, we present our proposed malicious website detection scheme. Firstly, we define the interaction measures of all feature subsets based on Choquet integral to quantify the aggregated impact of features on detection outcomes. Then, we develop the process of feature selection to obtain tradeoff between detection accuracy and overhead, so that the classification model is generated with multivariate regression.

A. Interaction Measures Among Features

This subsection introduces the properties of feature interaction measures. The relationships among website features, such as URL features, network layer features and application layer features can be quite complex, involving different relations to jointly determine the detection outcomes. The significance of each feature cannot be computed individually, and the probability function can hardly be formulated with traditional linear models. In this case, the interaction measure among features is defined as $\mu : P(X) \rightarrow R$, where $P(X)$ is the power set of X, i.e. the set of all subset of X, and when the input is null set then $\mu(\phi) = 0$. The set function μ is used to quantify the aggregated influences of feature combinations and can be negative to represent negative impacts on objective value. The larger $|\mu|$ implies higher significance of the combination. The interaction measure is also non-additive, so that for two feature sets A and B, $A \subseteq B \subseteq X$ does not necessarily imply $\mu(A) < \mu(B)$. This property is to reflect the fact that sometimes adding more features into the subset decreases their joint influence because features may have opposite impacts to objective value, causing lower detection accuracy even though more features are considered.

The properties of interaction measure are illustrated by Fig. 2, which shows that the full set of three features is not necessarily the most significant combination when determining the malicious probability. If we only consider x_1 and x_2 for prediction, the accuracy is expected to be higher because their combined impact to classification result is the largest among

Authorized licensed use limited to: University of Cincinnati. Downloaded on January 23,2025 at 01:48:42 UTC from IEEE Xplore. Restrictions apply.

all subsets of features. By applying interaction measures, we can generate non-linear function for precise classification and quantitatively evaluation of the importance of various features to a malicious website. The proposed scheme applies Choquet integral to formulate interaction measures and provide nonadditivity to the model. Choquet integral measures the interactions among features with non-additive measure μ defined over power set of features, which can be expressed as

$$z = (C)\int_X f\ d\mu, \tag{2}$$

where $X = \{x_1, x_2, \ldots, x_N\}$ is a set the features, μ is the interaction measure based on subset of X, and f is the tuple of observed values on X that is determined by all of the feature values belonging to the considered subset.

B. Detection with Website Feature Selection

This subsection discusses the process of feature selection and overhead-accuracy tradeoff balancing to obtain the classification model with multivariate regression. Suppose the training data contains N features and Q samples, it can be organized as matrix X as shown in Eqn. 3, where x_{ij} denotes the observed value of sample i and feature j, and y_i denotes website label of sample i. Min-max normalization is firstly applied to the dataset to eliminate the influence of various feature scales on the quantification of feature interactions

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{Q1} & x_{Q2} & \cdots & x_{QN} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}. \tag{3}$$

Given the definition of Choquet integral and normalization of training data, the interaction measures of all feature subsets can be obtained with a multivariate regression model

$$\hat{Y} = e + \int_X f\ d\mu + N(0, \delta^2), \tag{4}$$

where $\hat{Y}$ is the vector of estimated labels; μ is the vector of independent variables of the model representing the significance of feature subsets; e is the regression constant that represents the bias of model when no features is used for classification; $\int_X f\ d\mu$ is Choquet integral computed on set of X with respect to interaction measures μ; $N(0, \delta^2)$ is the normally distributed random perturbation of regression model; and the variance δ^2 is the measure of regression residue error. To obtain Choquet integral, the tuple f needs to be computed for all feature subsets by extending the matrix of training data from $Q-by-N$ matrix X into $Q-by-2^N$ augmented matrix Z. For each row:

$$z_{q0} = 1, z_{qk} = \max\{\min_{n\in\{n|k_n=1\}}\{x_{qn}\} - \max_{n\in\{n|k_n=0\}}\{x_{qn}\}, 0\} \tag{5}$$

where $q = 1, \ldots, Q$ and $k = 1, \ldots 2^N - 1$, k_n is the n^{th} lowest bit of binary representation of k. For example, when $k = 3$, its binary representation is $00\ldots011$, so that $k_1 = k_2 = 1$ and $k_3 = \cdots = k_N = 0$. The values of k can be used to represent different subsets of features by converting k to binary form. In this case, each element of matrix Z can be considered as a composite variable merging all features involved in the corresponding subset. The interaction measure for each subset can be expressed as $\{e, \mu_1, \mu_2, \ldots, \mu_{2^N-1}\}$ so that all possible feature combinations are considered, and the measure for null set μ_0 can be replaced by regression constant e in regression model, which indicates the bias when no features are applied.

By applying Eqn. 5, the augmented matrix Z can be generated based on original data matrix X, which extends its column number to the number of feature subsets. The values of g in Eqn. 4 can be obtained from the entries of matrix Z as the polynomial coefficients, which correspond to the interaction measures of feature subsets, respectively. The regression model of Eqn. 4 can be expressed in matrix form.

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_Q \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1(2^N-1)} \\ 1 & z_{21} & \cdots & z_{2(2^N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{Q1} & \cdots & z_{Q(2^N-1)} \end{bmatrix} * \begin{bmatrix} e \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{2^N-1} \end{bmatrix} + N(0, \delta^2). \tag{6}$$

By minimizing the regression residue δ^2, we can obtain the solution of $\{e, \mu_1, \mu_2, \ldots, \mu_{2^N-1}\}$ as

$$\mu = (Z^T Z)^{-1} Z^T y, \tag{7}$$

so that the interactions of all feature subsets can be quantified, which indicate the contributions of feature subsets on detection outcomes. By selecting subsets with larger interaction, we can adjust the number of features applied to train classification model to reduce modeling overhead while maintaining detection accuracy. The overhead-accuracy tradeoff can be balanced for different computation devices and application scenarios.

V. Performance Evaluation

A. Simulation Settings

In this section, we evaluate the effectiveness and efficiency of the proposed scheme with respect to detection accuracy and computational time per model training. The simulation dataset contains 1781 samples with 216 malicious websites and 1565 legit websites. Website features are collected to reveal malicious and benign characteristics including URL information, network layer and application layer behaviors, which are summarized in Table. I. This full set is applied as the basis for feature selection, which can be expanded in future work to validate the proposed scheme in a higher-dimensional space. Since the dataset is heavily imbalanced, we compute the balanced accuracy and F1-score of detection outcomes as performance metrics. The simulation considers the scenario that in real-time applications, new malicious websites keeps being created on the Internet, requiring iterative update of classification model. To validate the proposed scheme for iterative model training, we shuffle the training samples randomly and split the dataset into 26 episodes. For each episode, new data is added into the training set and use the remaining samples as the testing set to update the classification model and derive

Authorized licensed use limited to: University of Cincinnati. Downloaded on January 23,2025 at 01:48:42 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Description of website features

Index	feature description
1	the number of characters in URL
2	the number of special characters in URL
3	the operative system of the server got from response
4	number of TCP packets exchanged between the server and client
5	number of the ports detected and different to TCP
6	total number of IPs connected to the client
7	the number of bytes transferred
8	the number of packets sent from client to the server
9	the number of packets sent from server to the client
10	the number of DNS packets generated during the communication

TABLE II: Feature subsets ranked with interaction measures

Rank	Feature subset	No. of features	Interaction measure magnitude
1	{3,5-10}	7	519600
2	{2-4,5-8,10}	8	462800
3	{1-10}	10	375400
4	{3,4,6-10}	6	253100
5	{3,6,7,9,10}	5	244700

new balanced accuracy, F1-score and computational time of training. Since at least one episode is required for training and testing, the process is repeated for 25 times, so that the average values and variances of the performance metrics can be obtained.

B. Simulation Results

As presented in Table. II, the relative significances of feature subsets in the classification model are quantified as interaction measures, which indicate the aggregated impacts of involved features on determining whether a website is malicious or not. In some cases, a subset with more features has less interaction than a smaller subset, because the involved features have weaker correlation with each other. To evaluate the effectiveness of feature selection, we compare the accuracy of proposed scheme with two existing detection schemes against malicious websites, which utilize deep reinforcement learning (RL) [16] and extreme learning (EL) technique [17], respectively. All schemes are required to apply the same number of features, for which the proposed scheme selects feasible feature subset with the largest interaction and the benchmark schemes select features from the full set randomly. The random feature selection of each benchmark scheme repeats for 5 times, so that the proposed scheme is processed for 25 model updates (add one episode of data per update) and benchmark schemes are processed for 125 updates. The average accuracy values of all model updates and 95% confidence intervals are shown in Fig. 3. The F1-scores of whole dataset for malicious website detection are shown in Fig. 4. Since the proposed scheme selects website features with the largest interaction on detection outcomes, maximum information carried by training data can be preserved for the generated detection model. The result validates that based on systematic characterization of feature contributions, proper feature selection is capable of reducing computational complexity of model training while maintain the detection accuracy.

Besides balanced accuracy and F1-score, the computational complexity of the proposed scheme is also evaluated. In the

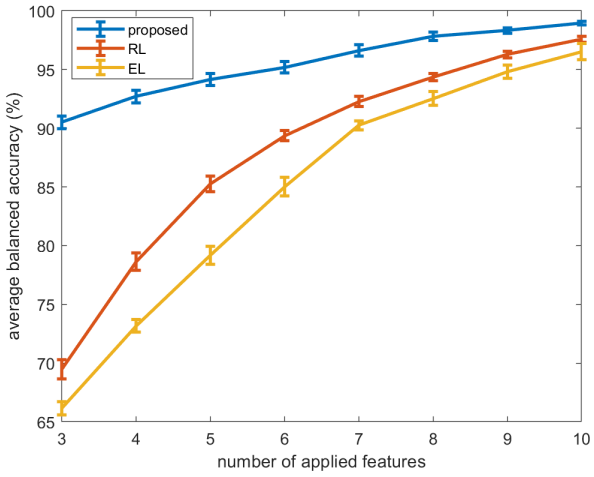

Fig. 3: Average balanced accuracy vs. number of applied features: the proposed scheme maintains high detection accuracy after interaction-based feature selection compared to RL and EL, which select features from dataset randomly. As the required model dimension decreases, the difference in accuracy becomes larger. The accuracy variance of the proposed scheme is also lower than RL and EL, showing more consistent detection performance.

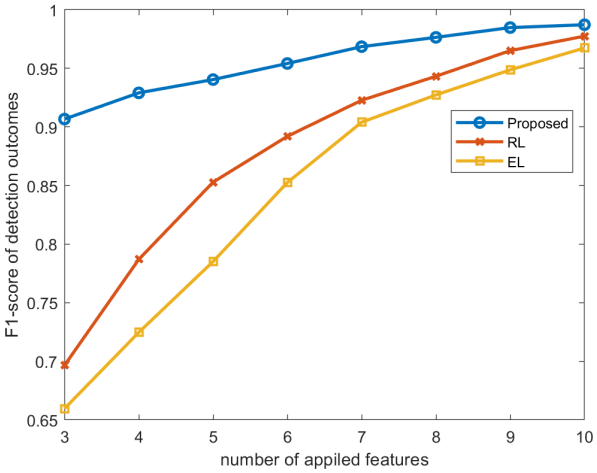

Fig. 4: Average F1-score vs. number of applied features: the proposed scheme achieves the best balance between precision and recall compared to RL and EL, which select features from dataset randomly. As the required model dimension decreases, the difference in F1-score becomes larger and proposed scheme provides the most reliable detection model.

simulation, a new episode of data is added into the training set in each model update to simulate the scenario that new malicious websites keep being established in real networks. The classification model is updated iteratively and the average computational time per training is computed as metric of scheme complexity. Compared with RL and EL, the initial training of the proposed scheme has larger computational overhead to traverse all subsets of features for interaction measures. The number of operations increases exponentially with the number of applied features. However, after feature selection of the initial training, the proposed scheme requires much less overhead for the following model updates, so that the average computational time can decrease to an acceptable level. Fig.

Authorized licensed use limited to: University of Cincinnati. Downloaded on January 23,2025 at 01:48:42 UTC from IEEE Xplore. Restrictions apply.

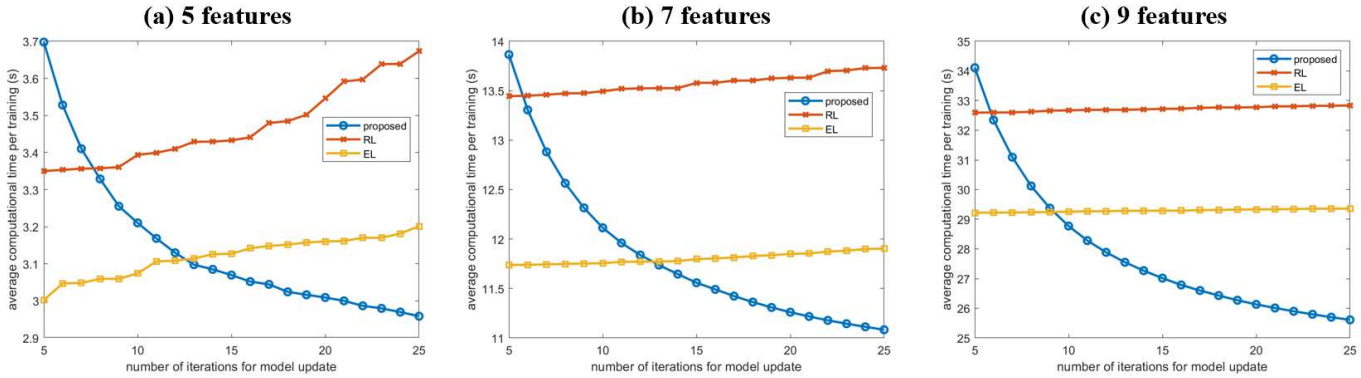

Fig. 5: Average computational time per training vs. number of model updates: the comparison is presented with the number of updates from 5 to 25. Under different cases of feature selection, the overhead of the proposed scheme is high for the initial training, and then converges rapidly to an acceptable level compared to RL and EL. As the number of applied features decreases, the computational overhead of the proposed scheme also drops substantially.

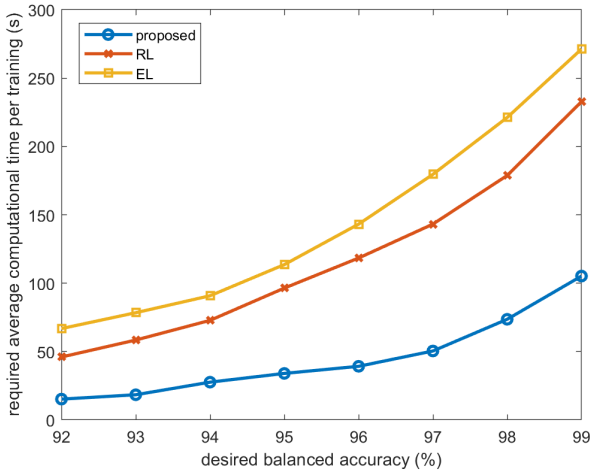

Fig. 6: Average computational time per training vs. achieved balanced accuracy: by setting the evaluated schemes to achieve the same accuracy, the computational time consumed by the proposed scheme is substantially lower than RL and EL.

5 presents the comparison of computational overheads by requiring the proposed and benchmark schemes to apply 5, 7, and 9 features for model training, respectively. In real applications of malicious website detection, emerging websites are created in the network with rapid rate, so that the detection model has to be updated iteratively with the recently collected data to remain adaptive to malicious characteristics. As the number of times to update model increases with new websites emerging, the average computational time of the proposed scheme keeps decreasing and tends to converge to a much lower level than RL and EL. The simulation results efficiency of the proposed scheme in the scenario of real-time malicious website detection, where frequent update is necessary to obtain an adaptive generated model for the latest malicious website patterns utilized by the attackers.

C. Further Discussions

As discussed previously, the operations to traverse all feature subsets for characterizing corresponding interactions do not cause obvious overhead in realistic scenarios that require iterative model update. Then we can validate the reduction of computational complexity of the proposed scheme by making the proposed and benchmark schemes achieve the same accuracy. Fig. 6 shows the average computational time required to achieve different desired accuracies. As desired accuracy increases, the reduction of computational overhead achieved by the proposed scheme becomes more substantial compared to RL and EL. This improvement is essential for devices with limited computational power, such as smartphones and wearable devices, which require lightweight detection scheme to provide timely response.

In addition, the overhead-accuracy curve of the proposed scheme also show that users are capable of balancing the overhead-accuracy tradeoff for malicious website detection. In different cases of model updates, various user demands for detection accuracy and constraints of computing resources can be fulfilled by the proposed scheme. In comparison, balancing the overhead-accuracy tradeoff via proper feature subset selection can be difficult for existing feature-based detection schemes, such as RL and EL, that are integrated with dimension reduction algorithms. The dimension reduction algorithms utilize transformed dimensions for modeling without the original physical meanings of website features. This requires dimension reduction process to be repeated every time emerging website data are collected, or the user demands and resource constraints change. Therefore, the proposed scheme achieves more lightweight malicious website detection while guaranteeing the desired accuracy.

In the future, a promising direction for our work is to extend the proposed non-additive measure based feature selection technique to more high-dimensional scenarios. In realistic network security problems, such as malicious website detection, a large number of features are expected to be collected and considered. Since the total number of feature subsets increases exponentially with the number of features, the characterization of feature interactions in high-dimensional problems can be challenging. We will expand the training dataset in the future work to provides a more complete basis for feature selection and overhead-accuracy tradeoff balancing. To measure the

Authorized licensed use limited to: University of Cincinnati. Downloaded on January 23,2025 at 01:48:42 UTC from IEEE Xplore. Restrictions apply.

sheer volume of feature interactions, more powerful solution algorithm will be investigated to handle the excessive computational overhead and overfitting problems.

VI. Conclusion

In this paper, we have proposed a lightweight detection scheme against malicious website with optimal feature selection based on interaction measures to balance overhead-accuracy tradeoff. Specifically, we have formulated interaction measures among website features based on Choquet integral to bring non-additive property to classification model. Then we have derived the interaction measures by applying multivariate regression to quantify the aggregated impacts of feature combinations on detection outcomes. Through proper feature selection, the overhead for data collection and model training can be reduced substantially while maintaining high detection accuracy. In addition, the systematic and quantitative analysis of feature significance enables users to achieve various overhead-accuracy tradeoffs by adjusting the selected feature subsets to fulfill requirements of different network security applications. As new malicious websites keep being created on the Internet, the training dataset and classification model need to be updated iteratively. In real-time applications, fast response is required for malicious website detection. The relatively low overhead of our proposed scheme takes an advantage in fast training and giving quick response compared to the existing works.

Acknowledgment

This work is supported by the University of Nebraska Foundation.

References

[1] Verizon, "2020 data breach investigations report," 2020. [Online]. Available: https://enterprise.verizon.com/en-gb/resources/reports/dbir/

[2] Cisco, "Cisco 2018 annual cybersecurity report," 2018. [Online]. Available: https://www.cisco.com/c/dam/m/digital/elq-cmcglobal/witb/acr2018/acr2018final.pdf

[3] J. Hong, T. Kim, J. Liu, N. Park, and S. W. Kim, "Phishing url detection with lexical features and blacklisted domains," in *Adaptive Autonomous Secure Cyber Systems*. Springer, 2020, pp. 253–267.

[4] M. Akiyama, T. Yagi, T. Yada, and Y. Kadobayashi, "Analyzing the ecosystem of malicious url redirection through longitudinal observation from honeypots," *Computers & Security*, vol. 69, pp. 155–173, 2017.

[5] X. Yan, Y. Xu, B. Cui, S. Zhang, T. Guo, and C. Li, "Learning url embedding for malicious website detection," *IEEE Transactions on Industrial Informatics*, pp. 6673–6681, 2020.

[6] T. Shibahara, Y. Takata, M. Akiyama, T. Yagi, K. Hato, and M. Murata, "Evasive malicious website detection by leveraging redirection subgraph similarities," *IEICE Transaction on Information and Systems*, vol. 102, no. 3, pp. 430–443, 2019.

[7] S. Ndichu, S. Kim, and S. Ozawa, "Deobfuscation, unpacking, and decoding of obfuscated malicious javascript for machine learning models detection performance improvement," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 3, pp. 184–192, 2020.

[8] S. Priyanga, K. Krithivasan, S. Pravinraj, and S. Shankar, "Detection of cyberattacks in industrial control systems using enhanced principal component analysis and hypergraph based convolution neural network (EPCA-HG-CNN)," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 4394–4404, 2020.

[9] A. M. Sardarabadi and A. J. van der Veen, "Complex factor analysis and extensions," *IEEE Transactions on Signal Processing*, vol. 66, no. 4, pp. 954–967, 2017.

[10] S. Marchal, J. François, R. State, and T. Engel, "Phishstorm: Detecting phishing with streaming analytics," *IEEE Transactions on Network and Service Management*, vol. 11, no. 4, pp. 458–471, 2014.

[11] H. Wang, L. Yu, S. Tian, Y. Peng, and X. Pei, "Bidirectional lstm malicious webpages detection algorithm based on convolutional neural network and independent recurrent neural network," *Applied Intelligence*, vol. 49, no. 8, pp. 3016–3026, 2019.

[12] B. Altay, T. Dokeroglu, and A. Cosar, "Context-sensitive and keyword density-based supervised machine learning techniques for malicious webpage detection," *Soft Computing*, vol. 23, pp. 4177–4191, 2019.

[13] J. Cao, Q. Li, Y. Ji, Y. He, and D. Guo, "Detection of forwarding-based malicious urls in online social networks," *International Journal of Parallel Programming*, vol. 44, no. 1, pp. 163–180, 2016.

[14] L. Lv, W. Wang, Z. Zhang, and X. Liu, "A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine," *Knowledge-Based Systems*, p. 105648, 2020.

[15] A. Astorino, A. Chiarello, M. Gaudioso, and A. Piccolo, "Malicious url detection via spherical classification," *Neural Computing and Applications*, vol. 28, no. 1, pp. 699–705, 2017.

[16] S. Smadi, N. Aslam, and L. Zhang, "Detection of online phishing email using dynamic evolving neural network based on reinforcement learning," *Decision Support Systems*, vol. 107, pp. 88–102, 2018.

[17] Y. Shi, G. Chen, and J. Li, "Malicious domain name detection based on extreme machine learning," *Neural Processing Letters*, vol. 48, no. 3, pp. 1347–1357, 2018.

Authorized licensed use limited to: University of Cincinnati. Downloaded on January 23,2025 at 01:48:42 UTC from IEEE Xplore. Restrictions apply.