DEEP LEARNING FOR PHISHING

Safa ORHAN
Computer Engineering Student
Istanbul Kultur University
Istanbul, Turkey

Eyüp USTA
Computer Engineering Student
Istanbul Kultur University
Istanbul, Turkey

I. DEEP LEARNING FOR PHISHING ATTACK DETECTION [3]

Approaches Used = Deep Neural-Network, Feed-Forward Deep Neural-Network, Recurrent Neural-Network, Convolutional Neural-Network, Restricted Boltzmann Machine, Deep Belief Network, Deep Auto-Encoder.

1. *Authors in Subasi et al.: [2]*
   - ✓ Suggested a comparison of Adaboost and multi boosting to detect phishing website.
   - ✓ They used the UCI machine learning repository dataset with 11,055 samples and 30 features.
   - ✓ Ensemble models improve the exhibition of the classifiers in terms of precision, F-measure, and ROC region.
   - ✓ Experimental results reveal that by utilizing ensemble models, it is possible to recognize phishing pages with a precision of 97.61%.

2. *Authors in Abdelhamid et al.: [1]*
   - ✓ Proposed a comparison based on model content and features.
   - ✓ They used a dataset from PhishTank, containing around 11,000 examples.

- ✓ They used an approach named enhanced dynamic rule induction (eDRI) and claimed that dynamic rule induction (eDRI) is the first algorithm of machine learning and DL which has been applied to an anti-phishing tool.
- ✓ This algorithm passes datasets with two main threshold frequencies and rules strength. The training dataset only stores "strong" features and these features become part of the rule while others are removed.

II. DETECTION OF PHISHING URLs BY USING DEEP LEARNING APPROACH AND MULTIPLE FEATURES COMBINATIONS [4]

1. *Correa Bahnsen et al.: [5]*

The researchers of used a recurrent neural network with character level embeddings and a 2,000,000 URLs dataset. They achieved a 93,40% F1 score.

2. *Wei et al.: [6]*

The authors of used word level embeddings with a convolutional neural network and a dataset of 1,523,966 samples. They achieved an 86.63% accuracy of classifying URLs into the phishing and legitimate ones.

3. *Yang, Zhao et al.: [7]*

In the research on character level embeddings and a network with a mix of convolutional and recurrent layers were used. The authors gathered a dataset of 1,021,758 samples and achieved a 98.61% classification accuracy.

4. *Wang et al.: [8]*

The authors of experimented with character level embeddings and a network of convolutional and recurrent layers. They used a dataset of 5,118,727 URLs samples and achieved an F1 score of 95.52%.

5. *Saxe and Berlin: [9]*

In the research of the authors used character level embeddings and a convolutional neural network to classify URLs and achieved an AUC (Area under the curve) of 99,30%. A dataset that was used contained 19,067,879 samples of URLs.

6. *Le, Pham et al.: [10]*

The authors of used a combination of character and word level embeddings with a convolutional neural network. They used a dataset of 15,000,000 samples and achieved an AUC of 99,29%.

7. *Sahingoz et al.: [11]*

We propose the use of Artificial Neural Network and Deep Neural Network based system for classifying the incoming URLs. The experimental results show that both these approaches result satisfactory accuracy rate and DNN with 40*20 hidden layer structure produce best solution with about 96% of accuracy. The latency of the execution time of the algorithm is also an important metric for selection of the detection algorithms. As seen from the results use of Alexa Ranking results a great increase in the execution time, although it has a great importance for detection of phishing. Therefore, according to aim of the system this feature can be disabled for decreasing the execution time. As the Future works, to decrease the execution time and increase the efficiency of the system, the power of the Graphics Programming Units can be used. Additionally, the other approaches of Deep Learning, such as

recurrent neural networks, convolutional neural networks and LSTM can be tested for increasing the performance of the system.

REFERENCES

[1] Abdelhamid, N., Thabtah, F., Abdel-jaber, H. (2017). Phishing detection: A recent intelligent machine learning comparison based on models content and features. In 2017 IEEE international conference on intelligence and security informatics (ISI) (pp. 72–77). IEEE.

[2] Subasi, A., Molah, E., Almkallawi, F., & Chaudhery, T. J. (2017). Intelligent phishing website detection using random forest classifier. In 2017 International conference on electrical and computing technologies and applications (ICECTA) (pp. 1–5). IEEE.

[3] Basit, A., Zafar, M., Liu, X., Javed, A.R., Jalil, Z., Kifayat, K. (2020). A comprehensive survey of AI-enabled phishing attacks detection techniques. (pp. 139–154). Springer.

[4] Rasymas, T., Dovydaitis, L. (2020). Detection of Phishing URLs by Using Deep Learning Approach and Multiple Features Combinations. (pp. 471–483).

[5] Correa Bahnsen, A., Contreras Bohorquez, E., Villegas, S., Vargas, J., González, F. A. (2017). Classifying Phishing URLs Using Recurrent Neural Networks.

[6] Wei, B., Hamad, R. A., Yang, L., He, X., Wang, H., Gao, B., Woo, W. L. (2019). A Deep-Learning-Driven Light-Weight Phishing Detection Sensor. Sensors, 19(19), 4258.

[7] Yang, P., Zhao, G., Zeng, P. (2019). Phishing website detection based on multidimensional features driven by deep learning. IEEE Access, 7, 15196–15209.

[8] Wang, W., Zhang, F., Luo, X., Zhang, S. (2019). PDRCNN: Precise Phishing Detection with Recurrent Convolutional Neural Networks. 2019, 15.

[9] Saxe, J., Berlin, K. (2017). eXpose: A Character-Level Convolutional Neural Network with Embeddings for Detecting Malicious URLs, File Paths and Registry Keys.

[10] Le, H., Pham, Q., Sahoo, D., Hoi, S. C. H. (2018). URLNet: Learning a URL Representation with Deep Learning for Malicious URL Detection.

[11] Sahingoz, O.K., Baykal, S.I., Bulut, D. (2018). Phishing Detection from URLs by Using Neural Networks. Computer Science & Information Technology (CS & IT), 41–54.