Literature Review: Phishing Detection using Machine Learning and Deep Learning

Phishing remains a prevalent cybersecurity threat, and recent research emphasizes the role of intelligent systems, particularly machine learning (ML) and deep learning (DL), in combating such attacks. This literature review highlights findings from two key studies, one focusing on traditional ML models and the other on deep learning frameworks.

1. Traditional ML-Based Phishing Detection

Abdelhamid et al. (2017) in their paper *"Phishing Detection: A Recent Intelligent Machine Learning Comparison Based on Models, Content, and Features"* conduct a comparative study of eight ML algorithms for phishing website detection. The authors argue that phishing classification is well suited for supervised ML techniques due to its binary nature (phishing vs. legitimate).

Key contributions:

- Evaluated models include Naive Bayes, C4.5, SVM, AdaBoost, and rule-based classifiers (e.g., RIDOR, OneR, eDRI).
- A large dataset of over 11,000 websites from sources like Phishtank and MillerSmiles was used.
- Features such as URL_of_Anchor and SSL_Final_State were found to be highly influential.

Findings:

- eDRI and RIDOR produced compact rule based models with high accuracy and interpretability, making them suitable for novice users.
- C4.5 had high accuracy but generated overly complex models with hundreds of rules, reducing practical usability.
- Rule based models were recommended for real world deployment due to their simplicity and robust performance.

This study demonstrates that rule learning approaches can outperform complex classifiers when usability, interpretability, and phishing detection rate are balanced.

2. Deep Learning-Based Phishing Detection

Sahingoz et al. (2024) present *"DEPHIDES: Deep Learning Based Phishing Detection System"*, focusing on URL based detection using various deep learning architectures. The system was trained on an extensive dataset of over 5.1 million URLs (2.3M phishing, 2.8M legitimate), which is the largest publicly shared phishing dataset to date.

Key methods:

- Five DL models were explored: Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Bidirectional RNNs, and Attention Networks.
- A character level embedding approach was used for vectorization, enhancing the language independence and robustness of the system.

Highlights:

- CNN achieved the best performance with a phishing detection accuracy of 98.74%.
- The system is designed to detect zero day phishing attacks, outperforming traditional ML models by avoiding reliance on third party services like blacklists or Whois.
- Its real time detection capability and scalability make it ideal for proactive cybersecurity infrastructure.

This study illustrates the effectiveness of deep learning, especially CNNs, in handling complex, large scale phishing detection tasks with minimal human intervention.

Summary

Aspect	Abdelhamid et al. (2017)	Sahingoz et al. (2024)
Approach	Traditional ML (e.g., C4.5, eDRI, SVM)	Deep Learning (CNN, RNN, Bi-RNN, Attention)
Dataset	~11,000 URLs	~5.1 million URLs
Key Techniques	Rule-based models, decision trees	Character-based embeddings, URL classification
Best Performing Model	eDRI (simple, accurate, interpretable)	CNN (high accuracy: 98.74%)
Zero-Day Attack Capability	Limited	Strong
Model Interpretability	High (e.g., rule sets for users)	Moderate (less interpretable than rule sets)