Cyber-Phishing Website Detection Using Fuzzy Rule Interpolation

Mohammad Almseidin 1,*,†, Mouhammad Alkasassbeh 2,†, Maen Alzubi 3,† and Jamil Al-Sawwa 4,†

1 Department of Computer Science, Aqaba University of Technology, Aqaba 11191, Jordan
2 Department of Computer Science, Princess Sumaya University for Technology, Amman 11941, Jordan;
   m.alkasassbeh@psut.edu.jo
3 Department of Computer Science, University of Miskolc, 3515 Miskolc, Hungary; alzubi@iit.uni-miskolc.hu
4 Department of Computer Science, Tafila Technical University, Tafila 66110, Jordan; Jalsawwa@ttu.edu.jo
* Correspondence: msoudi@aut.edu.jo
† These authors contributed equally to this work.

Abstract: This paper introduces a novel detection method for phishing website attacks while avoiding the issues associated with the deficiencies of the knowledge-based representation and the binary decision. The suggested detection method was performed using Fuzzy Rule Interpolation (FRI). The FRI reasoning methods added the benefit of enhancing the robustness of fuzzy systems and effectively reducing the system’s complexity. These benefits help the Intrusion Detection System (IDS) to generate more realistic and comprehensive alerts in case of phishing attacks. The proposed method was applied to an open-source benchmark phishing website dataset. The results show that the proposed detection method obtained a 97.58% detection rate and effectively reduced the false alerts. Moreover, it effectively smooths the boundary between normal and phishing attack traffic because of its fuzzy nature. It has the ability to generate the required security alert in case of deficiencies in the knowledge-based representation. In addition, the results obtained from the proposed detection method were compared with other literature results. The results showed that the accuracy rate of this work is competitive with other methods. In addition, the proposed detection method can generate the required anti-phishing alerts even if one of the anti-phishing sparse rules does not cover some input parameters (observations).

Keywords: phishing websites; intrusion detection; fuzzy rule interpolation; intrusion datasets

1. Introduction

The web is now worldwide, and contemporary societies use the web to carry out many activities, including sharing information, social media, and many financial experiments, including buying, selling, promoting, etc. Phishing attacks might be a regular attack on the web. It is a social engineering manner in which customers are lured onto incorrect pages to obtain their personal information or sensitivity. The amount of distinguishing phishing occurrences reported to the organization in the final quarter of 2018 in the Anti-Phishing Working Gather (APWG) survey was 211,032 [1] and increased by 12 percent, resulting in 239,910 reports [2] in the last quarter 2018. Moreover, the Microsoft Security Insights study (Volume 24) subsequently showed that phishing attacks had been a success in 2018, and it projected that they would continue to expand [3]. When phishing attacks are known, the greatest obstacle is the quest for tactics. Phishers are constantly developing their technologies and can build websites that shield themselves from different detection forms. Similarly, defining robust, efficient, and up-to-date methods of phishing exploration is necessary if the phishers are to oppose the adaptive methods used [4,5].

An analysis of anti-phishing strategies literature can be separated into the following approaches: deny-list, heuristic, and fuzzy rules. There are various features and challenges in each answer. The deny-list approach retained a catalog of dubious or malicious URLs.
and compiled with distinctive methods such as Google secure browsing, PhishTank, and user voting. Thus, the browser looks up the deny list as a web page opened to alert the user if the web page has been identified. Finally, users have to save a deny list on a machine or a server [6,7]. Deny lists are used to classify websites as deceptive or legal means. Although these tactics have low false-positive levels, newly generated malicious URLs need to be classified [8,9].

A detailed review of web page material is done in the content-based method. They categorize and eliminate functionality from website contents and third-party administrators, like browsers and DNS servers. Nevertheless, these techniques are ineffectual because they expose the past of the user’s navigation through an immense amount of training and reliance on third parties servers [4]. Based on an evolutionary algorithm, the investigative process uses multiple discriminatory functions to describe the architectures of phished web pages and examine them. The technique used to handle these features plays an essential role in the effective and correct classification of web pages [10]. Since fuzzy logic provides the transitional level of principles, the fuzzy rule-based method distinguishes web pages according to the level of phishing displayed by a specific collection of algorithms [11,12]. The fuzzy approach helps ambiguous considerations to be treated. Fuzzy reasoning coordinates human experts to grasp these variables and their relations. In addition, the fugitive logic uses language considerations to explain phishing characteristics and the probability of the physical website [13,14].

Therefore, developing a detection method against cyber phishing websites is seen as an urgent demand. One of the most effective security solutions for detecting various attacks is the Intrusion Detection System (IDS). However, defining an acceptable detection method against various attacks is not easy. For example, many difficulties are associated with designing and implementing a detection method for IDS, such as boundary issues. The boundary issue appears to be a serious challenge as there are no clear boundaries between normal and attack packets. There is no formal threshold to determine the security status of the packets. The boundary problem appeared clearly by adapting the typical crisp detection mechanisms [15–17]. The fuzzy reasoning methods overcome the boundary problem due to their fuzzy nature and provide more comprehensive alerts.

This work proposes a detection model for cyber-phishing website attacks using fuzzy rule interpolation. The suggested detection model was designed and implemented using the Incircle fuzzy rule interpolation method. Adapting the fuzzy system instead of the binary decision system offers the required extension of the binary decision to the continuous space. Hence, the level of attack could be easily measured. Moreover, the suggested detection model uses a simple method to optimize the fuzzy parameters, and a few anti-phishing sparse rules were used against the cyber-phishing websites' attacks. In this work, we break down the implementation of the proposed detection model into four main steps:

- To identify observable features suitable for detecting the cyber-phishing website’s attacks.
- To identify how the extension of the binary decision to the continuous space handles the intrusion boundaries problem.
- To implement the proposed detection model as a detection mechanism for cyber phishing attacks using fuzzy rule interpolation.
- To compare the results of the proposed detection model with other recent literature results, which had used the same test-bed environment with different algorithms and techniques for detecting cyber phishing website attacks.

This paper is organized as follows: Section 2 introduces the relevant related works to detect phishing website attacks. Section 3 presents the studied Incircle fuzzy rule interpolation method followed by the design and implementation of the suggested detection method in Section 4. Section 5 introduces the accomplished results compared with other literature results. Finally, Section 6 concludes the paper.
2. Related Works

This section presents some of the recent relevant works on developing a detection model against phishing websites. It provides a technical discussion about the shortcomings of the current anti-phishing methods and how the suggested detection handles these shortcomings.

In [18], K. L. Chiew et al. use a new logo icon to establish the website’s identity through the coordination of real and fake websites. The method suggested consists of two steps, logo mining and clarification of identification. The first step will describe the suitable logo using machine learning algorithms. Although Google’s search engine uses the image look in the second phase to retrieve the fake identity, it will validate it. As the logo and weblink relation is one kind, the domain is called the logo uniqueness. Thus we can detect phishing and actual web pages by contrasting Google’s web address and the one detected from the website inquiry. The test results reveal that the malware detection accuracy increased during the logo extraction process and is higher than in exploration phases based on semantic properties. Two disparate datasets of 1140 phishing from PhishTank and legal web pages from Alexa were used to test this framework. The framework focused on 23 traits, picking the most responsive 8. It will take time to access all the functions.

From another perspective, a couple of experiments have merged a heuristic and machine-learning algorithm to boost the classification method of web pages by Solanki, J et al. in [19]. A clarifier and a convincing algorithm are used to provide a precise design to discriminate between phishing and trusted websites. They suggested a heuristic exploration technique for the phishing industry, which was used to identify the malware site. The model deletes and uses URL-based characteristics from the start. These attributes are used with machine learning algorithms at that stage, and whether the website is phished or legitimate, it will be realized. Ten functions on the input URL dataset were included in the framework. The findings are graded as normal or abnormal. The supporting Vector Machine algorithm is used for the results of the remote characteristics, and the value is determined for TN, FN, FP, FN, and TP. The F1 calculation value and the exact value of 96% were also determined. PhishTank and the Yahoo catalog contain 200 URLs from real and phishing websites. The URLs in this dataset are compiled.

Lee, J.L et al. in [20] have again introduced a heuristic-based phishing detection procedure in addition to a machine-learning algorithm using URLs. The suggested approach has revoked the URL of the client has requested web pages. Five computer teaching methods are used to pick a classifying system with the most feasibility of using URL-based features: Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbor, Decision Tree, Random Tree, and Random Forest. Data were collected from PhishTank (3000 phishing sites) and DMOZ (3000 valid web pages) to test and train a classifier. Twenty-six features based on URLs are omitted and included. The test findings show that the leading Random Forest (RF) classifier is 98.23% correct.

Abdelhamid et al. proposed a heuristic way for phishing URLs to be isolated from URLs in [10]. They are evaluated using datasets of more than 16,000 phishes and 31,000 malware-free URLs. They are also evaluated. A collection of 138 features were used to identify URLs for phishing. Features are clasped into four groups: lexical, keyword-based, reputational, and search engine-based features. In addition, Support Vector Machines (SVM with RBF part), SVM with linear kernel, Multi-Layer Perceptron (MLP), Random Forest, Logistic Regression (LR), and C-4.5 were executed in seven separate classifications. Random Forest obtained higher accuracy and a lower error rate according to experimental data.

In the analysis of the work of Al-Taha, A in [5], the author suggested merging two algorithms, K-Nearest Neighbors (KNN) and the Supportive Vector Machine algorithm. The implementation of KNN and SVM as the classification method was the first used. The dataset used for the test was taken from previous studies. It includes over 1353 samples from multiple sources. Each sample record consists of nine elements and the website’s phishing, genuine, or suspect class, respectively. The clarity of KNN is further synchronized with the adequacy of SVM, irrespective of their drawbacks. The specificity of the technique
presented is 90.04%. It also implies that Naïve Bays should be coupled with the help of the vector machine by means of a swift, efficient malware investigating the potential. For the recognition of websites, NB has been used. SVM is used to reclassify the websites so long as the websites are not remembered and are still suspect. The data collection for the learning of the students is provided by PhishTank, which includes 600 websites, 400 legitimate websites, 100 legitimate web pages, and the rest as a dataset. The test results show that the method proposed has been carried out with a high detection precision and low recognition time.

Discussion

The previous works have made convincing contributions and, at the same time, supported the urgent need to implement a suitable detection method against phishing website attacks. From another perspective, the previous detection methods still have common gaps summarized as follows:

- Some of the previous detection methods adapt the binary decision to detect phishing website attacks. However, the binary decision is considered a severe problem for detecting abnormal traffic because there is no formal threshold to determine if the traffic is normal or abnormal.
- The level of attack could not be determined according to the binary decision. Therefore, the anti-phishing alerts need to be more readable and understandable.
- Some of the previous detection methods recorded large values of false alerts, which could lead to time and resource consumption.
- The fuzzy-based detection methods of the previous works required tedious data preprocessing to obtain the anti-phishing rules.
- The fuzzy-based detection methods of the previous works could not handle the issue associated with the deficiencies of the knowledge-based representation.

In response to the previous issues, this paper introduces a detection method against phishing website attacks using fuzzy rule interpolation. The aim behind using the fuzzy rule interpolation reasoning method is summarized as follows:

- The strength of the fuzzy rule interpolation is derived from the fuzzy concept and interpolation technique.
- Using the fuzzy rule interpolation, the issue related to the deficiencies of the knowledge-based representation could be handled by the interpolation techniques.
- The fuzzy rule interpolation effectively smooths the boundaries between normal and abnormal traffic. Therefore, there are no binary decisions.
- Adapting the fuzzy rule interpolation instead of the binary decision system offers the required extension of the binary decision to the continuous space. Hence, the level of attack could be easily measured.

3. Incircle Fuzzy Rule Interpolation

Alzubi and Kovacs proposed a new fuzzy interpolative reasoning method called “Incircle-FRI”, which is defined for a Convex and Normal Fuzzy set (CNF) [21]. The proposed Incircle-FRI follows the geometrical considerations for performing fuzzy interpolation. It considers producing Convex and CNF for all rules and observations. The Incircle-FRI method was constructed based on the triangular fuzzy number. The aim behind using the triangular fuzzy number is twofold: the fact that it is widely used in the fuzzy rule base system and the simplicity of the triangular fuzzy number. Assuming that the triangular fuzzy set $A = (a_1, a_2, a_3; H)$, presents a triangle with the coordinates of vertices $A = (a_1, 0)$, $B = (a_2, H)$, and $C = (a_3, 0)$ with $H$ (as shown in Figure 1). In the case of the $H$ value equal to 1, then the fuzzy number is normal.
The interpolation procedure within the Incircle-FRI method requires two preliminary steps:

- The first step is to determine the main notations of the Incircle-FRI triangular fuzzy number for the fuzzy rules and the observation’s fuzzy sets.
- The second step is to calculate the interpolation consequent of the Incircle-FRI method in the fuzzy triangular set.

The main notations of the Incircle-FRI triangular fuzzy number were used to determine the Gergonne Point (GP) as a reference point of the fuzzy set (Equation (1)). The main sides of the triangular are titled SD1, SD2, and SD3 as shown in Figure 1 for Incircle Notations. In addition, the tangents length and vertices of the triangle with its Incircle titled PS1, PS2, and PS3, which present the “fuzziness sides” (Equation (2)).

Assuming that there are two flanking disjointed fuzzy rules, $A_1 \rightarrow B_1$ and $A_2 \rightarrow B_2$ with the observation $A^\ast$ found in between, as shown in Figure 2. Therefore, the consequent fuzzy set $B^\ast$ is expected to be found in between the consequent of the flanking rules. The Incircle-FRI method consistently produces triangular CNF fuzzy conclusions by holding the same rate of weights among the observation and the two rule antecedents, the conclusion

---

**Figure 1. Triangular Fuzzy Number Notations.**

The interpolation procedure within the Incircle-FRI method requires two preliminary steps:

- The first step is to determine the main notations of the Incircle-FRI triangular fuzzy number for the fuzzy rules and the observation’s fuzzy sets.
- The second step is to calculate the interpolation consequent of the Incircle-FRI method in the fuzzy triangular set.

The main notations of the Incircle-FRI triangular fuzzy number were used to determine the Gergonne Point (GP) as a reference point of the fuzzy set (Equation (1)). The main sides of the triangular are titled SD1, SD2, and SD3 as shown in Figure 1 for Incircle Notations. In addition, the tangents length and vertices of the triangle with its Incircle titled PS1, PS2, and PS3, which present the “fuzziness sides” (Equation (2)).

Assuming that there are two flanking disjointed fuzzy rules, $A_1 \rightarrow B_1$ and $A_2 \rightarrow B_2$ with the observation $A^\ast$ found in between, as shown in Figure 2. Therefore, the consequent fuzzy set $B^\ast$ is expected to be found in between the consequent of the flanking rules. The Incircle-FRI method consistently produces triangular CNF fuzzy conclusions by holding the same rate of weights among the observation and the two rule antecedents, the conclusion
and the two corresponding rules consequent with the reference points, and with the “fuzziness sides”.

Algorithm 1 illustrates the Incircle fuzzy rule interpolation method process.

**Algorithm 1 FRI-Incircle Algorithm**

1. Compute the main notations (GP<sub>x</sub>, PS1, and PS3) of the fuzzy rule-based and observation by (Equations (1) and (2)).
2. Determine the adjacent rules to observation.
3. Calculate the weights between the adjacent triangular fuzzy rules, Rule<sub>i</sub>, and observation A* by:
   \[ W_i = 1 - \frac{|GP_xA^* - GP_xA_i|}{GP_xA_2 - GP_xA_1} \]  
   \[ (3) \]
4. Calculate the reference point (GP<sub>x</sub>) of the fuzzy interpolative reasoning result B* by:
   \[ GP_xB^* = \sum_{i=1}^{2} (W_i \times GP_xB_i) \]  
   \[ (4) \]
5. Calculate the fuzziness sides of the B* fuzzy set, which is based on fuzziness sides of the fuzzy adjacent rule bases and observation by:
   \[ PS_M(B^*) = \left\{ \begin{array}{ll}
   PS_M(A^*) \times \sum_{i=1}^{2} (W_i \times \frac{PS_M(B_i)}{PS_M(A_i)}), & \text{if } \exists PS_M(A_i) > 0 \\
   PS_M(A^*), & \text{if } \forall PS_M(A_i) = 0
   \end{array} \right\} \]  
   \[ (5) \]
6. Finally, the fuzzy interpolative reasoning conclusion B* for the triangular fuzzy set can be collected as:
   \[ B_1^* = GP_xB^* - B((PS1))^* \]
   \[ B_2^* = GP_xB^* \]
   \[ B_3^* = GP_xB^* + B((PS3))^* \]  
   \[ (6) \]

As an extension of the Incircle-FRI, it could be applied with trapezoidal and hexagonal membership functions, in which the trapezoidal fuzzy set can be represented by two triangular fuzzy numbers AL = (a1, a2, Mp; H) and AR = (Mp, a3, a4; H), where Mp presents the mid-point of the trapezoidal fuzzy set. Additionally, the hexagonal can be represented by two triangular fuzzy numbers AL = (a1, a2, a3) and AR = (a4, a5, a6) as shown in Figure 3.

The aim behind using the Incircle-FRI method instead of other reasoning methods is summarized as follows:

- The Incircle-FRI method is characterized by its low computational time and complexity which offers the required quick response for real-time applications such as intrusion detection.
- The Incircle-FRI method follows the fuzzy rule interpolation conditions.
- The Incircle-FRI method supports different types of membership functions for both antecedents and consequent parts, and that includes triangular, trapezoidal, singleton, and hexagonal shaped.
- The Incircle-FRI method facilitates similarity propagation of the fuzziness and core between the observation and conclusion.
4. The Suggested FRI-Incircle Detection Method

In this section, the complete architecture of the proposed FRI-Incircle detection method is presented in detail, besides the main functions and the requirements of the implementation.

4.1. Phishing Websites Environment Preparation

To provide a suitable and realistic environment for testing and evaluating the proposed FRI-Incircle approach, the well-known benchmark phishing dataset from the University of California Irvine (UCI) repository [22] is used. We examine the following periods between January 2015 and May 2015 and between May 2017 and June 2017 to extract the necessary data from phishing websites. Exactly 5000 phishing websites were extracted based on PhishTank2 (https://www.phishtank.com/, accessed on 7 March 2022) and OpenPhish3 (https://www.openphish.com/, accessed on 7 March 2022). In addition, we used another 5000 legal websites based on URLs from Alexa (https://www.alexa.com/, accessed on 7 March 2022) and the Common Crawl5 (http://commoncrawl.org/, accessed on 7 March 2022) archive [7,14,23].

The complete dataset can be downloaded from [24]. The extracted phishing website dataset contains many input parameters (48), and in general, not all of them are relevant to studying the behavior of phishing attacks. Therefore, we used the top five input parameters generated by the latest phishing website detection methods in [14,23,25]. These input features are:

- **PctExtHyperlinks**: it counts the percentage of external hyperlinks in web page HTML source code.
• **PctExtResourceUrls**: it counts the percentage of external resource URLs in web page HTML source code.

• **NumNumericChars**: it counts the number of numeric characters in the webpage URL.

• **PctExtNullSelfRedirectHyperlinksRT**: it counts the percentage of hyperlinks in HTML source code that uses different domain names, starts with #, or using “JavaScript ::void(0)”. Apply rules and thresholds to generate the value.

• **PctNullSelfRedirectHyperlinks**: it counts the percentage of hyperlinks fields containing an empty value, a self-redirect value such as #, the URL of the current webpage, or some abnormal value such as “file://E:/”.

These five input parameters were used as input parameters of the FRI-Incircle detection model. The general structure of the proposed detection model is shown in Figure 4 starting with collecting the website data from the internet and URL database. These data are passed directly to the pre-processing phase to extract the required top five parameters and handle any issues related to the missing data.

Subsequently, the data are passed to the FRI-Incircle model to examine recent URL traffic and look for details about phishing sites. The FRI-Incircle model generated more understandable results due to its fuzzy nature. The results are more readable for the administrator. Furthermore, the strength of the proposed detection model results from the combination of the two concepts, fuzzy and interpolation. The proposed detection model was thus able to smooth the boundaries between legitimate and phishing websites. In addition, it could interpolate the conclusion in the case of the missing rule base.

![Figure 4. FRI-Incircle Detection Method Architecture.](image)

In general, with classical reasoning methods, unless certain situations are clearly defined using fuzzy rules, the consequences cannot be derived. Therefore, the inference of the consequences of the fuzzy system requires a complete fuzzy rule base. There is a lack of expert knowledge in the area of network security [16,26,27]. Therefore, we used fuzzy rule interpolation instead of classical reasoning methods in this work. Fuzzy rule interpolation overcame the lack of knowledge-based handling of different reasoning algorithms and techniques. In case of sparse fuzzy rules which were not covered in all possible situations, FRI methods offer the capability to generate the possible inference, even in case of a lack of definitions and information of existing knowledge representation, as shown in Figure 5. This benefit could be beneficial in partial heuristically solved applications. Suppose that there is an observation x (see Figure 5) that is not explicitly defined in a fuzzy rule base. The FRI reasoning methods could interpolate the results. This advantage can be advantageous in partially heuristically solved applications [26,28–31]. Equation (7) shows the union of the antecedent part of fuzzy rules in the case of sparse rule, supp refers to support. The support of a fuzzy set indicates the set of all elements within the universe of discourse whose degree of membership is greater than zero. Further, $X_i$ is the $i$th input universe of discourse, $A_{ik}$ is the $k$th set of the partition of $X_i$ [32].

$$\bigcup_{i=1}^{k} \text{supp}(A_i) \subset X \rightarrow \bigcup_{i=1}^{k} \text{supp}(A_i) = 0$$ (7)
The implementation of the FRI-Incircle detection method is divided into two phases as follows:

- **Design and optimization phase**: During this phase, the following requirements should be met. Define the expected inputs and outputs as fuzzy sets. Determine and optimize each membership function for each input/output. In addition, extract the necessary anti-phishing fuzzy rules.

- **Testing and Validating Phase**: In this phase, the suggested FRI-Incircle method would be tested and evaluated based on various performance metrics in order to determine the efficiency of the suggested method. Additionally, the proposed anti-phishing detection method would be compared to other anti-phishing detection approaches that employed the same testbed environment.

To prevent phishing attacks, the general structure of anti-phishing schemes shown in Figure 6 should be considered [13]. These anti-phishing schemes start from the bottom up with phishing detection, user training, and phishing prevention. The proposed FRI-Incircle method considered the anti-phishing schemes where it can effectively prevent phishing sites and provide more understandable and readable anti-phishing fuzzy rules. In addition, the output of the suggested detection method could be more readable for the administrator to take the required actions to prevent phishing websites.

**Figure 5. Sparse Fuzzy Rule.**

**Figure 6. Anti-Phishing Schemes.**

4.2. **Design and Optimization Phase**

At the early stage of designing the FRI-Incircle detection method, the imported dataset was used to design and optimize the general structure of the detection method. The phishing website dataset includes a large number of records, and it contains a large number of input parameters (48). In general, not all of them are relevant to studying phishing
attacks’ behavior. Therefore, we used the top 5 input parameters generated by the latest phishing website detection methods in [14,23,25]. For simplicity, 10,000 records were used to design and optimize the suggested detection method. The proposed structure of the FRI-Incircle method is introduced by defining the required input parameters. These parameters were chosen as input parameters. The other parameters were skipped and considered irrelevant to detect the phishing website attack.

The proposed detection method was designed and constructed using the Sparse Fuzzy Model Identification (SFMI) introduced by Johanyak in [33]. The FRI inference engine is designed to detect phishing websites by generating an alert reflecting the dangerous level of the attack. Three main components control the dangerous level of the attack: the top five input parameters, anti-phishing fuzzy rules, and the interpolation technique of the suggested detection model. The design of the proposed method focuses on two primary factors: first, it uses the necessary advantages of the interpolation technique to reduce the complexity of the fuzzy system. This advantage could generate a system with a small number of fuzzy rules. Second, introduce a detection method to generate the required alerts for each expected observation. The Rule Base Extension using Set Interpolation (RBE-SI) method, which Johanyak and Kovacs introduced in [34] used. It is worth mentioning that the RBE-SI method and other spare rules extraction methods can be downloaded from [35].

Consequently, for the following input parameters: PctExtHyperlinks, PctExtResourceUrls, and NumNumericChars there are four linguistic terms to show their values within the imported phishing dataset. In addition, for the following input parameters: (PctNullSelfRedirectHyperlinks, and PctExtNullSelfRedirectHyperlinksRT) there are three linguistic terms to show their values within the imported phishing dataset. Table 1 shows the linguistic terms that are used for each input parameter.

| Index | Input Parameters                  | Linguistic Terms       |
|-------|----------------------------------|------------------------|
| 1     | PctExtHyperlinks                 | vLow, Low, Medium, High|
| 2     | PctExtResourceUrls               | vLow, Low, Medium, High|
| 3     | NumNumericChars                  | vLow, Low, Medium, High|
| 4     | PctExtNullSelfRedirectHyperlinksRT| Low, Medium, High       |
| 5     | PctNullSelfRedirectHyperlinks    | Low, Medium, High       |

The trapezoidal membership functions were chosen during the design and optimization phase to represent the extracted fuzzy set parameters. The reasoning engine of the suggested detection method was performed by the Incircle FRI method [21]. The output response of the proposed FRI-Incircle method was derived from the interval between 0 and 1. This interval could lead to defining the attack’s diversity. Figure 7 illustrates the optimized antecedent partitions of the suggested FRI-Incircle method.

The inference engine of the proposed detection method offered the anti-phishing warnings even in the case of some input parameters that were not directly defined in the fuzzy rules. The extracted fuzzy rules were sparse. Therefore, these fuzzy rules were designed and extracted for interpolation techniques. However, in the case of classical reasoning methods, these sparse rules could not be implemented, and no results could be obtained. A total of twenty-two anti-phishing sparse rules were generated to detect phishing website attacks. Table 2 presents a sample of the optimized anti-phishing sparse rules.
5. Experiments and Results

This section introduces the second phase of designing the FRI-Incircle method, which is the testing and validating phase. The general structure of the testing and validating process is presented in Figure 8. The suggested FRI-Incircle method was tested and evaluated according to two schemes as follows:

- Testing the optimized fuzzy sets and anti-phishing sparse rules against the testing data, which is already labeled.
- Investigating the interpolation technique by testing the observations that were not covered by the sparse rules.

The evaluation and validation processes of the anti-phishing sparse rules and the fuzzy sets parameters were performed by calculating the false-positive and false-negative rate parameters. These lists of parameters and other performance metrics in Table 3 allow the comparison of the proposed detection method to other methods used in the same testbed environment. It is worth mentioning that the preliminary response of the proposed method was compared along with the actual values of benign and phishing attacks to extract the performance metric parameters. Table 3 shows the obtained performance metric parameters.

Table 3. The Performance Metrics of Suggested FRI-Incircle Method.

| Num | Performance Metrics            | Value   |
|-----|--------------------------------|---------|
| 1   | Sensitivity                    | 0.9780  |
| 2   | Specificity                    | 0.9732  |
| 3   | Precision                      | 0.9780  |
| 4   | Negative Predictive Value      | 0.9732  |
| 5   | False Positive Rate            | 0.0268  |
| 6   | False Negative Rate            | 0.0220  |
| 7   | Computation Time               | 180 min |
| 8   | Accuracy                       | 0.9758  |
The fuzzy nature of the proposed detection method, along with the interpolation technique, facilitates smoothing the boundaries between benign and malicious traffic. The binary decision is considered a critical issue in the case of intrusion detection, which could lead to a decrease in the efficiency of the detection methods [15,17,36]. Furthermore, the suggested detection method effectively reduces false alerts, leading to time and resource consumption. Figure 9 illustrates The Receiver Operating Characteristic (ROC) curve obtained during the experiments.

The experiments showed that the FRI-Incircle method showed 97.58% as the detection rate. In addition, the performance metric parameters conclude that the proposed method effectively reduces false-positive warnings.

From another perspective, there are several methods that were implemented against phishing website attacks as discussed in Section 2. Nevertheless, these detection methods were implemented either based on a binary decision or they required tedious data preprocessing to extract the required detection rules. Even the fuzzy-based anti-phishing methods of the previous detection methods required complete fuzzy rules to generate the required
anti-phishing alerts. However, the deficiencies of the knowledge-based representation appeared as a constant challenge.

To summarize the accomplished results, the performance of the proposed detection model achieved satisfactory values and, at the same time, supports the idea that implementing the fuzzy rule interpolation could be a promising method against phishing website attacks. In addition, the obtained results of the proposed detection method were compared with other literature results. Table 4 presents the comparison accuracy results with other literature detection methods.

Table 4. FRI-Incircle Method vs. Other Algorithms.

| Comparison | Algorithm       | Acc. |
|------------|-----------------|------|
| [7] (2019) | SVM             | 0.92 |
| [7] (2019) | Naive Bayes     | 0.90 |
| [7] (2019) | KNN+SVM         | 0.90 |
| [6] (2019) | Multilayer Perceptron | 0.91 |
| [6] (2019) | J48             | 0.95 |
| [6] (2019) | Random Forest   | 0.95 |
| [34] (2022)| BaggedBFTree    | 0.96 |
| [34] (2022)| BaggedNBTre     | 0.95 |
| [34] (2022)| BoostedBFTree   | 0.97 |
| [35] (2022)| Functional Tree | 0.96 |
| The Proposed Method | Incircle-FRI | 0.97 |

It could be concluded that the suggested detection method is competitive with other detection methods. In addition, the suggested detection method is characterized by its fuzzy nature and interpolation technique. Therefore, the issue related to the binary decision and smoothing the boundaries between benign and phishing traffic was covered by extending the binary decision to the continuous space. The suggested detection method could be a suitable method to detect phishing websites attacks because the following key points characterize it:

- The suggested method was designed based on the strength of fuzzy nature and interpolation techniques.
- The fuzzy nature effectively facilitates the boundaries between benign and phishing attack traffic.
- The output response of the suggested method could be more readable and understandable.
- The output response of the suggested method could lead to defining the attack’s diversity.
- The interpolation technique was able to generate the results (phishing warnings) even in the case of incomplete fuzzy rules.

6. Conclusions

This paper has introduced a novel method to detect phishing website attacks based on fuzzy rule interpolation. The proposed method was designed and optimized using the Rule Base Extension using Set Interpolation (RBE-SI) method. The well-known benchmark phishing attacks dataset was used to evaluate the proposed method. Twenty-two anti-phishing sparse rules were generated to detect phishing website attacks. The results show that the proposed method obtained an acceptable detection rate and effectively reduced the false alerts. Because of its fuzzy nature, the output response of the suggested method is more readable and understandable. The strength of the proposed detection method is derived from its fuzzy nature and interpolation technique. These benefits facilitate smoothing the boundaries between benign and malicious traffic. Moreover, it helps generate more realistic and comprehensive alerts in case of phishing attacks. Furthermore, it handles the deficiencies in knowledge-based representation. In addition, the proposed detection method was compared with other literature results. The results showed that the accuracy rate of this work is competitive with other methods.
For future works, more phishing attack datasets may be included to justify using the fuzzy rule interpolation and the method for generating the input and rules that are generally effective and applicable to a broad spectrum of use cases. From another perspective, other optimization methods could be used to extract the required sparse rules.

Author Contributions: Conceptualization, M.A. (Mohammad Almseidin) and M.A. (Mouhammad Alkasassbeh); methodology, M.A. (Mohammad Almseidin) and J.A.-S.; software, M.A. (Maen Alzubi) and M.A. (Mouhammad Alkasassbeh); validation, M.A. (Mohammad Almseidin) and M.A. (Maen Alzubi). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. APWG. Phishing Activity Trends Report: 4th Quarter 2016. In Proceedings of the Anti-Phishing Working Group: APWG Symposium on Electronic Crime Research, Phoenix, AZ, USA, 25–27 April 2017; Volume 12.
2. APWG. Phishing Activity Trends Report 1 Quarter. In Proceedings of the Anti-Phishing Working Group: APWG Symposium on Electronic Crime Research, San Diego, CA, USA, 15–17 May 2018; pp. 1–12.
3. Cavit, D.; MM, J.S.; Arroyo, J.; Faulhaber, J.; Pecelj, D.; Seifert, C.; Gullotto, V.; Penta, A.; Simorjay, F.; Wu, S.; et al. Microsoft Security Intelligence Report; Microsoft: Redmond, WA, USA, 2010; Volume 10.
4. Shirazi, H.; Bezawada, B.; Ray, I. “Know Thy Domain Name” Unbiased Phishing Detection Using Domain Name Based Features. In Proceedings of the Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies, Indianapolis, IN, USA, 13–15 June 2018; pp. 69–75.
5. Obeidat, I.; Hamadneh, N.; Alkasassbeh, M.; Almseidin, M.; AlZubi, M. Intensive pre-processing of kdd cup 99 for network intrusion classification using machine learning techniques. Int. J. Interact. Mob. Technol. 2019, 13, 70–83. [CrossRef]
6. Altaher, A. Phishing websites classification using hybrid svm and knn approach. Int. J. Adv. Comput. Sci. Appl. 2019, 8, 90–95. [CrossRef]
7. Zuraiq, A.; Mohammad, A.; Al-Kasassbeh, M.; Alnidami, N. Phishing detection based on machine learning and feature selection methods. Int. J. Interact. Mob. Technol. 2019, 13, 171.
8. Chen, Y.S.; Yu, Y.H.; Liu, H.S.; Wang, P.C. Detect phishing by checking content consistency. In Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014), Redwood City, CA, USA, 13–15 August 2014; pp. 109–119.
9. Alkasassbeh, M.; Al-Naymat, G.; Hassanat, A.; Almseidin, M. Detecting distributed denial of service attacks using data mining techniques. Int. J. Adv. Comput. Sci. Appl. 2016, 7, 436–445. [CrossRef]
10. Abdelhamid, N.; Ayesh, A.; Thabtah, F. Phishing detection based associative classification data mining. Expert Syst. Appl. 2014, 41, 5948–5959. [CrossRef]
11. Kumar, K.M.; Alekhy, K. Detecting phishing websites using fuzzy logic. Int. J. Adv. Res. Comput. Eng. Technol. (IJARCE) 2016, 5, 2413–2417.
12. Almseidin, M.; Alzubi, M.; Kovacs, S.; Alkasassbeh, M. Evaluation of machine learning algorithms for intrusion detection system. In Proceedings of the 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 14–16 September 2017; pp. 000277–000282.
13. Aburrous, M.; Hossain, M.A.; Dahal, K.; Thabtah, F. Intelligent phishing detection system for e-banking using fuzzy data mining. Expert Syst. Appl. 2010, 37, 7913–7921. [CrossRef]
14. Abuzuraiq, A.; Alkasassbeh, M.; Almseidin, M. Intelligent Methods for Accurately Detecting Phishing Websites. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 7–9 April 2020; pp. 85–90.
15. Almseidin, M.; Alkasassbeh, M.; Kovacs, S. Fuzzy rule interpolation and snmp-mib for emerging network abnormality. arXiv 2018, arXiv:1811.08954.
16. Almseidin, M.; Piller, I.; Al-Kasassbeh, M.; Kovacs, S. Fuzzy automaton as a detection mechanism for the multi-step attack. Int. J. Adv. Sci. Eng. Inf. Technol. 2019, 9, 575–586. [CrossRef]
17. Almseidin, M.; Kovacs, S. Intrusion detection mechanism using fuzzy rule interpolation. arXiv 2019, arXiv:1904.08790.
18. Chiew, K.L.; Chang, E.H.; Tiong, W.K. Utilisation of website logo for phishing detection. Comput. Secur. 2015, 54, 16–26. [CrossRef]
19. Solanki, J.; Vaishnav, R.G. Website phishing detection using heuristic based approach. In Proceedings of the Third International Conference on Advances in Computing, Electronics and Electrical Technology, Kuala Lumpur, Malaysia, 11–12 April 2015; pp. 87–92.

20. Lee, J.L.; Kim, D.H.; Chang-Hoon, L. Heuristic-based approach for phishing site detection using url features. In Proceedings of the Third International Conference on Advances in Computing, Electronics and Electrical Technology-CEET, Kuala Lumpur, Malaysia, 11–12 April 2015; pp. 131–135.

21. Alzubi, M.; Kovacs, S. Interpolative fuzzy reasoning method based on the incircle of a generalized triangular fuzzy number. *J. Intell. Fuzzy Syst.* 2020, 39, 709–729. [CrossRef]

22. Mohammad, R.; Thabtah, F.A.; McCluskey, T. Phishing Websites Dataset. 2015. Available online: http://eprints.hud.ac.uk/id/eprint/24330/ [accessed on 20 March 2022].

23. Chiew, K.L.; Tan, C.L.; Wong, K.; Yong, K.S.; Tiong, W.K. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. *Inf. Sci.* 2019, 484, 153–166. [CrossRef]

24. Tan, C.L. Phishing Dataset for Machine Learning: Feature Evaluation. Published: 24 March 2018, Version 1. Available online: https://data.mendeley.com/datasets/h3cgnj8hft/1 [accessed on 20 March 2022].

25. He, M.; Horng, S.J.; Fan, P.; Khan, M.K.; Run, R.S.; Lai, J.L.; Chen, R.J.; Sutanto, A. An efficient phishing webpage detector. *Expert Syst. Appl.* 2011, 38, 12018–12027. [CrossRef]

26. Al-Kasassbeh, M.; Almseidin, M.; Alrfou, K.; Kovacs, S. Detection of IoT-botnet attacks using fuzzy rule interpolation. *J. Intell. Fuzzy Syst.* 2020, 39, 421–431. [CrossRef]

27. Alzubi, M.; Almseidin, M.; Lone, M.A.; Kovacs, S. Fuzzy Rule Interpolation Toolbox for the GNU Open-Source OCTAVE. In Proceedings of the 2019 17th International Conference on Emerging eLearning Technologies and Applications (ICETA), Stary Smokovec, Slovakia, 21–22 November 2019; pp. 16–22.

28. Almseidin, M.; Al-Sawwa, J.; Alkasassbeh, M. Anomaly-based Intrusion Detection System Using Fuzzy Logic. In Proceedings of the 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 14–15 July 2021; pp. 290–295. [CrossRef]

29. Almseidin, M.; Al-Kasassbeh, M.; Kovacs, S. Detecting slow port scan using fuzzy rule interpolation. In Proceedings of the 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, 9–11 October 2019; pp. 1–6.

30. Altarawneh, G.A.; Hassanat, A.B.; Tarawneh, A.S.; Carfì, D.; Almuhaimeed, A. Fuzzy Win-Win: A Novel Approach to Quantify Win-Win Using Fuzzy Logic. *Mathematics* 2022, 10, 884. [CrossRef]

31. Hassanat, A.B.; Tarawneh, A.S.; Abed, S.S.; Altarawneh, G.A.; Alrashidi, M.; Alghamdi, M. RDPVR: Random Data Partitioning with Voting Rule for Machine Learning from Class-Imbalanced Datasets. *Electronics* 2022, 11, 228. [CrossRef]

32. Kovács, S. Fuzzy Rule Interpolation. In *Encyclopedia of Artificial Intelligence*; IGI Global: Hershey, PA, USA 2009; pp. 728–733.

33. Johanyak, Z.C. Sparse fuzzy model identification matlab toolox-rulemaker toolbox. In Proceedings of the 2008 IEEE International Conference on Computational Cybernetics, Stara Lesná, Slovakia, 27–29 November 2008; pp. 69–74.

34. Johanyak, Z.C.; Kovacs, S. Sparse Fuzzy System Generation by Rule Base Extension. In Proceedings of the 2007 11th International Conference on Intelligent Engineering Systems, Budapest, Hungary, 29 June–2 July 2007; pp. 99–104. [CrossRef]

35. Johanyak, Z.; Tikk, D.; Kovacs, S.; Wong, K.W. Fuzzy Rule Interpolation Matlab Toolbox—FRI Toolbox. In Proceedings of the 2006 IEEE International Conference on Fuzzy Systems, Vancouver, BC, Canada, 16–21 July 2006; pp. 351–357. [CrossRef]

36. Hassanat, A.B.; Ali, H.N.; Tarawneh, A.S.; Alrashidi, M.; Alghamdi, M.; Altarawneh, G.A.; Abbadi, M.A. Magnetic Force Classifier: A Novel Method for Big Data Classification. *IEEE Access* 2022, 10, 12592–12606. [CrossRef]