CERIAS

The Center for Education and Research in Information Assurance and Security

Expanding Phish-NET: Detecting Phishing Emails Using Natural Language Processing

Student: Bryan R. Lee & Gilchan Park / Advisor: Julia M. Taylor

Abstract

Phishing is a potentially disruptive action that causes substantial financial loss to Internet users. One of the most popular ways of carrying out a phishing attack is Through email. Many businesses use typical spam filters such as blacklist-based or URL analysis techniques to protect users from some potentially malicious emails, but these alone are not enough. There have been quite a few attempts at creating a reliable, robust phishing email detection systems based on analyzing the content of the emails. For example, CANTINA[3], phishGILLNET[1] and Phish-Net [2] are proposed methods for content-based phishing detection.

Phish-Net is a phishing detection utility that analyzes three parts of an email to determine whether or not it contains a phishing attack: the header, the text, and any links the email contains. The purpose of this research is to expand the text analysis portion of Phish-Net in determine whether it is possible to improve its email analysis capabilities. The text analysis portion of Phish-Net takes into account actionable verbs that tempt the user into performing an action.

In this study, the new algorithm includes not only actionable verbs, but also other parts of speech so that it can catch any other actionable words in phishing emails.

Process

- ✓ **Text-score**: the analysis score of the email **itself**.
- ✓ **Context-score**: the analysis score of the email **comparing with the previous ones**.
- ✓ A score of **1 represents phishing** and **0 stands for legitimate.**

< The Previous algorithm >

Find *actionable* verbs.

- Actionable verbs : verbs that tempt the user into performing an action.

E.g., click, follow, visit, go, update, apply, etc.

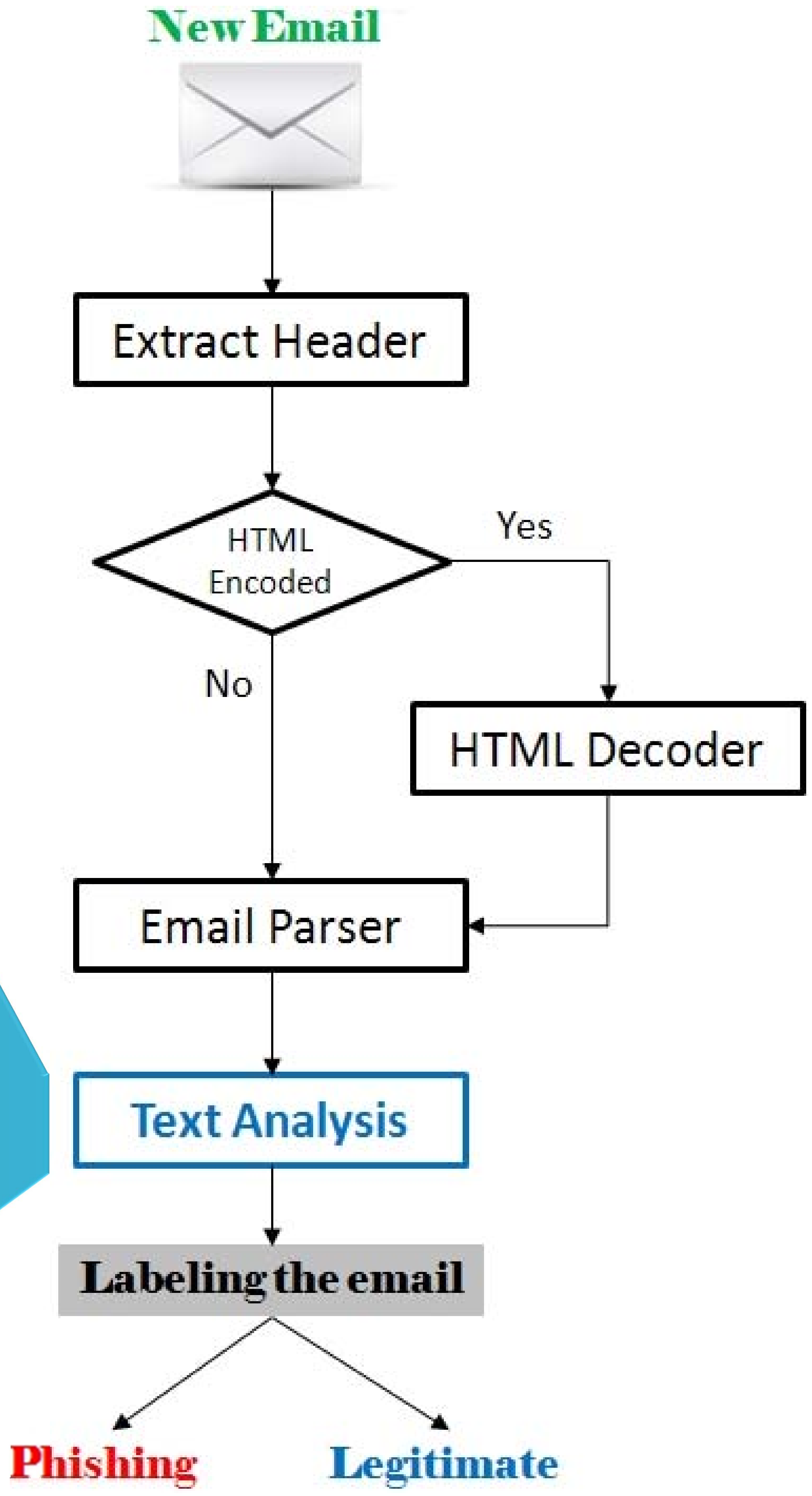
< The New algorithm >

Find *actionable* words.

- Actionable words: any parts of speech that tempt the user into performing an action.

E.g., click, clicking, clicked, clickable, follow, followed, following, update, updated, updating, apply, applied, etc.

- Synonyms of chosen words are also taken into consideration.
- Text-score of each actionable word is generated.
  Then, Text-score(e) = Max{score(w) | w ∈ e}
- Compare the new email with the previous ones.
  Then, Context-score(e) is generated.
- Combine Text-score(e) & Context-score(e) = Final-score(e)



Expected Results

- Higher phishing detection rate than the previous algorithm by considering not only verbs, but also other part of speech.
  In the previous results, only 51.3% (without context) and 65.8% (with context) of phishing detection was successful by text analysis.
- Examination of what factors increase a false positive rate.
  In the previous results, only 79% (without context) and 85.1% (with context) of legitimate emails was marked as legitimate by text analysis.

References

[1] Ramanathan, V., & Wechsler, H. (2012). phishGILLNET—phishing detection methodology using probabilistic latent semantic analysis, AdaBoost, and co-training. *EURASIP Journal on Multimedia and Information Security*, *2012(1)*, 1-22.

[2] Verma, R., Shashidhar, N., & Hossain, N. (2012). Detecting Phishing Emails the Natural Language Way. *Computer Security–ESORICS 2012*, 824-841.

[3] Zhang, Y., Hong, J. I., & Cranor, L. F. (2007). Cantina: a content-based approach to detecting phishing web sites. In *Proceedings of the 16th international conference on World Wide Web. ACM,* 639-648.

CERIAS

PURDUE UNIVERSITY