Detecting and Characterizing SMS Spearphishing Attacks

Mingxuan Liu, Yiming Zhang, Baojun Liu, Zhou Li, Haixin Duan, Donghong Sun
What is SMS Spearphishing Attacks?
What is SMS Spearphishing Attacks?

Spearphishing Short Messages:

(1) Contains victim’s personal information
(2) Contains attacker’s contact method

Example of Spearphishing SMS:

Dear Alice: Your reservation for flight MU**** from Fuzhou to Nanjing on February 24 has been forced to be cancelled due to mechanical problems. Please contact China Eastern Airlines 0371-65****19 for refund or ticket change [Eastern Airline]
Spearphishing Attack: A Persistent Problem

Spearphishing attacks are widely spread via Email, social media and even telephone. Billions of dollars have been lost due to spearphishing attacks in recent years!!

From: “Alice”<alice@company.com>
To: “Bob”<bob@victim.com>
Subject: Company

Hi, Bob
Here is the new contract in this link (which is a phishing link.)

Regards,
Alice

However, SMS Spearphishing attack is very different from these attacks via other channels!!
Threat Model of SMS Spearphishing Attack

I. Luring Phase
Attacker: Collect Alice’s personal Info
Victim: Scammed by SMS

II. Exploiting Phase
Attacker: Send Spearphishing SMS
Victim: Scammed by SMS
**Dataset: A Real-World Dataset**


**Example of a spam message log**

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Sender Phone Number</th>
<th>IMEI (hashed)</th>
<th>IMSI (hashed)</th>
<th>IP address of mobile client</th>
<th>Message content</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-02-01 15:34:01</td>
<td>400****3371</td>
<td>HASH_1</td>
<td>HASH_2</td>
<td>172.xxx.xxx.231</td>
<td>尊敬的Name，你的航班MU****，福州飞往南京，已经被取消。退订请联系东方航空，0371-65****19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dear Name, your Flight MU****, from Fuzhou to Nanjing, has been cancelled. Please contact Eastern Airline 0371-65****19</td>
</tr>
</tbody>
</table>

◆ Discussion of **ethical** considerations (shown in our paper).
Detecting Spearphishing Messages
Collect Ground-truth Dataset

Origin Dataset (31.96 M) → 50k random samples → Subset for Manually Inspect → manually label → 2 Researchers Labeled Independently → labeled data → Ground-Truth Dataset with 1,196 spearphishing messages

Key Observations

- “Luring”: Personal information of victims.
  - Name, Flight Info., Plate Nu., Bank Card Nu. and ID Card Nu.

  - URL, CellPhone, Hotline, Phone, QQ, WeChat

- “Syntactic”: Syntactic relationship of personal pronouns.
  - I am NAME (Attacker) from Lottery company. This is my phone number: 17***31.
Detection System of SMS Spearphishing Attack

Step I: Data Pre-Processing

“Sanitize” the text content of fraudulent messages before forwarding them to the next module.

<table>
<thead>
<tr>
<th>Origin Format</th>
<th>Sanitized Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1git</td>
<td>digit</td>
</tr>
<tr>
<td>hell0</td>
<td>hello</td>
</tr>
<tr>
<td>微信</td>
<td>微信 (Wechat)</td>
</tr>
</tbody>
</table>

Step II: Entity Recognition

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Identify Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name Entity Recognition</td>
</tr>
<tr>
<td>Flight, License Plate, License Plate, License Plate</td>
<td>Regular Matching</td>
</tr>
<tr>
<td>URL, CellPhone, Hotline, Phone, QQ, Wechat</td>
<td>Regular Matching</td>
</tr>
</tbody>
</table>
Our detection system reports 90,801 spearphishing messages, with 96.16% precision and 75.33% recall in ground-truth dataset.
Measurement of SMS Spearphishing
# Categories of Spearphishing SMS

<table>
<thead>
<tr>
<th>Category</th>
<th>Example of message content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Scam</td>
<td>Dear Mr/Mrs. NAME, your BOC credit card has been suspended due to an overdue payment. Please contact 86137****765.</td>
</tr>
<tr>
<td>Lawsuit Scam</td>
<td>[CMB] NAME, your credit card has been deemed to be seriously overdue. We will formally prosecute you after 24 hours! Please contact: 86239****7834</td>
</tr>
<tr>
<td>Social Scam</td>
<td>NAME, how are you recently? I changed WeChat. Please add my new WECHAT 266****491.</td>
</tr>
<tr>
<td>Employment Scam</td>
<td>NAME, hello! Your resume in 5.com investment has been accepted. Please contact the QQ: 33****471.</td>
</tr>
<tr>
<td>Insurance Scam</td>
<td>[China Life Insurance] Dear NAME, our company has issued a commercial insurance policy for your car, PLATE. Visit http://***.cc/bX8Vg for details</td>
</tr>
<tr>
<td>Fortune-telling Scam</td>
<td>[Lingji Culture] NAME. Full analysis of fortune in next ten years is coming! See the future and prevent bad luck: <a href="https://s.k****a.cn/dbgc8">https://s.k****a.cn/dbgc8</a></td>
</tr>
<tr>
<td>Gambling Phishing</td>
<td>Hi, NAME. Yabo Sports join hands with Wuhan, register on f8.cn and you can get recharge and get masks, come on, Wu Han!</td>
</tr>
<tr>
<td>Promotional Spam</td>
<td>[Tantan application] NAME, someone loves you, do you want to accept? It is only 3 Km away from you! Click tan****pp.com.</td>
</tr>
<tr>
<td>Other</td>
<td>Mr. NAME, do you want to purchase brand products at a discount? Please contact QQ: 324 *** 558.</td>
</tr>
</tbody>
</table>
Campaign Analysis

Clustering Campaigns

Messages embedded with the same contact or from the same content template are belonging to the same campaign.

• A total of 11,475 campaigns were reported.

Overall Analysis of Campaigns

• Long-Tail Distribution: a few campaigns hold the majority of the market share.

• Interesting Strategies.
Strategies of Spearphishing: Test-Sending

4 Campaigns employ “test-sending” strategy, which totally sent 539 test messages and 9,062 formal spearphishing messages and affected 2,275 victims.
The strategy indicates that spearphishing attackers would track victims for a continuous period, which is more costly than one-time scams by also bringing high profits.
Strategies of Spearphishing

Multi-Semantic Evasion

[Guangfa Bank] Dear Alice: Bob remitted 2.88 RMB to your account with the message: Tiancai Gambling service in tc***.com will send bonus to new users.

Semantic 1: Financial
Semantic 2: Gambling

Global Affair Integration

Hi, NAME. Yabo Sports join hands with Wuhan, register on f**8.cn and you can get recharge and get masks, come on, Wu Han!
Profound Real-World Impact of SMS Spearphishing

Victim Coverage
Around **24k victims** were endangered by spearphishing spam SMS from Dec. 2019 to Mar. 2020, covering all the provinces in China.

Follow-Up Domain Visits

[China Life Property and Casualty Insurance] Dear insured Name, our company has issued a commercial insurance policy for your car Yue MP***3, policy number is 805****612, the insurance period is 2020-03-19-2021-03-18, the premium amount is 1305.55 RMB. For more information, please see [http://9**.cc/b***g](http://9**.cc/b***g)

1,392 (94.50%) domains are marked as malicious, of which 105 are phishing-related.
Mitigation Recommendation

Defense of SMS Spearphishing

Security Software
- Enable our detection system and provide UI alerts to users.

Internet Service Provider (ISP)
- Enhance the audits of content submitted to their SMS platforms.

Attacking strategies we found would assist multiple parties in designing new detection methods of spearphishing SMS.
Summary

1. A first systematical measurement study on SMS Spearphishing ecosystem
   From a country-level dataset, with 31.96M real-world data
   Design a NLP-based detection system

2. Understand the strategy and behavior of spammers
   From Macro-Level and Micro-Level
   Still growing, with new categories emerging and various evasion strategies

3. Recommendation for mitigation SMS Spearphishing attacks
   Attacking strategies can assist in designing new detection methods
Detecting and Characterizing SMS Spearphishing Attacks

Mingxuan Liu, Yiming Zhang, Baojun Liu, Zhou Li, Haixin Duan, Donghong Sun

Email: liumx18@mails.tsinghua.edu.cn