AFRL-AFOSR-JP-TR-2024-0021

Neuro-Symbolic Integration for Detecting Phishing Attacks

SUNG-BAE CHO
YONSEI UNIVERSITY UNIVERSITY-INDUSTRY FOUNDATION
50 YONSEI-RO, SEODAEMUN-GU
SEOUL, SEOUL, 03722
KOR

12/12/2023
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Asian Office of Aerospace Research and Development
Unit 45002, APO AP 96338-5002

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
20231212	Final	START DATE 20210930	END DATE 20230929

4. TITLE AND SUBTITLE
Neuro-Symbolic Integration for Detecting Phishing Attacks

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	FA2386-21-1-4085	61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Sung-Bae Cho

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
YONSEI UNIVERSITY UNIVERSITY-INDUSTRY FOUNDATION
50 YONSEI-RO, SEODAEMUN-GU
SEOUL, SEOUL 03722
KOR

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AOARD UNIT 45002 APO AP 96338-5002	AFRL/AFOSR IOA	AFRL-AFOSR-JP-TR-2024-0021

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Phishing attack is a type of social engineering attack often used to steal user's data, including login credential and credit card numbers. Unsuspecting users visit a website via compromised URL and become victims of devastating loss, unaware of what they are doing. Considering the fatality of phishing attacks that are emphasized by many organizations, the inductive learning approach using reported malicious URLs has been verified in the field of deep learning.
- The deep learning-based methods, mainly focused on the fitting of a classification task via historical URL observations, have a limitation of recall due to the extremely diverse
set of phishing URLs. In the field of the information security, an experienced engineer can cope with the phishing attacks by leveraging the stereotype. To detect phishing attacks, an approach that utilizes expert knowledge is promising.
- Phishing URL patterns from cyber security experts
a. Keyword index
b. Length difference between phishing and benign URLs
c. Distribution difference of characters constituting phishing and benign URLs
à The pattern information has been utilized to detect phishing URLs in cyber security field. Therefore, the expert knowledge can be used to prevent phishing attacks.
- We propose a novel integration method of deep learning and logic programmed domain knowledge to use the real-world constraints. We design neural and logic classifiers and
propose the joint learning method of each component based on the traditional neurosymbolic integration.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR	10
U	U	U		

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
AKIRA NAMATAME	3152277010

Standard Form 298 (Rev.5/2020)
Prescribed by ANSI Std. Z39.18

Final Report for AOARD Grant FA2386-21-1-4085

Neuro-Symbolic Integration for Detecting Phishing Attacks

12/04/2023

Name of Principal Investigators: Sung-Bae Cho

- e-mail address: sbcho@yonsei.ac.kr
- Institution: Yonsei University
- Mailing Address: 50 Yonsei-ro, Sudaemoon-gu, Seoul 03722, South Korea
- Phone: +82-2-2123-2720
- Fax: +82-2-2123-8636

Report Due Date: 12/28/2023

Abstract: Considering the fatality of phishing attacks that are emphasized by many organizations, the inductive learning approach using reported malicious URLs has been verified in the field of deep learning. However, the deep learning-based methods mainly focused on the fitting of a classification task via historical URL observation show a limitation of recall due to the characteristics of zero-day attack. To model the nature of a zero-day phishing attack in which URL addresses are generated and discarded immediately, an approach that utilizes the expert knowledge is promising. We introduce the integration method of deep learning and logic programmed domain knowledge to inject the real-world constraints. We design neural and logic classifiers and propose the joint learning method of each component based on the neuro-symbolic integration. Extensive experiments on three real-world datasets consisting of 222,541 URLs show the highest recall among the latest deep learning methods, despite the hostile class-imbalanced condition. We demonstrate that the optimized weighting between neural and logic components has an effect of improving the recall over 3%p compared to the existing methods.

Overview

- Phishing attack is a type of social engineering attack often used to steal user's data, including login credential and credit card numbers. Unsuspecting users visit a website via compromised URL and become victims of devastating loss, unaware of what they are doing. Considering the fatality of phishing attacks that are emphasized by many organizations, the inductive learning approach using reported malicious URLs has been verified in the field of deep learning.
- The deep learning-based methods, mainly focused on the fitting of a classification task via historical URL observations, have a limitation of recall due to the extremely diverse set of phishing URLs. In the field of the information security, an experienced engineer can cope with the phishing attacks by leveraging the stereotype. To detect phishing attacks, an approach that utilizes expert knowledge is promising.
- Phishing URL patterns from cyber security experts
    a. Keyword index
    b. Length difference between phishing and benign URLs
    c. Distribution difference of characters constituting phishing and benign URLs

  → The pattern information has been utilized to detect phishing URLs in cyber security field. Therefore, the expert knowledge can be used to prevent phishing attacks.
- We propose a novel integration method of deep learning and logic programmed domain knowledge to use the real-world constraints. We design neural and logic classifiers and propose the joint learning method of each component based on the traditional neuro-

symbolic integration.

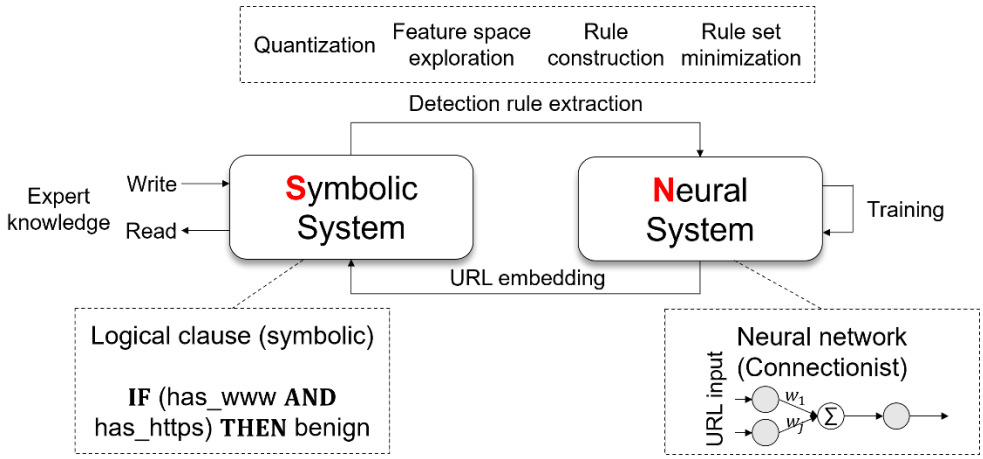


Figure 1. Overview of neuro-symbolic integration in phishing URL detection

Research Objectives

- Develop an integration method of deep learning component and logic component programmed with real-world constraints (Figure 1)
    a. Study on research trends related to phishing attacks and its detection rules
    b. Address limitations of existing phishing detector based on inductive methods
    c. Implement deep learning model to detect phishing URLs based on CNN-LSTM
    d. Implement deductive model to detect phishing URLs based on expert knowledge to protect against phishing attacks
    e. Develop and verify the integration method that can calibrate neural-based phishing detector using symbolic knowledge
- Optimize the symbolic reasoning based on domain knowledge and deep learning for detecting phishing URL
    a. Extract, refine and select expert knowledge to model features of phishing attacks
    b. Optimize the structure of deep learning model that can detect malicious URLs by modelling the distribution of characters and words in benign URLs
    c. Optimize integration method of deep learning and domain knowledge to fully exploit two different approaches and utilize them to detect unknown phishing attacks
        i. Calibration of neural-based phishing detector: define a novel loss function leveraging the output of the deep learning model which can refer to the symbolic reasoning output
        ii. Learning the confidence of expert knowledge: Develop a method to learn the confidence of phishing inference rules based on expert knowledge and optimize the weights based on observations

Accomplishments

List of Publications

[1] S.-J. Bu and S. B. Cho, "Triplet-trained graph transformer with control flow graph for few-shot malware classification," *Information Sciences*, vol. 649, p. 119598, 2023. (SCI, IF: 8.233)

[2] S.-J. Bu, H.-B. Kang, and S.-B. Cho, "Ensemble of deep convolutional learning classifier system based on genetic algorithm for database intrusion detection," *Electronics*, vol. 11, no. 5, p. 745, 2022. (SCIE, IF: 2.397)

[3] J.-Y. Kim and S.-B. Cho, "Obfuscated malware detection using deep generative model based on global/local features," *Computers & Security*, vol. 112, p. 102501, 2022. (SCI, IF: 4.438)

[4] S.-J. Bu and S.-B. Cho, "Phishing URL detection with prototypical neural network disentangled by triplet sampling," *Int. Conf. on Computational Intelligence in Security for Information Systems*, pp. 132-143, 2023.

[5] S.-J. Bu and S.-B. Cho, "A fuzzy transformer network with neuro-fuzzy loss for phishing URL detection," *The 20th World Congress of the International Fuzzy Systems Association*, 2023.

[6] K.-W. Park and S.-B. Cho, "A Vision Transformer Enhanced with Patch Encoding for Malware Classification," *Int. Conf. on Intelligent Data Engineering and Automated Learning*, pp. 289-299, 2022.

[7] H.-J. Moon, S.-J. Bu, and S.-B. Cho, "Directional graph transformer-based control flow embedding for malware classification," *Int. Conf. on Intelligent Data Engineering and Automated Learning*, pp. 426-436, 2021.

[8] K.-W. Park, S.-J. Bu, and S.-B. Cho, "Evolutionary optimization of neuro-symbolic integration for phishing URL detection," *Int. Conf. on Hybrid Artificial Intelligent Systems*, pp. 88-100, 2021.

[9] S.-J. Bu and S.-B. Cho, "Integrating deep learning with first-order logic programmed constraints for zero-day phishing attack detection," *Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 2685-2689, 2021.

Major Activities

- Develop a neuro-symbolic phishing detection model
    a. Develop a neuro-symbolic framework to detect phishing URLs including optimized expert knowledge to combine each other's complementary strengths
- Collect and pre-process expert knowledge for phishing URL detection
    a. Analyze the cyber security research trends and investigate the expert knowledge
    b. Construct domain knowledge for the phishing URL detection
    c. Optimize phishing URL detection rules to utilize expert knowledge and calibrate the deep learning model
- Collect the various benchmark datasets and pre-process them
    a. Collect URL addresses using website crawling algorithm
    b. Construct a URL database by categorizing known attacks and benign URLs
    c. Develop an optimized pre-processing method for phishing URL detection

Impacts

- Develop a practical neuro-symbolic method that achieves the best performance (recall) under class imbalance condition which is the main difficulty to detect phishing URLs
    a. (Neural) Deep learning structure specialized in modeling global/local characteristics of cyber attacks
    b. (Neuro-symbolic) Genetic optimization of neural network for intrusion detection

- Develop and verify a novel balancing method between neural and symbolic decisions using expert knowledge and observations simultaneously
- Develop a prototype system for phishing attack detection by utilizing the expert knowledge and enormous (10M instances of) URL observations

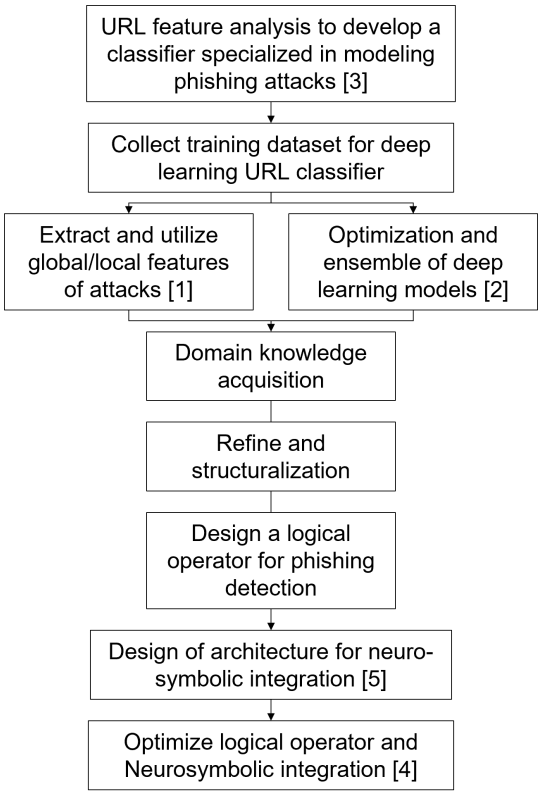


Figure 2. Flowchart of the research

Changes

Changes in approach: None
Problems or delays: None
Expenditure impacts (delays in hiring staff or favorable developments): None
Significant changes in the use or car of human subjects, vertebrate animals and biohazards: None
Changes to the primary place of performance from originally proposed: None

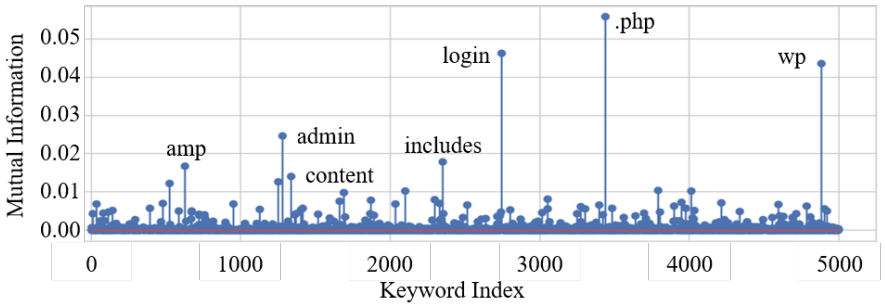
Technical Updates

Introduction. The flow of research is shown as in Figure 2. We design an optimized deep learning-based URL classifier and collect domain knowledge to calibrate the output of the deep learning classifier. The logical classifier is integrated and optimized by designing a novel loss function of deep learning classifier.

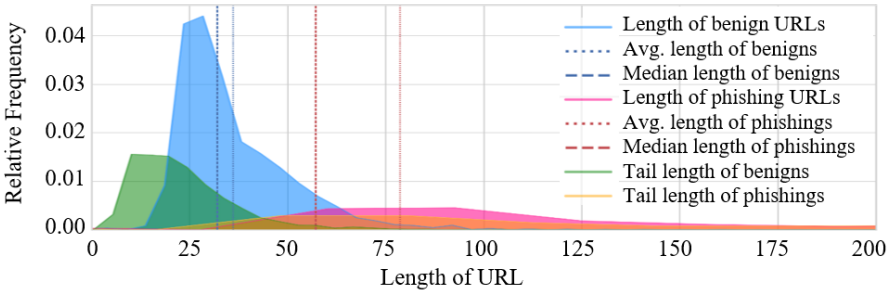
Develop neural component

- **Analyze URL features to discriminate the benign and phishing attacks.** We have found three major statistics in the distribution of characters that consist of URLs in Figure 3 to focus on the difficulties inherent in the field of URL modeling. In Figure 3a, we quantify the effect of specific subdomains on the characteristics of phishing URLs as mutual information. As security experts pointed out, it is confirmed that keywords such as wp, admin, and content from default settings in the personal server and php keyword can be used as abnormal features of phishing URLs. Figure 3b, and c show that phishing URLs are particularly longer than benign URLs and have a composition

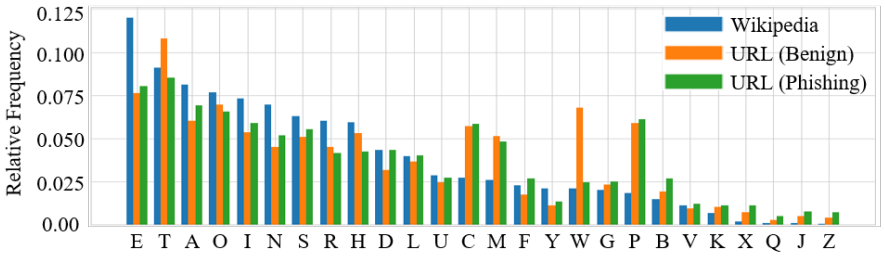
that is significantly different from the alphabetic distribution constituting natural language.



(a) Mutual information for each keyword by decomposing the subdomain of the phishing URLs



(b) Character-level length features of benign and phishing URLs



(c) Distribution of characters constituting phishing and benign URLs

Figure 3. Three main statistics supporting the strong need for neuro-symbolic integration in the task of phishing URL detection: (a) mutual information by keyword; (b) availability of the URL length feature; (c) character distribution that separates benign and phishing URLs.

- **Enhanced feature extraction for phishing URL detection**
  a. Deep character-level analysis: Leveraging the insights from our study [6], we have advanced a CNN architecture capable of discerning subtle character-level anomalies in URLs. This model employs multiple convolutional layers to detect patterns like unusual character combinations or encoded strings, which are common in phishing URLs.
  b. Sophisticated word-level feature recognition: Building on our research in intrusion pattern analysis [7], we have developed an LSTM component designed to identify word sequences and lexical features indicative of phishing. This includes recognizing deceptive domain names or misleading path structures within URLs.
- **Advanced URL pattern recognition techniques**
  a. Malware-inspired URL analysis: Drawing parallels from our malware classification research [3], we have applied sophisticated pattern recognition algorithms to phishing URL detection. This involves identifying structural similarities between malware signatures and phishing URLs, utilizing deep

learning to unravel these intricate patterns.

b. Temporal and spatial feature extraction: We have refined our bidirectional LSTM model to capture not only the temporal sequence of URL characters and words but also their spatial arrangement. This dual-focus approach, inspired by our neural network research, allows for a comprehensive analysis of URL structures, enhancing the detection accuracy for sophisticated phishing schemes.

Develop symbolic component

- **Develop rule-based phishing detection method using domain expertise**
  a. Comprehensive integration of cybersecurity expert knowledge: Our work on rule-based systems [9] has led to the development of an extensive set of detection rules. These rules are derived from a deep understanding of phishing tactics, such as the use of misleading subdomains, the presence of IP addresses instead of domain names, and the specific usage of HTTPS protocols in phishing URLs.
  b. Refinement and optimization of detection rules: Applying evolutionary algorithms from our research [8], we have meticulously optimized these rules. This includes not only improving their accuracy but also ensuring their adaptability to evolving phishing strategies. We have employed techniques such as rule weighting and dynamic threshold adjustment to enhance the system's responsiveness to new phishing patterns.
- **Improve symbolic logic for phishing URL detection**
  a. Detailed structural analysis of URL components: Our symbolic logic research has enabled us to break down URLs into their constituent parts for detailed analysis. This involves examining each segment of a URL – from the protocol to the query string – using a set of logical rules tailored to identify phishing indicators in each segment.
  b. Dynamic and adaptive rule application: In recognition of the ever-evolving nature of phishing attacks, our method features an adaptive logic framework. This framework dynamically adjusts the application of detection rules based on real-time analysis of emerging phishing tactics, ensuring the system remains effective against new and sophisticated phishing methods. We have integrated machine learning algorithms to continually update and refine the rule set, ensuring it stays abreast of the latest trends in phishing techniques.
  c. Organize the expert knowledge to mitigate the conflict among rules: To structuralize symbolic components, we have collected rules that have proven fairness and validity with expert knowledge for detecting phishing URLs. For example, in domain knowledge, a longer URL than 54 characters tends to hide the suspicious part, and it is likely to be classified as a phishing URL. The rules are categorized into four criteria: address-based, abnormal request-based, domain-based, and script-based. We utilize such expert knowledge from former research, and 12 significant rules are evenly selected from each criterion. A total of 12 rules are collected and organized in the form of first-order logic (FOL) as shown in Table 1. For example, the knowledge that ‘the input URL has IP address and contains both '#' and '%' within the URL’ is defined as follows:

$$IPAddress(x_i) \wedge \exists c: \left(Special_{\#}(c) \wedge Special_{\%}(c)\right) \Longrightarrow Phishing\ (x_i)$$

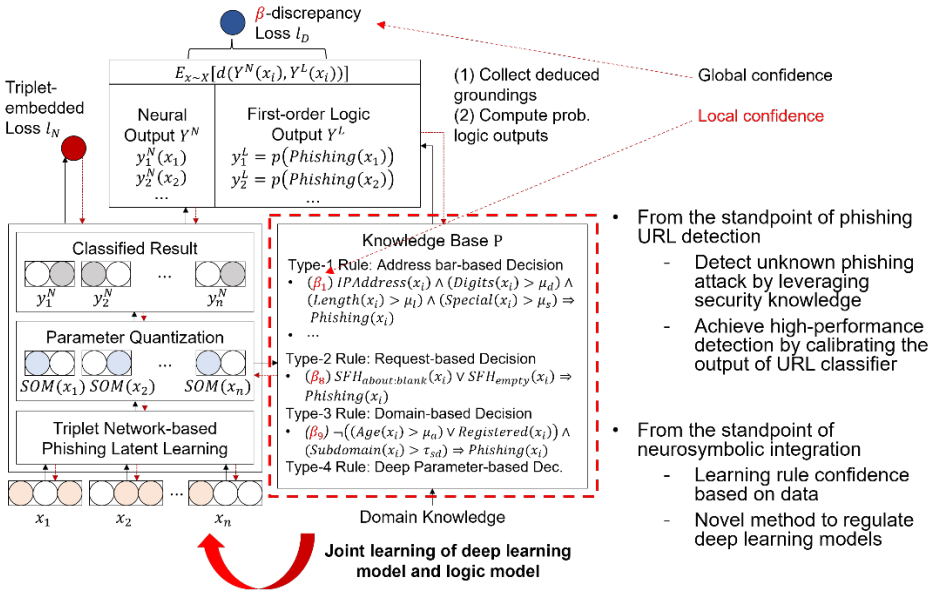


Figure 4. Joint learning of deep learning and logic programmed constraints implemented with triplet network and discrepancy loss function.

Table 1. Address, request, domain, script-based expert knowledge for phishing URL detection

Type	Rule features	First-Order Logic (FOL) expression
Address -bar	Usage of IP address	**1**. $IPAddress(x_i) \wedge (Digits(x_i) > \mu_d) \wedge (Length(x_i) > \mu_l) \wedge (Special(x_i) > \mu_s) \Rightarrow Phishing\ (x_i)$
	Digit count	
	Character length	**2**. $IPAddress(x_i) \wedge \exists c: (Special_{\#}(c) \wedge Special_{\%}(c)) \Rightarrow Phishing\ (x_i)$
	Special character count	**3**. $Length(x_i) > \tau_l) \wedge (Digits(x_i) > \tau_d) \Rightarrow Phishing\ (x_i)$
		4. $(Dot(x_i) > \tau_d) \Rightarrow Phishing\ (x_i)$
Abnormal request	URL request count	**5**. $(Request_{src}(x_i) > \mu_r) \Rightarrow Phishing\ (x_i)$
	SFH	**6**. $SFH_{abount:blank}(x_i) \vee SFH_{empty}(x_i)) \Rightarrow Phishing\ (x_i)$
Domain	Age of domain	**7**. $HTTPS(x_i) \wedge Typosquatted(x_i) \Rightarrow Phishing\ (x_i)$
	Registered top domain	**8**. $HTTP(x_i) \wedge (Special(x_i) > \mu_s) \Rightarrow Phishing\ (x_i)$
	Subdomain count	**9**. $\neg(Registered(x_i)) \wedge (Subdomain(x_i) > \tau_{sub}) \Rightarrow Phishing\ (x_i)$
	Typo-squatted domain	**10**. $\neg(\mathrm{Age}(x_i) > \mu_a) \wedge (Subdomain(x_i) > \tau_{sub}) \Rightarrow Phishing\ (x_i)$
	Suspicious TLD	**11**. $TLD(x_i) \Rightarrow Phishing\ (x_i)$
Script	Usage of mouseover	**12**. $\neg\mathrm{Rightclick}(x_i) \wedge Mouseover(x_i) \wedge Popup(x_i) \Rightarrow Phishing\ (x_i)$
	Usage of pop-up	
	Disabling right click	

Table 2. Source and description of the three benchmark phishing URL datasets

Source	URL Label	Instances	e.g. (accessed date: 20 June 2022)
ISCX-URL-2016	Benign	35,000	http://metro.co.uk/2015/05
	Phishing	9000	http://standardprincipal.pt/
	Malware	11,000	http://9779.info/%E5%88%
	Spam	12,000	http://adverse*s.co.uk/scr/cl
PhishStorm	Benign	47,682	en.wikipedia.org/wiki/Walkingdead
	Phishing	47,859	nobell.it/70ffb52d079109dc
PhishTank	DMOZ Open Directory Project (Benign)	45,000	http://geneba**.org/ftp/
	OpenDNS (Phishing)	15,000	http://droopbxoxx.com/@@@

Develop a neuro-symbolic integration method to detect phishing attacks

- **Neural component**
  a. Triplet-embedded neural architecture: Following insights from our research on graph transformers and deep convolutional learning [4], we have implemented a triplet-embedded neural network. This structure excels in learning latent features of phishing attacks, distinguishing between phishing and benign URLs with high precision.
  b. Quantization of neural outputs: Utilizing the self-organizing map (SOM) technique, inspired by our work on obfuscated malware detection [1], we convert neural outputs into quantized symbols. These symbols represent complex URL features, capturing subtleties that are crucial for accurate phishing detection.
- **Symbolic component**
  a. Integration of expert knowledge: Drawing from our research on fuzzy logic and neuro-fuzzy systems [5], we have encoded expert cybersecurity knowledge into a set of logical rules. These rules are applied to the quantized symbols, enabling the system to reason about the phishing likelihood of URLs.
  b. Deductive reasoning with local confidence: The logic component employs these rules, formulated in first-order logic, to deduce the nature of URLs. This process is informed by our work on evolutionary optimization [2], allowing the system to adjust its reasoning based on local confidence metrics.
- **Integration and calibration**
  a. Neuro-symbolic integration for phishing URL detection: We extract the quantized symbols from deep learning component and utilize it to calibrate the logic components. Figure 4 illustrates the overall architecture of the proposed method that consists of triplet-embedded neural component, logic operator, and a beta-discrepancy loss. The neural component learns the latent space of the phishing attacks and classifies the phishing and benign URLs as neural output y_i^N. The quantized deep parameter represented by the SOM is utilized as a symbol y_i^C of the logic component. We build a knowledge base consisting of four types of rules from expert knowledge. The logic component deduces the logic output y_i^L from the quantized symbols of the neural component based on 10 rules expressed in first-order logic with its local confidence β_k respectively. Finally, the neural output and logic output are tuned for the purpose of minimizing the discrepancy loss l_D defined by the distance metric $d(y_i^N, y_i^L)$. The β-discrepancy loss function l_{BD} is a linear combination of the classification loss l_N and the discrepancy loss l_D, and the global confidence β_G adjusts the weight of the logic integration.
  b. Beta-discrepancy loss for calibration: In line with our study on integrating deep learning with logic constraints, we have implemented a beta-discrepancy loss function. This function optimizes the integration by minimizing the discrepancy between neural and logic outputs, using a distance metric for calibration.
  c. Global confidence adjustment: The global confidence parameter balances the influence of neural and symbolic components, ensuring a harmonious and effective integration.

Collect the various benchmark datasets and preprocess them. We validate the proposed method with the benchmark URL database. For extensive evaluation, three real-world URL datasets consisting of 1,048,576 URLs were collected as summarized in Table 2. The ISCX-URL-2016 dataset aims at the four-way classification task consisting of benign, phishing,

malware, and spam URLs, and has a 3:1 class imbalance as a characteristics of malicious URL modeling. Web-accessible Phishstorm and Phishtank datasets provide known phishing attack cases. Unlike the Phishstorm dataset where class sampling was performed, Phishtank does not provide a benign URL. We have collected benign URLs from the Open Directory Project, resulting in 95,541 and 60,000 URLs.

Experimental results and discussion

- **Improved accuracy and recall.** Our neuro-symbolic integration method achieves an accuracy of 0.9785 and a recall of 0.9610 for the ISCX-URL-2016 dataset, outperforming the base network models and the latest deep metric learning approaches.
- **Handling misclassifications.** Despite the high performance, we have observed misclassification cases, particularly with benign URLs having complex character sequences (as shown in Table 3). This indicates a need for further refinement in feature extraction, which can be addressed by extending our neuro-symbolic approach to fully leverage both character-level and word-level URL features (as shown in Table 4).

Table 3. 10-fold cross-validation of accuracy and recall with the state-of-the-art methods.

Benchmark Dataset	**ISCX-URL-2016**		**PhishStorm**		**PhishTank**	
Metrics	**Acc.**	**Recall**	**Acc.**	**Recall**	**Acc.**	**Recall**
Base Network						
Character-CNN [10]	0.9363	0.8909	0.9016	0.8565	0.8852	0.8034
LSTM	0.9175	0.8803	0.8777	0.8440	0.8544	0.7865
CNN-LSTM	0.9424	0.9015	0.9229	0.8785	0.9070	0.8374
Comparative Studies						
URLNet [11]	0.9450	0.9390	0.9395	0.8864	0.9226	0.8785
Texception [12]	0.9765	0.9462	0.9710	0.9227	0.9319	0.9075
Triplet network [13]	0.9505	0.9064	0.9473	0.8902	0.9081	0.8469
Proposed Method						
Logic programming	0.7338	0.6365	0.7013	0.6250	0.6463	0.5837
CNN-LSTM based triplet network	0.9655	0.9364	0.9565	0.9100	0.9110	0.8572
Logic integration (β_G=1.00)	0.9765	0.9552	0.9755	0.9330	0.9398	0.9202
Logic integration (β_G Optimized)	**0.9785**	**0.9610**	**0.9832**	**0.9464**	**0.9540**	**0.9265**

Table 4. Summary of the correctly classified and misclassified cases.

Component	URL (accessed date: 12-04-2023)	Classification Result	Label	Predict
Neuro-symbolic Integration	http://longsdale.dk/5DSJKYT-5D78G45-3148TYUI1-FBZF6E15-1L46GSRFD-SC1DF6JDG4/default/38af09dca62acf475045bf..	Correct	Phishing	Phishing
	http://www.danville-va.gov/l_fire.asp?menuid=2820&sub1menuid=2832&sub2menuid=2886&cid=3537	Correct	Phishing	Phishing
	https://bit.ly/2VWEUsp	Misclassified	Phishing	Benign
Neural	http://www.musk-space.tech/bitcoin/index.html	Correct	Phishing	Phishing
	http://longsdale.dk/5DSJKYT-5D78G45-3148TYUI1-FBZF6E15-1L46GSRFD-SC1DF6JDG4/default/38af09dca62acf475045bf..	Misclassified	Phishing	Benign
Symbol	https://mypreferences.allstate.com/activity-handler?cid=EMC-C-A-eA-RMass-171101&pl=QUNUSU9OPVVOU1VCfEVNQUlMPVlySEtzTWY0RlRpejN4YVQ5NFpIT1pqWnRCc1ZXZk1ZVHZDeHVuTnQ5RHc9fF (Case-S1)	Correct	Phishing	Phishing
	http://www.danville-va.gov/l_fire.asp?menuid=2820&sub1menuid=2832&sub2menuid=2886&cid=3537	Misclassified	Benign	Phishing

Attachments

2023_INS_SJB.pdf: published paper for [1].
2022_Electronics_SJB.pdf: published paper for [2].
2022_COSE_JYK.pdf: published paper for [3].
2023_ICISIS_SJB.pdf: published paper for [4].
2023_IFSA_SJB.pdf: published paper for [5].
2022_IDEAL_KWP.pdf: published paper for [6].
2021_IDEAL_HJM.pdf: published paper for [7].
2021_HAIS_KWP.pdf: published paper for [8].
2021_ICASSP_SJB.pdf: published paper for [9].

References

[1] S.-J. Bu and S. B. Cho, "Triplet-trained graph transformer with control flow graph for few-shot malware classification," *Information Scienc*es, vol. 649, p. 119598, 2023. (SCIE) (IF: 8.233)

[2] S.-J. Bu, H.-B. Kang, and S.-B. Cho, "Ensemble of deep convolutional learning classifier system based on genetic algorithm for database intrusion detection," *Electronics*, vol. 11, no. 5, p. 745, 2022. (SCIE) (IF: 2.397)

[3] J.-Y. Kim and S.-B. Cho, "Obfuscated malware detection using deep generative model based on global/local features," *Computers & Security*, vol. 112, p. 102501, 2022. (SCIE) (IF: 4.438)

[4] S.-J. Bu and S.-B. Cho, "Phishing URL detection with prototypical neural network disentangled by triplet sampling," *Int. Conf. on Computational Intelligence in Security for Information Systems*, pp. 132-143, 2023.

[5] S.-J. Bu and S.-B. Cho, "A fuzzy transformer network with neuro-fuzzy loss for phishing URL detection," *The 20th World Congress of the International Fuzzy Systems Association*, 2023.

[6] K.-W. Park and S.-B. Cho, "A Vision Transformer Enhanced with Patch Encoding for Malware Classification," *Int. Conf. on Intelligent Data Engineering and Automated Learning*, pp. 289-299, 2022.

[7] H.-J. Moon, S.-J. Bu, and S.-B. Cho, "Directional graph transformer-based control flow embedding for malware classification," *Int. Conf. on Intelligent Data Engineering and Automated Learning*, pp. 426-436, 2021.

[8] K.-W. Park, S.-J. Bu, and S.-B. Cho, "Evolutionary optimization of neuro-symbolic integration for phishing URL detection," *Int. Conf. on Hybrid Artificial Intelligent Systems*, pp. 88-100, 2021.

[9] S.-J. Bu and S.-B. Cho, "Integrating deep learning with first-order logic programmed constraints for zero-day phishing attack detection," *Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 2685-2689, 2021.

[10] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," *Adv. in Neural Information Processing Systems*, pp. 649-657, 2015.

[11] H. Le, Q. Pham, D. Sahoo, and S. C. Hoi, "URLNet: Learning a URL representation with deep learning for malicious URL detection," *arXiv preprint arXiv:1802.03162*, 2018.

[12] F. Tajaddodianfar, J. W. Stokes, and A. Gururajan, "Texception: a character/word-level deep learning model for phishing URL detection," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 2857-2861, 2020.

[13] S. Novoselov, V. Shchemelinin, A. Shulipa, A. Kozlov, and I. Kremnev, "Triplet loss based cosine similarity metric learning for text-independent speaker recognition," *Interspeech*, pp. 2242-2246, 2018.