Abstract – The analysis of document, titled "AntiPhishStack: LSTM-based Stacked Generalization Model for Optimized Phishing URL Detection," will cover various aspects of the document, including its methodology, results, and implications for cybersecurity. Specifically, the document's approach to using Long Short-Term Memory (LSTM) networks within a stacked generalization framework for detecting phishing URLs will be examined. The effectiveness of the model, its optimization strategies, and its performance compared to existing methods will be scrutinized.

The analysis will also delve into the practical applications of the model, discussing how it can be integrated into existing cybersecurity measures and its potential impact on reducing phishing attacks. The document's relevance to cybersecurity professionals, IT specialists, and stakeholders in various industries will be highlighted, emphasizing the importance of advanced phishing detection techniques in the current digital landscape.

This summary will serve as a valuable resource for cybersecurity experts, IT professionals, and others interested in the latest developments in phishing detection and prevention.

I. Introduction

The paper titled "LSTM-based Stacked Generalization Model for Optimized Phishing" discusses the escalating reliance on revolutionary online web services, which has introduced heightened security risks, with persistent challenges posed by phishing attacks.

Phishing, a deceptive method through social and technical engineering, poses a severe threat to online security, aiming to obtain illicit user identities, personal account details, and bank credentials. It's a primary concern within criminal activity, with phishers pursuing objectives such as selling stolen identities, extracting cash, exploiting vulnerabilities, or deriving financial gains.

The study aims to advance phishing detection with operating without prior phishing-specific feature knowledge. The model leverages the capabilities of Long Short-Term Memory (LSTM) networks, a type of recurrent neural network that is capable of learning order dependence in sequence prediction problems. It leverages the learning of URLs and character-level TF-IDF features symmetrically, enhancing its ability to combat emerging phishing threats.

II. Methodology and Significance of the Study

It presents a novel model for detecting phishing sites. The significance of this study lies in its advancement of phishing detection techniques, specifically through the introduction of a two-phase stack generalized model named AntiPhishStack.

This model is designed to detect phishing sites without requiring prior knowledge of phishing-specific features, which is a significant improvement over traditional phishing detection systems that rely on machine learning and manual features.

This research contributes to the ongoing discourse on symmetry and asymmetry in information security and provides a forward-thinking solution for enhancing network security in the face of evolving cyber threats.

The data source used in the study includes two benchmark datasets comprising benign and phishing or malicious URLs. These datasets are used for experimental validation of the model. The datasets are referred to as DS1 and DS2 within the paper, with DS1 including benign Yandex sites and PhishTank phishing sites, and DS2 consisting of benign sites from common-crawl, the Alexa database, and phishing sites from PhishTank.

III. Key Components

According to the methodology the proposed model operates in two phases (two-phase stack generalized model):

- **Phase I**: The model learns URLs and character-level TF-IDF features symmetrically. These features are trained on a base machine learning classifier, employing K-fold cross-validation for robust mean prediction.
- **Phase II**: A two-layered stacked-based LSTM network with five adaptive optimizers is used for dynamic compilation, ensuring premier prediction on these features.
- Additionally, the symmetrical predictions from both phases are optimized and integrated to train a meta-XGBoost classifier, contributing to a final robust prediction.

A. URL Features

- **URL Structure**: The paper emphasizes that attackers often create phishing URLs that appear legitimate to users. Attackers use URL jamming tactics to deceive users into disclosing personal information.
- **Lightweight Features**: The research aims to detect phishing websites using lightweight features, specifically a weight factor URL token system, which allows for quick detection without accessing the website's content.

- **Weight Calculation**: The paper provides a formula for calculating the weight W_i for i-th distinct word in a URL, which is used to assign a weight value to each URL for phishing prediction.
- **URL Components**: The paper describes the components of a URL, including the protocol, host IP address or resource location, major domains, top-level domains (TLD), port number, path, and optional fields like inquiry.
- **Phishing Indicators**: Several sub-features are identified as indicators of phishing, such as the use of an IP address instead of a domain name, the presence of the '@' symbol, the "//" symbol, domain name prefixes and suffixes separated by the "-" sign, and the use of multiple sub-domains.
- **HTTPS and Certificate Age**: The paper notes that most legitimate sites use HTTPS, and the age of the certificate is crucial. A trustworthy certificate is required.
- **Favicon**: The favicon can be used to redirect clients to dubious sites when layered from external space.
- **Sub-features Analysis**: The paper provides an analysis of sub-features like the IP address, '@' symbol, "//" symbol, domain name prefixes and suffixes, HTTPS, and favicon, explaining how these features can be used to identify phishing websites

B. Character Level Features

- **TF-IDF for Character-Level Features**: The paper utilizes Term Frequency-Inverse Document Frequency (TF-IDF) at the character level to determine the relative importance of characters within URLs across the corpus of URLs being analyzed.
- **TF-IDF Calculation**: The TF-IDF score is composed of two parts: Term Frequency (TF), which is the normalized count of a term within a document, and Inverse Document Frequency (IDF), which is the logarithm of the ratio of the total number of documents to the number of documents containing the term.
- **Levels of TF-IDF**: The paper mentions that TF-IDF vectors can be generated at different levels, such as word level, character level, and n-gram level, with the character level being particularly relevant for this study.
- **Limitations of TF-IDF**: The paper acknowledges that while TF-IDF is useful for extracting prominent keywords, it has limitations, such as failing to extract misspelled terms, which can be problematic since URLs may contain nonsensical words.
- **Character-Level TF-IDF**: To address the limitations of TF-IDF for URLs that may contain misspelled or nonsensical words, the study employs a character-level TF-IDF approach with a maximum feature count of 5000.
- **Natural Feature Learning**: The model treats URL strings as character sequences, which are considered natural features that do not require prior feature knowledge for the model to learn effectively.
- **Stack Generalization for Feature Extraction**: The model uses stack generalization to extract local URL features from the character sequences, and a meta-classifier is designed for the final prediction.
- **Advantages of the Approach**: This approach allows the proposed model to train on URL character sequences as natural features, which simplifies the learning process and potentially improves the model's ability to detect phishing URLs without prior feature knowledge

C. Stack generalization model

- **Two-Phase Approach**: The model is divided into two phases. Phase I uses machine learning classifiers to generate a mean prediction, while Phase II employs a two-layered LSTM-based stack generalized model optimized for premier prediction in phishing site detection.
- **Integration of Predictions**: The mean prediction from Phase I is combined with the premier prediction from Phase II. A meta-classifier, specifically XGBoost, is then used to deliver the final prediction.
- **Stack Generalization Technique**: The model uses stack generalization, an ensemble learning methodology that integrates various machine learning algorithms and deep learning models, to enhance detection impact.
- **Model Flow**: The model's flow includes collecting datasets, dividing them into training and testing sets, constructing the stack generalization model's phases, and merging predictions for the ultimate prediction.
- **Feature Importance**: The model emphasizes the importance of URL and character-level TF-IDF features, which are learned symmetrically to detect phishing web pages.
- **Significant Advantages**: The model offers several advantages, including independence from prior feature knowledge, strong generalization ability, and independence from cybersecurity experts and third-party services.
- **Enhanced Phishing Detection**: The model aims to intelligently identify new phishing URLs previously unidentified as fraudulent, demonstrating robust performance on benchmark datasets.

D. Experiments

It presents the experimental validation of the proposed model. The model was tested on two benchmark datasets, which comprised benign and phishing or malicious URLs.

- The model demonstrated exceptional performance in detecting phishing sites, achieving an accuracy of 96.04%. This result was notably higher compared to existing studies.

- The model was assessed through various matrices, including AUC-ROC curve, Precision, Recall, F1, mean absolute error (MAE), mean square error (MSE), and accuracy.
- A comparative analysis with baseline models and traditional machine learning algorithms, such as support vector machine, decision tree, naïve Bayes, logistic regression, K-nearest neighbor, and sequential minimal optimization, highlighted the superior phishing detection efficiency of the model.
- The model was found to be effective in identifying new phishing URLs that were previously unidentified as fraudulent.
- The model operates without prior phishing-specific feature knowledge, which is a significant advantage in achieving advancements in cybersecurity

E. Optimizer evaluation on LSTM

- **Optimizer Performance**: The paper evaluates the performance of five different adaptive optimizers: AdaDelta, Adam, RMSprop, AdaGard, and SGD (Stochastic Gradient Descent), to determine which is best suited for the proposed anti-phishing model.
- **Epochs and Learning Rate**: Different numbers of epochs are considered to implement the 2-layered LSTM with different optimizers. The learning rate, a crucial hyperparameter, is adjusted for each optimizer to control the speed at which the model learns.
- **Accuracy, MSE, and MAE**: The paper reports the accuracy, mean squared error (MSE), and mean absolute error (MAE) for each optimizer with the LSTM-based stack generalization model on two datasets (DS1 and DS2).
- **Results on Datasets**: The AdaGard optimizer provided the highest accuracy with the lowest MSE and MAE on DS1, while the Adam optimizer achieved the highest accuracy on DS2.
- **Precision-Recall Curves**: Precision-recall curves are presented for each feature set, indicating the trade-off between precision and recall for the different optimizers.
- **Optimizer Selection**: The analysis suggests that the learning rate significantly contributes to the success of the proposed model with the adaptive optimizers. The Adam optimizer is highlighted for its performance with a specific learning rate when the 2-layered LSTM is employed with 100 epochs.
- **Comparative Analysis**: The average performance of the optimizers on DS1 and DS2 is compared, with DS2 showing slightly better accuracy.
- **Significance of Optimizers**: The evaluation of optimizers is crucial for the model's accuracy, which is a key component of machine learning and artificial intelligence, responsible for molding the model to acquire the most accurate results possible

IV. KEY FINDINGS

The model's design allows it to effectively identify new phishing URLs previously unidentified as fraudulent, thus reducing the likelihood of false negatives. The use of K-fold cross-validation and a two-layered LSTM network helps to mitigate overfitting and improve the model's ability to correctly classify phishing sites, thereby reducing the likelihood of false positives.

- **Development of model**: a novel mode introduced via two-phase stack generalized model designed to detect phishing sites effectively.
- **Learning of URLs and character-level TF-IDF features symmetrically**: This model leverages the learning of URLs and character-level TF-IDF features symmetrically. This enhances the model's ability to combat emerging phishing threats.
- **Two-phase operation**: In Phase I, features are trained on a base machine learning classifier, employing K-fold cross-validation for robust mean prediction. Phase II employs a two-layered stacked-based LSTM network with five adaptive optimizers for dynamic compilation, ensuring premier prediction on these features.
- **Integration of predictions (Meta-XGBoost Classifier)**: The symmetrical predictions from both phases are optimized and integrated to train a meta-XGBoost classifier, contributing to a final robust prediction.
- **Independence from prior phishing-specific feature knowledge**: The model operates without prior phishing-specific feature knowledge, which is a significant advancement in phishing detection that showing strong generalization ability and independence from cybersecurity experts and third-party services.
- **High performance**: Experimental validation on two benchmark datasets, comprising benign and phishing or malicious URLs, demonstrates the model's exceptional performance, achieving a notable 96.04% accuracy compared to existing studies
- **Independence from cybersecurity experts and third-party services:** This model autonomously extracts necessary URL features, eliminating the reliance on cybersecurity experts. It also demonstrates independence from third-party features such as page rank or domain age
- **Strong generalization ability**: The URL character-based features are utilized for more robust generalization and check-side accuracy, and the multi-level or low-level features are combined in the hidden layers of the neural network to attain effective generalization

- **Prior feature knowledge independence**: The approach taken in this work treats URL strings as character sequences, serving as natural features that require no prior feature knowledge for the proposed model to learn effectively
- **Enhancing Network Security**: The research adds value to the ongoing discourse on symmetry and asymmetry in information security and provides a forward-thinking solution for enhancing network security in the face of evolving cyber threats.

V. Benefits and Limitations of the Study

Comparatively, traditional phishing systems, reliant on machine learning and manual features, struggle with evolving tactics. Other models, such as the CNN-LSTM model and the end-to-end deep learning architecture grounded in natural language processing techniques, have shown limitations in their generalization on test data and their dependency on existing knowledge of phishing detection. The model, in contrast, shows strong generalization ability and independence from prior feature knowledge, making it a robust and effective tool for phishing detection.

The benefits of the study compared to traditional phishing systems include:

- **Prior Feature Knowledge Independence**: The proposed model does not require prior phishing-specific feature knowledge, which allows it to adapt to new and evolving phishing tactics more effectively than traditional systems that rely on predefined features.
- **Strong Generalization Ability**: The model uses URL character-based features for robust generalization and check-side accuracy, which enables it to generalize across different phishing threats better than traditional systems that may not adapt as well to variations in phishing URLs.
- **Independence from Cybersecurity Experts and Third-Party Services**: The model autonomously extracts necessary URL features, reducing the reliance on cybersecurity experts and third-party services like page rank or domain age, which traditional systems may depend on.
- **High Accuracy**: The model has demonstrated exceptional performance, achieving a notable 96.04% accuracy on benchmark datasets, which is a significant improvement over traditional phishing detection systems.
- **Adaptability to Evolving Threats**: The model's design allows it to learn from the data it processes, making it potentially more adaptable to the continuously evolving tactics used by phishers, unlike traditional systems that may require manual updates to stay effective.

Limitations of the study include:

- **Real-World Application**: The paper does not discuss the model's performance in real-world scenarios where phishing tactics are constantly evolving.
- **Performance on Other Datasets**: The model's performance has been validated on two benchmark datasets, but it's unclear how it would perform on other datasets or in different contexts.
- **Feature Reliance**: The model's reliance on URL and character-level TF-IDF features may limit its ability to detect phishing attempts that use other tactics.
- **Computational Resources**: The paper does not discuss the computational resources required to implement the model, which could be a potential limitation for some users.

The proposed model has several limitations in terms of scalability and performance.

- Firstly, the model's reliance on Long Short-Term Memory (LSTM) networks can lead to computational inefficiency. LSTM networks are known for their high computational and memory requirements, which can limit the model's scalability when dealing with large datasets or in real-time applications.
- Secondly, the model's two-phase approach, which involves training features on a base machine learning classifier and then employing a two-layered stacked-based LSTM network, can be time-consuming and computationally intensive. This could potentially limit the model's performance in real-time phishing detection scenarios.
- Lastly, while the model is designed to operate without prior phishing-specific feature knowledge, this could also be a limitation. The model may struggle to accurately detect new or sophisticated phishing attempts that exploit features not considered in the model's training.

VI. Implications for Future Research

- **Model Generalization**: The model's ability to operate without prior phishing-specific feature knowledge suggests that future research could explore the development of more generalized models that can adapt to various types of cyber threats without extensive retraining.
- **Deep Learning Techniques**: The success of the LSTM-based model indicates that deep learning techniques have significant potential in cybersecurity applications. Future research could further investigate the integration of different neural network architectures and their effectiveness in threat detection.
- **Feature Extraction**: The use of character-level TF-IDF features and URL analysis in the model demonstrates the importance of feature extraction in phishing detection. Research could focus on identifying new features and methods of extraction to improve detection rates.

- **Stack Generalization**: The two-phase approach used in the model, which combines machine learning classifiers and LSTM networks, showcases the benefits of stacked generalization. Future studies could explore other combinations of algorithms and models to enhance predictive performance.
- **Benchmark Datasets**: The use of benchmark datasets for model validation in this study underscores the need for comprehensive and up-to-date datasets in cybersecurity research. Future work could involve creating and maintaining datasets that reflect the latest threat landscapes.

VII. Main Contribution to Cybersecurity

- **Prior Feature Knowledge Independence**: The model's ability to learn from URL strings as character sequences without the need for prior feature knowledge simplifies the detection process and makes it more adaptable to new and unknown phishing attacks.
- **Strong Generalization Ability**: The model's use of URL character-based features for robust generalization and check-side accuracy, combined with the integration of multi-level features in the neural network, contributes to its effectiveness in generalizing across different phishing threats.
- **Independence from Cybersecurity Experts and Third-Party Services**: By autonomously extracting necessary URL features, the model reduces reliance on cybersecurity experts and third-party services, making it a self-sufficient tool for phishing detection.
- **Enhanced Detection Accuracy**: The model's experimental validation on benchmark datasets demonstrated exceptional performance, with a notable accuracy of 96.04%, which is higher than that of existing studies.
- **Contribution to Symmetry in Information Security**: The research adds to the discourse on symmetry and asymmetry in information security by providing a model that can symmetrically learn and detect phishing URLs, thereby enhancing network security against evolving cyber threats.

VIII. Potential Future Research Directions

- **Improving Generalization Ability**: The model has a strong generalization ability, utilizing URL character-based features for robust generalization and check-side accuracy. Future research could focus on further enhancing this ability, particularly in the context of evolving phishing tactics and techniques.
- **Enhancing Independence from Cybersecurity Experts and Third-Party Services**: The model autonomously extracts necessary URL features, eliminating reliance on cybersecurity experts and third-party services. Future research could explore ways to further improve this independence, potentially through the development of more sophisticated feature extraction techniques.
- **Optimizing the Stacked Generalization Model**: The model uses a two-phase stacked generalization model, with the first phase generating a mean prediction and the second phase utilizing a two-layered LSTM-based stack generalized model optimized for premier prediction in phishing site detection. Future research could focus on optimizing this model, perhaps through the use of different machine learning algorithms or techniques.
- **Enhancing Accuracy**: While the model has demonstrated high accuracy in detecting phishing sites, future research could focus on ways to further enhance this accuracy, particularly in the context of zero-day attacks and other advanced phishing techniques.
- **Expanding the Model to Other Cybersecurity Applications**: The model could potentially be adapted for other cybersecurity applications beyond phishing detection.