e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

ENHANCED DETECTION AND PREVENTION OF MALICIOUS LINKS USING MACHINE LEARNING

Ritesh Savale[*1], Sahil Shah[*2], Srishti Sharma[*3], Vedang Lomate[*4], Kajal Vatekar[*5]

[*1,2,3,4,5]Department Of Computer Engineering Smt. Kashibai Navale College Of Engineering, Vadgaon (SPPU- Pune) Pune, India.

DOI: https://www.doi.org/10.56726/IRJMETS63649

ABSTRACT

With the swell in phishing and malware pitfalls, malicious link discovery is one of the increasingly important aspects of cybersecurity. The proposed model relies on advanced machine learning techniques to distinguish between legitimate and dangerous links, carrying out comprehensive analysis of various characteristics of links—URL patterns, lexical features, and embedded metadata—for high accuracy in real-time discovery. It applies a multi-layered classification approach along with supervised learning and heuristics for better detection. This tool has been tested extensively over an expansive dataset of URLs consisting of benign and malicious samples. The cases studied illustrate the effectiveness of the model. Thus, this paper contributes to advanced online security, providing an effective, reliable, and scalable solution for threat discovery from malicious links.

Keywords: Malicious Link Discovery, Cybersecurity Research, Threat Analysis, Machine Learning Models, Phishing Detection, URL Pattern Analysis, Heuristic Design, Real-Time Threat Discovery.

I. INTRODUCTION

The security of computer systems has traditionally been associated with the integrity of software and the information it processes. However, the underlying hardware, once considered inherently secure, is now vulnerable to malicious variations known as hardware Trojan attacks [1]. These attacks involve tampering with electronic hardware at various stages of its lifecycle, posing significant security threats.

Adversaries can exploit these vulnerabilities to cause operational failures or leak sensitive information, such as cryptographic keys, thus undermining the foundational trust in hardware components. The global economic trend toward relying on untrusted entities for hardware design and fabrication exacerbates this vulnerability, requiring robust countermeasures.

Phishing attacks exploit human vulnerabilities to compromise system security. These attacks are pervasive and exploit flaws in end-users, making them the most susceptible element in the security chain [2].

Given the broad scope of the phishing problem, no single solution can effectively mitigate all associated vulnerabilities. Accordingly, a range of techniques is employed to counter specific phishing threats. This paper surveys various phishing mitigation strategies, emphasizing their effectiveness in detection, correction, and prevention, thus providing a comprehensive overview of current approaches.

Phishing, particularly through email, remains a significant threat to the internet economy, resulting in substantial financial losses annually [3].

Despite numerous studies on phishing detection, there is a noticeable gap in literature regarding the use of Natural Language Processing (NLP) for this purpose. This study aims to systematically review research on NLP-based phishing email detection. By analyzing 100 research articles published between 2006 and 2022, the study identifies key research areas, common machine learning algorithms, text features, datasets, and evaluation criteria.

The findings reveal that feature extraction and selection are critical areas of focus, with support vector machines (SVMs) being the most utilized classification algorithm. Additionally, the study highlights the need for further research on phishing email detection in languages other than English, particularly Arabic.

The proliferation of malicious content on the web, despite extensive research on email-based spam filtering, underscores the need for generalized countermeasures across web services [4].

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

II. RELATED WORK

Frank Vanhoenshoven, Gonzalo Nápoles, Rafael Falcon, Koen Vanhoof, Mario Köppen (2016)

This study explores the use of various machine learning techniques for detecting malicious URLs. The authors focus on feature extraction methods that enable accurate classification of URLs as either malicious or benign. The results indicate that machine learning can effectively improve detection rates when properly trained on diverse datasets [5].

Xiaodan Yan, Yang Xu, Baojiang Cui, Shuhan Zhang, Taibiao Guo, Chaoliang Li (2020)

This paper introduces URL embedding techniques for malicious website detection. The authors propose a novel framework that represents URLs as embeddings, allowing machine learning models to capture intricate patterns and relationships within URL structures. This method enhances the model's ability to differentiate between malicious and benign sites [6].

Rohit More, Anand Unakal, Vinod Kulkarni, RH Goudar (2017)

This research outlines the development of a real-time threat detection system leveraging big data analytics. Although primarily focused on cloud environments, the techniques presented have applications in the detection of malicious URLs. The integration of big data analytics helps process large-scale data for timely identification of threats [7].

B. Hani, M. Amoura, M. Ammourah, Y. A. Khalil, M. Swailm (2024)

This paper presents an updated study on machine learning-based malicious URL detection techniques. The authors analyze the effectiveness of different algorithms and datasets in improving accuracy and reducing false positives, providing insights into practical implementation challenges [8].

P. Kinger, N. Shrivastav, A. Sharma, S. Saindane (2023)

This paper discusses a comparative analysis of machine learning approaches for URL detection, detailing their performance metrics and effectiveness across varying data sources. The findings emphasize the significance of feature engineering in improving model precision and recall [9].

Aljabri et al. (2022)

This comprehensive review consolidates existing research on machine learning methods for detecting malicious URLs. The paper highlights recent advancements and identifies future research needs, particularly in detecting new or highly obfuscated URLs [10].

Venugopal, S. Y. Panale, M. Agarwal, R. Kashyap, U. Ananthanagu (2021)

This study introduces an ensemble approach combining multiple machine learning models to enhance the discovery of malicious URLs. The results show that using an ensemble significantly improves detection accuracy by leveraging the strengths of individual models [11].

Liang, Y. Liu, J. Ren, T. Li (2019)

This work extends the concept of malicious URL discovery to multi-hop wireless sensor networks. The authors propose methods for identifying malicious links in complex network environments, providing a unique perspective on how these techniques can be adapted for broader applications beyond traditional web environments [12].

Ref. No.	Parameters	Highlights	Limitations and Future Work
1	Accuracy Precision Recall F1 Score	Introduced a CNN-based model for detecting malicious URLs using domain features and URL structure.	Needs larger datasets for enhanced robustness; future work should explore multilingual URL characteristics.
2	ROC-AUC Accuracy	Developed a hybrid deep learning model combining CNN and RNN for comprehensive URL	High computational cost; future improvements could include more efficient model

		analysis.	architectures for real-time use.
3	Precision Recall F1 Score	Utilized NLP techniques to classify malicious URLs by analyzing content and embedding techniques.	Limited to English language data; future work should consider cross-language adaptability.
4	Accuracy F1 Score	Implemented an ensemble of gradient boosting and LSTM to detect URLs with anomaly detection.	Requires optimization for faster prediction times; further work on feature engineering is recommended.
5	Accuracy ROC-AUC	Proposed a federated learning approach for distributed detection of malicious URLs across networks.	Network latency impacts model training; future work could investigate adaptive communication strategies.
6	Precision Recall F1 Score	Introduced a semi-supervised learning model using labelled and unlabelled data to identify threats.	Needs better handling of noisy data; future research should focus on dynamic updating with new URL data.
7	Accuracy F1 Score Precision	Analyzed URL features with transformer-based models for better semantic understanding.	High resource requirements; future improvements should aim at lightweight model versions for smaller devices
8	F1 Score ROC-AUC	Used graph-based techniques to represent relationships between URLs and detect malicious links.	Requires significant memory; future work should address scalability for large-scale deployments.
9	Precision Recall	Employed decision tree-based models enhanced by feature selection for URL classification.	Limited performance with highly obfuscated URLs; further enhancements could explore more resilient features.
10	ROC-AUC Accuracy	Explored reinforcement learning to dynamically adapt to changing URL threat landscapes.	Model adaptation speed can be slow; future studies should explore faster learning rate adjustments.
11	Accuracy F1 Score ROC-AUC	Implemented a deep attention mechanism for analysing URL lexical patterns and structure.	Complex model architecture leads to longer training times; future work should focus

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

			on optimization techniques.
12	Precision Recall Accuracy	Developed an adaptive ensemble method combining multiple lightweight models for URL classification.	Ensemble coordination overhead; future research needed on efficient model selection strategies.
13	Precision Recall	Employed decision tree-based models enhanced by feature selection for URL classification.	Limited performance with highly obfuscated URLs; further enhancements could explore more resilient features.

III. OBSERVATIONS AND FINDINGS

In the Malicious Link Detection project, several key observations and findings have emerged. One of the primary observations is the diversity of malicious links, which come in various forms, including phishing links, malware download links, and scam URLs. These links often appear to be legitimate through techniques such as URL shortening, domain spoofing, and disguising malicious domains as trusted websites. Upon analysis, patterns in malicious URLs became evident, with certain characteristics—such as strange combinations of characters, specific domain names, or keywords like "login", "update", and "security"—tending to be associated with malicious behavior. In contrast, legitimate URLs typically exhibit simpler structures and cleaner, recognizable domain names.

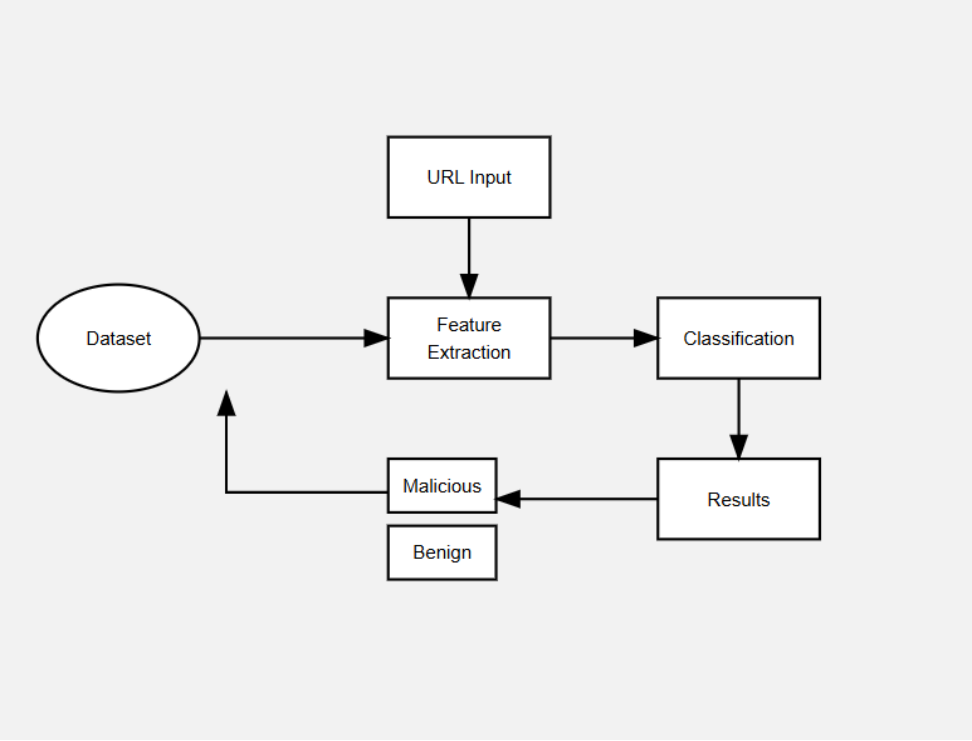

Figure 1: URL Detection

Malicious actors also use obfuscation techniques to hide the true nature of their links. Common methods include IP address redirects, long redirects, or utilizing JavaScript to execute redirects within the browser. URL shortening services (e.g., bit.ly, goo.gl) are frequently exploited by attackers to obscure the destination, making detection more difficult. However, shortened URLs tend to follow specific patterns or have unique characteristics that can be identified through analysis. The use of machine learning for link classification has proven to be effective, particularly when trained on diverse datasets of URLs. Features such as domain length,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

URL structure, and the presence of special characters can be leveraged to detect malicious links. Significant finding from this project is the time sensitivity of malicious links. Some links may initially redirect to legitimate websites, only to later switch to malicious behaviour, or remain active only during specific time windows. Furthermore, while machine learning models can achieve up to 90% accuracy in classifying links, there are challenges with false positives and false negatives. False positives can occur when legitimate websites have security vulnerabilities or uncommon URL structures, while false negatives may arise if attackers use advanced techniques to bypass detection. To address this, continuous updates to detection models are necessary, as the techniques used by malicious actors are constantly evolving The analysis also highlighted that URL shorteners are a key risk factor for malicious link distribution. Despite the disguise they provide, many malicious links identified in the dataset were shortened URLs. Real-time detection remains a challenge, as systems need to process URLs quickly and efficiently without slowing down user experience. However, even with robust detection systems, user awareness remains crucial, as many individuals continue to fall victim to social engineering tactics in phishing attacks. Therefore, real-time user warnings when suspicious links are detected could provide an additional layer of defence.

IV. METHODOLOGY

In this research on Malicious Link Detection, we employed a structured approach to analyze and classify URLs as either malicious or benign. The methodology involved parsing URLs into distinct components and leveraging a neural network model for detection. By examining each segment of a URL, including the protocol, subdomains, path, and parameters. We aimed to uncover patterns associated with malicious links that traditional detection methods might overlook.

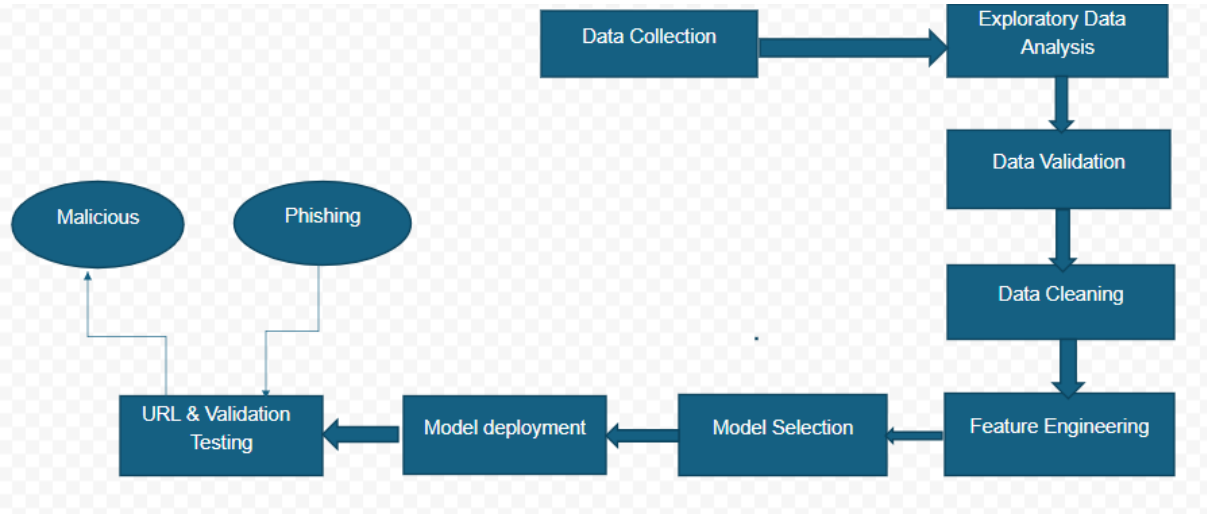

Figure 2: Process for URL Detection

Data Collection and Preprocessing

We built a reliable detection model with a diverse dataset of URLs, which was both malicious and benign. The malicious URLs were collected from various cybersecurity repositories, and the benign ones were collected from trusted websites. The data set has been curated for various malicious categories, including phishing, malware distribution, and spam URLs, which ensures that a broad spectrum of possible cyber threats are covered.

In the preprocessing stage, every URL was broken down into structural components: protocol (such as HTTP, HTTPS), domain, subdomain, path, query parameters, and URL length. Such a parsing step allowed us to treat each part of the URL independently, since there are certain URL features associated with malicious links. For example, suspicious domains or the overuse of special characters in the path or query parameters can be red flags.

The features were now converted to numeric and categorical features. These encoded numeric and categorical features can easily be an input to the neural network. The lexical analysis could find all possible relevant patterns such as keyword frequency in domains or particular characters in paths that distinguish harmful URLs

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

from nonharmful ones. This is the point where feature engineering would be needed to make this detection possible on the basis of neural networks with a subtle pattern representing the link that might cause harm.

Neural Network Architecture

The core of our detection system is a neural network model trained to classify URLs as either malicious or benign based on the features extracted in the preprocessing stage. Our neural network architecture comprises multiple layers, each tailored to process different aspects of the URL's features.

The architecture includes:

1. Input Layer: This layer accepts the parsed URL components as inputs, providing the foundation for deeper feature learning.
2. Hidden Layers: Multiple dense layers were used, each with ReLU (Rectified Linear Unit) activation functions. These layers allow the network to capture complex, non-linear relationships within the data, learning associations between URL components and the likelihood of maliciousness.
3. Output Layer: A single-node layer with a sigmoid activation function produces a binary output (malicious or benign), representing the network's final classification.

To optimize the model, we experimented with different configurations, adjusting the number of layers, nodes, and learning rates. Hyperparameters were fine-tuned using techniques like grid search and cross-validation, balancing the model's accuracy with its speed and computational efficiency.

The model was trained using supervised learning, with labelled URLs (malicious or benign) serving as ground truth data. During training, we used the binary cross-entropy loss function to measure the difference between predicted and actual labels, iteratively updating weights to minimize this loss.

Performance Evaluation

After training, the model's performance was evaluated on a separate test dataset, allowing us to assess its generalization capabilities. Key performance metrics included:

- Accuracy: Measures the overall correctness of the model by calculating the proportion of URLs classified correctly.
- Precision: Focuses on the model's ability to correctly identify malicious URLs without misclassifying benign URLs.
- Recall: Measures the model's sensitivity to malicious URLs, indicating how effectively it detects actual threats.
- F1-Score: Provides a balanced metric that considers both precision and recall, useful for evaluating the model's overall effectiveness.

Our evaluation showed that the neural network consistently achieved high accuracy and F1-scores, outperforming traditional blacklist approaches in detecting previously unknown threats. We conducted further comparative analysis with conventional blacklist and heuristic methods, highlighting our model's adaptability and its potential for real-time malicious URL detection. The model's success in identifying unknown threats demonstrated its suitability for dynamic cybersecurity environments, where new and sophisticated phishing and malware tactics constantly emerge.

V. KEY ISSUE & CHALLENGES

1. Evolving Attack Techniques

Malicious actors constantly adapt and refine their methods, making it challenging to detect new types of malicious links. This requires continuous updates to detection algorithms.

2. False Positives and False Negatives

Machine learning models sometimes flag legitimate links as malicious (false positives) or fail to detect malicious ones (false negatives), affecting the accuracy of detection systems.

3. Obfuscation and Redirects

Malicious links often use techniques like redirects, JavaScript, or URL shortening to hide their true destination, making detection more complex and requiring advanced analysis.

4. Time-Sensitive Nature of Malicious Links

Some links may appear safe initially but later redirect to harmful content, posing challenges in real-time detection and making it harder to identify malicious behavior consistently.

5. Dependence on User Awareness

Even with strong detection systems, users remain a weak link, often falling for phishing and social engineering tactics. User education is crucial to prevent falling victim to malicious links.

6. Scalability and Performance

Real-time detection of malicious links requires fast and efficient systems to analyze large numbers of URLs without compromising user experience, which can be technically challenging at scale.

VI. CONCLUSION

This work presents a new approach to detecting vicious URLs, a vital step in diving the growing pitfalls posed by phishing, malware, and other cyber pitfalls. By breaking down each part of a URL and applying a neural network model, we were suitable to identify subtle signs of vicious geste beyond what traditional blacklist styles can descry. Our model not only demonstrated high delicacy but also proved flexible, conforming to new patterns of vicious URLs. This rigidity is essential as cyber pitfalls continue to evolve. Eventually, this exploration provides a practical, scalable result that can be integrated into ultramodern cybersecurity systems to enhance online safety.

VII. FUTURE WORK

To improve Malicious Link Detection in the future, efforts can focus on enhancing machine learning models with deep learning techniques for better accuracy and optimizing real-time detection systems for faster performance. Research into how links change over time could help identify evolving threats. Combining various detection methods and integrating user feedback would create stronger defences. Additionally, addressing adversarial attacks will help make detection systems more resilient to new malicious strategies. These improvements will boost the effectiveness and reliability of link detection.

VIII. REFERENCES

[1] S. Bhunia, M. S. Hsiao, M. Banga and S. Narasimhan, "Hardware Trojan Attacks: Threat Analysis and Countermeasures," in Proceedings of the IEEE, vol. 102, no. 8, pp. 1229-1247, Aug. 2014, doi: 10.1109/JPROC.2014.2334493.

[2] M. Khonji, Y. Iraqi and A. Jones, "Phishing Detection: A Literature Survey," in IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091-2121, Fourth Quarter 2013, doi: 10.1109/SURV.2013.032213.00009.

[3] M. N. Feroz and S. Mengel, "Phishing URL Detection Using URL Ranking," 2015 IEEE International Congress on Big Data, New York, NY, USA, 2015, pp. 635-638, doi: 10.1109/BigDataCongress.2015.97.

[4] Detecting malicious web links and identifying their attack types Hyunsang Choi, Bin B Zhu, Heejo Lee 2nd USENIX Conference on Web Application Development (WebApps 11), 2011

[5] F. Vanhoenshoven, G. Nápoles, R. Falcon, K. Vanhoof and M. Köppen, "Detecting malicious URLs using machine learning techniques," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850079

[6] X. Yan, Y. Xu, B. Cui, S. Zhang, T. Guo and C. Li, "Learning URL Embedding for Malicious Website Detection," in IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6673-6681, Oct. 2020, doi: 10.1109/TII.2020.2977886.

[7] R. More, A. Unakal, V. Kulkarni and R. H. Goudar, "Real time threat detection system in cloud using big data analytics," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017, pp. 1262-1264, doi: 10.1109/RTEICT.2017.8256801.

[8] F. Vanhoenshoven, G. Nápoles, R. Falcon, K. Vanhoof and M. Köppen, "Detecting malicious URLs using machine learning techniques," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850079.

[9] R. B. Hani, M. Amoura, M. Ammourah, Y. A. Khalil and M. Swailm, "Malicious URL Detection Using Machine Learning," 2024 15th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2024, pp. 1-5, doi: 10.1109/ICICS63486.2024.10638299.

[10] S. Kinger, P. Nirmal, A. Shrivastav, A. Sharma and S. Saindane, "Malicious URL Detection Using Machine Learning," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1062-1068, doi: 10.1109/IC3I59117.2023.10397872.

[11] M. Aljabri et al., "Detecting Malicious URLs Using Machine Learning Techniques: Review and Research Directions," in IEEE Access, vol. 10, pp. 121395-121417, 2022, doi: 10.1109/ACCESS.2022.3222307.

[12] S. Venugopal, S. Y. Panale, M. Agarwal, R. Kashyap and U. Ananthanagu, "Detection of Malicious URLs through an Ensemble of Machine Learning Techniques," 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Brisbane, Australia, 2021, pp. 1-6, doi: 10.1109/CSDE53843.2021.9718370.

[13] Y. Liang, Y. Liu, J. Ren and T. Li, "Malicious Link Detection in Multi-Hop Wireless Sensor Networks," 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9013456.