A black background with a black square

Description automatically generated with medium confidence

Papers Review

Submitted to:  
  
Dr. Kami Vaniea,  
Department of Electrical and Computer Engineering,  
University of Waterloo

Submitted by:

Mohammad Kaosain Akbar  
*Prospective PhD Applicant, Electrical and Computer Engineering,  
University of Waterloo*

Contents:

1. List of Selected Papers

2. Paper summary (P1, P2)

3. Strength and Weakness (P1, P2)

4. Recommendation (P1, P2)

5. Personal Learning Feedback (P1, P2)

6. Disclosure and Appendix.

**1. List of Papers**

I have selected the following papers to review from your publication list.

1. Phishing codebook: A structured framework for the characterization of phishing emails
2. Using clustering algorithms to automatically identify phishing campaigns

I selected these two papers because they approach phishing from different but connected angles. One focuses on understanding the structure and patterns of phishing emails through a detailed framework, while the other studies how clustering techniques can automatically identify phishing campaigns. Together they show how human understanding and machine learning can complement each other in addressing cybersecurity challenges. Reviewing them helps me learn how researchers combine behavioral insights with technical methods to improve online safety and user awareness.

**(P1) Phishing codebook: A structured framework for the characterization of phishing emails**

**Paper Summary**

The following is the summary of the paper.

1. The paper introduces the Phishing Codebook, a structured framework designed to capture how people understand and respond to phishing emails rather than just focusing on technical details.
2. Researchers analyzed 503 phishing emails through multiple rounds of qualitative coding, refining their observations until they reached consistent results among different reviewers.
3. The final codebook includes eight major traits such as sender information, urgency, threats, and requested actions, representing the main cues humans notice when judging email legitimacy.
4. These traits reveal how attackers design their emails to exploit trust, pressure users, or imitate familiar organizations to gain sensitive information.
5. The framework can help organizations cluster similar phishing emails into campaigns, making it easier to identify large scale attacks quickly.
6. It also supports the creation of personalized user guidance tools that explain why an email looks suspicious, improving user awareness and decision making.
7. This work is novel because it combines human insight with structured analysis, laying the foundation for future hybrid systems that connect human reasoning with AI based phishing detection.

**Strength and Weakness of the Paper**

***Strengths:***

1. The study introduces a clear, structured, and human-centered framework that captures how users perceive phishing cues, bridging psychology and cybersecurity.
2. The iterative qualitative coding process ensures high reliability and reproducibility, with strong agreement among multiple coders.
3. The resulting codebook can be reused for both research and practical applications, such as training users or improving explainable phishing detection systems.

***Weakness:***

1. The dataset is limited to publicly available phishing emails, which may not represent the full diversity of real-world phishing attacks.
2. The coding relies heavily on manual interpretation, making it time-consuming and less scalable without automation.
3. The framework focuses mainly on email content and ignores technical elements like links, attachments, or metadata that also influence detection accuracy.

**Recommendation to improve the Paper**

In the following I have provided my personal recommendation for the paper.

* 1. The framework could be integrated with real time email filtering systems to test how well qualitative traits help in live detection.
  2. A follow up study could compare human coded results with AI generated annotations to evaluate consistency and scalability.
  3. Conducting user studies to evaluate how well the codebook improves people’s ability to recognize phishing emails could further validate its practical impact.

**Personal Learning Feedback**

As a machine learning researcher, I had no idea that such a qualitative methodology could be applied to build a structured framework for understanding phishing emails. This paper showed me how human-centered coding can capture behavioral and psychological aspects that purely technical models often miss. It was interesting to see how iterative qualitative coding can create a reproducible structure that IT and security experts can later use for training, detection, and even automation. It broadened my perspective on how human insight and machine learning can complement each other in cybersecurity research.

**(P2) Using clustering algorithms to automatically identify phishing campaigns**

**Paper Summary**

The following is the summary of the paper.

1. The paper explores how phishing emails can be automatically grouped into campaigns that share similar patterns or content. This helps in understanding large-scale phishing activities more effectively.
2. It aims to reduce the heavy manual effort of IT and security teams who currently analyze each phishing email separately to find links between them.
3. The study uses two unsupervised clustering methods, Mean Shift and DBSCAN, to identify natural groupings of related phishing emails without any prior labeling.
4. About seventy features are extracted from each email, including details like sender information, URL structures, message content, time of sending, and attachments. These features capture both technical and behavioral traits of phishing attempts.
5. The authors use Silhouette and Homogeneity scores to evaluate the quality of the clusters. These measures show how closely emails in a cluster are related and whether each cluster represents one distinct campaign.
6. The Mean Shift algorithm performs better than DBSCAN, producing cleaner and more accurate clusters that align closely with real phishing campaigns.
7. The research concludes that clustering can play a key role in improving phishing response systems, allowing analysts to detect campaigns faster, spot emerging threats, and manage large volumes of phishing reports more efficiently.

**Strength and Weakness of the Paper**

***Strengths:***

* 1. The paper presents a practical and scalable approach to identifying phishing campaigns using real-world data. Other researchers usually focuses towards profiling but the authors here worked towards finding clusters based on campaigns.
  2. It uses a various set of features from multiple aspects of emails, giving the clustering model a comprehensive understanding of phishing patterns.
  3. The results are validated both through evaluation metrics and human involvement, which strengthens the reliability and real-world usefulness of the approach.

***Weakness:***

* 1. The study relies on only two clustering algorithms and does not explore newer or more advanced machine learning methods that could improve accuracy.
  2. The dataset, while large, is limited to a single security company, which may reduce the generalizability of the results to other environments.
  3. The evaluation focuses mainly on clustering performance and but lacks explainability. Moreover, the computational resources used were not properly mentioned

**Recommendation to improve the Paper**

In the following I have provided my personal recommendation for the paper.

* 1. Future work could explore advanced or hybrid machine learning methods, such as deep clustering or semi-supervised approaches, to improve the accuracy and adaptability of phishing campaign detection.
  2. Utilization of NLP based approaches might also contribute towards accuracy improvement
  3. The study could be extended to include data from multiple organizations or regions to test how well the model generalizes across different phishing patterns and email infrastructures.

**Personal Learning Feedback**

From this paper, I learned how clustering can be used to identify phishing campaigns by grouping similar emails together. It was the first time I learned about the Mean Shift algorithm and how it can automatically determine the number of clusters based on data density. I also discovered that, apart from DBSCAN, there is another efficient algorithm like Mean Shift that can find clusters without predefining their number. Overall, the paper helped me understand how unsupervised learning can be applied in cybersecurity to make phishing detection more efficient.

**Appendix.**

Disclosure: No online large language models were used in writing this report.

I have also attached the annotated version of the paper which I utilized to write the report