University of Piraeus
Systems Security Laboratory (SSL)
Department of Digital Systems

Evolution of phishing email attacks and sophisticated Machine Learning detection solutions

Prof. Christos Xenakis

1st Open Annual Workshop on Future ICT

Presentation Outline

- Social Engineering attacks
- Phishing attacks analysis
- Phishing email detection solutions
- Literature overview and limitations
- Proposed phishing email detection methodology
- Experimental approach and results
- Conclusion

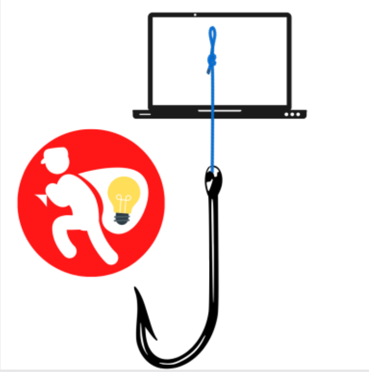
Social Engineering Attacks

- **Psychological manipulation** (or human hacking) to lure victims:
  - ➢ **Reveal sensitive information** (e.g., usernames/passwords)
  - ➢ **Click on malicious links**
  - ➢ **Download malicious attachments**
- Causes of **data breaches**
  - ➢ 70% to 90% through **social engineering**
  - ➢ 20% to 40% through **unpatched software**
- The most common way of applying social engineering is **PHISHING**

What is Phishing?

Phishing, is a **social engineering** technique

- It typically refers to use digital and online means: **email, websites, instant messaging, text messages**
- The intention is to maliciously gain **personal** or **financial** information
- By pretending to **be a trustworthy entity**.
- Most methods intend to **deceive end users** to willingly provide information
- Moreover, they intend to **get access to their device** without being aware of it.



University of Piraeus
Systems Security Laboratory (SSL)

Phishing over the years

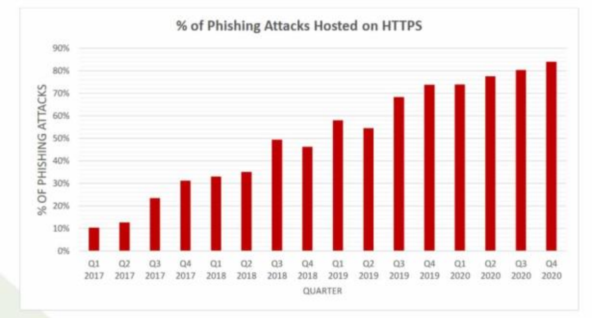
1987	1996	2001	2011	2020
1	2	3	4	5
A phishing technique was described theoretically	AOHell includes phishing functionality through social engineering	The first attack took place	RSA staff were successfully phished, leading to the master keys for all RSA security tokens	No.1 Cybercriminal strategy to begin a large-scale attack (e.g., APT)

Evolution of Phishing Attacks

- **Targeted phishing email attacks** instead of mass phishing campaigns
- **Spear phishing** campaigns targeting **executives, managers, administrators** of an organization
- **AI-as-a-service** for the **generation** of phishing emails
- **Hijacking** an email **reply chain**
  - Account takeover (data breach or earlier compromise)
  - Insert a phishing email to an existing conversation.

Recent Phishing Attack Statistics

- **26.2 billion €** of cumulative losses **in 2019** with **Business E-mail Comprise attacks**.
- **42,8%** of all **malicious attachments** were **Microsoft Office documents**.
- **30%** of phishing messages were **delivered** on **Mondays**.
- **32,5%** of all the **e-mails** used the keyword **'payment'** in the e-mail subject.
- **667% increase** in phishing scams in only **1 month** during the COVID-19 pandemic.
- **96%** of phishing attacks performed via **e-mails**
- In 2020 **75%** of organizations faced a phishing attack
- **64%** rise compared to **2019**



University of Piraeus
Systems Security Laboratory (SSL)

Phishing Email Detection Solutions

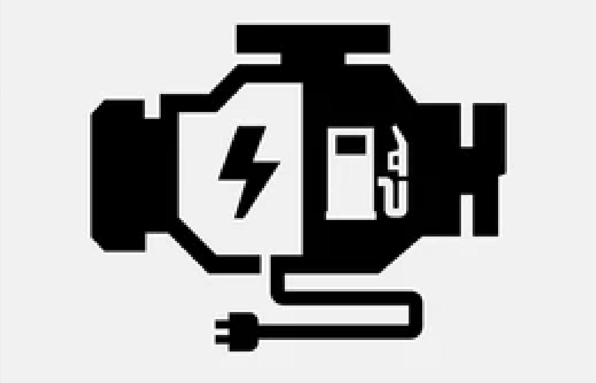
- Current phishing **e-mail detection solutions** are mostly based on **Machine/Deep Learning**
- Can be broadly grouped on two categories based on the extracted traits
  - **Content-based** focusing on traits extracted from the contents of emails
    - ✓ Headers
    - ✓ Hyperlinks
    - ✓ Most used words
  - **Text-based** focusing on traits extracted from the emails body text
    - ✓ Hand-crafted
    - ✓ Natural Language Processing methods (e.g., Word2Vec, TF-IDF)

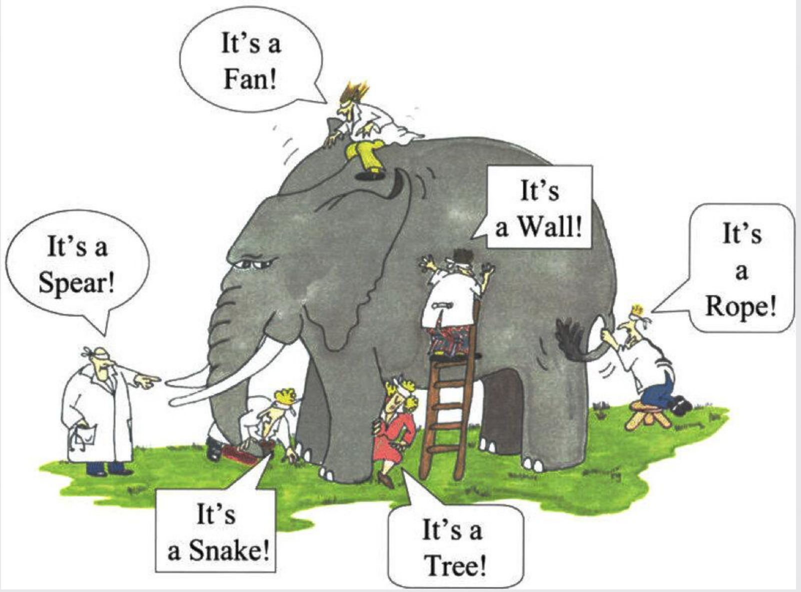
Limitations of the Literature

- Current solutions **cannot cope with the evolution** of phishing email attacks
- **Ensemble Learning** has not employed
- **Hybrid features** have not studied
- **Unrealistic** experiments
- **Obsolete datasets**
- **Poor evaluation experiments**
- **Outdated phishing e-mail samples**

HELPHED: Hybrid Ensemble Learning Phishing Email Detection

- **A novel phishing e-mail detection methodology**
- **Hybrid Features**
  - ➢Content-based
  - ➢Text-based
- **Ensemble Learning**
  - ➢Stacking Ensemble Classification
  - ➢Soft-voting Ensemble Classification





HELPHED - Hybrid Ensemble Learning Phishing Email Detection

- HELPHED includes **6 stages**
- S1: **Email Parsing**
  - The email's **header** and **body fields** are split and stored in **an array** in separated rows along with the email's class (phishing or benign).
- S2: **Content-based feature extraction**
  - 22 content-based features: Body, Syntactic, Header, URL, etc.
- S3: **Pre-processing**
  - Convert to lowercase, remove stopwords and punctuation
  - Replace Hyperlinks with fixed string
  - Tokenization, lemmatization

HELPHED - Hybrid Ensemble Learning Phishing Email Detection

- S4: **Textual Feature Extraction**
  - Word2Vec
- S5: **Feature Selection**
  - Mutual Information
- S6: **Ensemble Classification**
  - Method 1: Stacking Ensemble Learning
  - Method 2: Soft-voting Ensemble Learning

Experimental Approach

- **Dataset: 32,051** benign e-mails, **3,460** phishing e-mails
  - Real e-mails from **publicly available sources**
  - Realistic scenario were the **phishing emails are much less** than benign
- **Experiment 1:** Selection of base learners
  - The performance of **several well-known ML algorithms** were tested on **content-based** and **text-based features** separately
  - **Decision Tree** best performance on **content-based features**
  - **K-Nearest Neighbour** best performance on **text-based features**
- **Experiment 2:** Comparison of traditional ML-based classifiers with HELPHED on the hybrid features

Experimental Results

- HELPHED – **Method 2** achieved the best performance
  - 99.42% F1-score
  - 99.43% Classification Accuracy
- HELPHED **outperformed all the traditional ML-based classifiers**
- Very low training time (0.0313 Sec)

Classifier	F1-score	Accuracy	Precision	Recall	AUC	MCC	Training time (Sec)	Confusion Matrix
LR	0.858	0.9028	0.8609	0.9028	0.503	0.0468	72.998	9611 8 1028 7
GNB	0.8561	0.9014	0.815	0.9014	0.4992	0.0123	0.095	9604 15 1035 0
KNN	0.9398	0.9454	0.9416	0.9454	0.7628	0.6273	**0.01**	9517 102 480 555
DT	0.9806	0.9808	0.9805	0.9808	0.9376	0.8887	0.407	9534 85 120 915
MLP	0.9845	0.985	0.9849	0.985	0.93	0.9117	17.0050	9602 17 143 892
RF	0.9856	0.986	0.9805	0.9808	0.9323	0.918	4.038	9609 10 139 896
Method 1	0.9907	0.9907	0.9906	0.9907	0.969	0.9466	13.705	9580 39 60 975
Method 2	**0.9942**	**0.9943**	**0.9943**	**0.9943**	**0.9714**	**0.967**	0.0313	**9617 2** **59 976**

University of Piraeus
Systems Security Laboratory (SSL)

Conclusion

- Hybrid features better represent the emails
- The combination of **hybrid features** with **ensemble learning** improves the **phishing email detection performance**
- **HELPHED** accomplished **the best performance** on such a large and diverse dataset compared with previous works.

Paper/Year	F1-score (%)	Accuracy (%)	Feature Category	# Benign samples	# Phishing samples
Hamid et al. (2011) [14]	-	92	Content	2364	2230
Moradpoor et al. (2017) [16]	-	92.2	Content	6,656	7,714
Akinyelu et al. (2014) [18]	97.91	98.96	Content	1800	200
Islam et al. (2013) [42]	-	97	Content	N/A	N/A
Smadi et al. (2015) [19]	98.09	98.11	Content	4,559	4,559
Alhogail et al. (2021) [24]	98.5	98.2	Text	4,894	3,685
Gualberto et al. (2020) [26]	99.9	99.9	Text	4,150	2,279
Gualberto et al. (2020) [27]	100	100	Text	4,150	2,279
Fang et al. (2019) [28]	99.33	99.84	Text	7,781	999
Hiransha and Nidhin (2018) [30]	-	96.8	Text	5,088	612
Egozi et al. (2018) [32]	99	-	Text	7,689	1,210
Halgaš et al. (2019) [33]	98.63	98.91	Text	6,951	4,572
Unnithan et al. (2018) [36]	98	97	Text	7,781	997
Unnithan et al. (2018) [38]	-	88.4	Text	8,913	1,087
Yadav et al. (2017) [43]	-	98.02	Content	2,550	500
HELPHED (Soft-voting)	99.41	99.42	Hybrid	32,051	3,460

University of Piraeus
Systems Security Laboratory (SSL)
Department of Digital Systems

Thank you!

Prof. Christos Xenakis

NetPHISH

SECONDO

Cyber Security for Europe