Lies in the Air: Characterizing Fake-base-station Spam Ecosystem in China

Yiming Zhang, Baojun Liu, Chaoyi Lu, Zhou Li, Haixin Duan, Shuang Hao, Mingxuan Li, Ying Liu, Dong Wang and Qiang Li
What is a Fake-base-station?
How far is it from our lives?
Fake-base-station is Right by Your Side

Victim

FBS Operator/Spammer

FBS Spam Message

FBS Devices
利用伪基站发送300多万条信息 四川攀枝花警方抓
2020年上半年打击治理“黑广播”“伪基站”情况

2020年上半年，全国无线电管理机构查处“黑广播”违法犯罪案件806起（其中“黑广播”干扰民航案件18起），缴获“黑广播”设备660台（套）。查处“伪基站”违法犯罪案件6起，缴获“伪基站”设备1台（套）。
Fake-base-station: A Long-standing Problem

- 1G: Voice Call Only
- 2G: Text Message
- 3G: Internet
- 4G: Video Call
- 5G: Internet of Everything


- Legitimate Base Station
- Fake Base Station
- Network BTS
- Attacker

Root Cause: Lack of base station authentication under GSM(2G) network

An adversary could force the device to **downgrade from 3G/4G(5G) to 2G**.

FBS will be a **long-standing threat**!
Fake-base-station as a Spamming Channel

- In this work, we focus on the ability of FBSes to send spam messages to end-user devices from arbitrary phone numbers.

Techniques of FBS devices have been well studied. Several detecting methods have been proposed.

We still lack deep insights into the ecosystem powered by FBS.

* The data collection was implemented by our industrial partner, and we don’t consider it as our contribution in this work.
Outline of Our Work

• A **data-driven approach** to characterize FBS Spam ecosystem

**Data Collection**

• **279K real-world FBS messages**, 97 days

**Methodology**

• Machine-learning based **category classifier**
• Contact-information based **spam campaign classifier**

**Measurement**

• “Macro-level”: **behaviors of FBS spammers**
• “Micro-level”: **strategies of spam campaigns**

To understand the fraudulent activities and explore strategies in the FBS spam ecosystem

* The labeled ground-truth dataset (14K anonymized real-world FBS messages) has been released at https://github.com/Cyper-Z/FBS_SMS_Dataset.
Data Collection
Data Collection

• 279,017 FBS detection logs, Dec.1, 2018 to Mar.7,2019 (97 days)

---

Example of Collected Data Logs

<table>
<thead>
<tr>
<th>Timestamp (logged time)</th>
<th>Sender Phone Number</th>
<th>IMEI (hashed for anonymity)</th>
<th>IMSI (hashed for anonymity)</th>
<th>Information of Recently connected Base Station</th>
<th>IP address of mobile client</th>
<th>Message Content</th>
</tr>
</thead>
</table>

---

Detection based on FBS features
Move into Spam Inbox and notify user
Upload anonymized logs
Our dataset
Auxiliary Datasets

- **Passive DNS**: Query volume and active time of suspicious domains
- **WHOIS Database**: Registration behaviors of suspicious domains
- **Domain Blacklist**: Identify malicious domains
- **Bankcard Information**: Investigate fraudster’s bankcard information
- **BS Geo-location Database**: Geographic distributions of FBS spammers
Categorizing FBS Spam Messages
Overview of Data Processing

- Infer **business category** of FBS spam messages (self-labeled dataset, SVM)
- Infer **spam campaign** behind the spam activities (embedded contact information)

**Pre-Process**
- Contact Identification
- Entity Recognition

**Classifying Categories**
- TF-IDF
- Classifier

**Clustering Campaigns**
- Filter out messages labeled as “Others”

Overview of the data processing flow
Step 0: Collect Ground-truth

The Origin Dataset
27.9K Records

5% random samples
manually label
consistently labeled data

The Subset for
Manually Inspect
3 Researchers Label
Independently
The Ground-truth Dataset
14,077 Records

14 Categories of FBS Messages

Category
Illegal Promotion
Fraud
Advertisement
Other

Sub-category
Illegal in China

Gambling
Fake ID and invoice
Political propaganda
Escort service

Phishing (Bank)
Phishing (Other)
Financial fraud
Others

Retail
Loan service
Real estate
Network service
Others

Removed from dataset
Step I: Classify Business Categories

Fact: 78.3% FBS messages contain at least 1 contact identifier

Pre-process
8 types of contact information -> constant strings

Classification
- SVM: 96.90% Precision, 96.96% Recall, 96.87% F1-score
- NB: 95.23% Precision, 95.16% Recall, 95.06% F1-score
- LR: 94.90% Precision, 94.64% Recall, 94.32% F1-score
- RF: 75.63% Precision, 71.88% Recall, 72.89% F1-score

Five-fold cross validation results

Original message
尊敬的杨**，您的银行卡已被冻结，请登陆www.icbc**.com联系客服130xxxx8816或微信cn***。
【工商银行】

Pre-processed message
尊敬的NAME您的银行卡已被冻结，请登陆DOMAIN联系客服CELLPHONE或WECHAT工商银行

Classification
- TF-IDF
- Scikit-learn
- Evaluate on labeled set
- Choose SVM
Step II: Cluster Spam Campaigns

Previous Work

- IMC’10
- CCS’10
- ACM PER’09

Text Similarity

- LEET’08
- Wisec’16

Case-by-case

- IMC’11

Our Method

Leverage spam’s contact information in the FBS messages

- Assumption #1: Messages sharing the same contact information belong to the same campaign.
- Assumption #2: Contacts in a spam message belong to the same spam campaign.

Iteratively clustering spam campaigns

Discover 7,884 spam campaigns
Measuring the Patterns of FBS Spammers
Business: Profit-driven, Mostly Illegal

Ratio of three major spam categories

FBS Spam
SMS Gateway Spam [Wisec’16]

<table>
<thead>
<tr>
<th>Rank</th>
<th>Sub-Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fake ID and invoice</td>
</tr>
<tr>
<td>2</td>
<td>Bank Phishing</td>
</tr>
<tr>
<td>3</td>
<td>Gambling</td>
</tr>
<tr>
<td>4</td>
<td>Escort Service</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Sub-Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Payday Loan</td>
</tr>
<tr>
<td>2</td>
<td>Job Advertisement</td>
</tr>
<tr>
<td>3</td>
<td>Contest</td>
</tr>
<tr>
<td>4</td>
<td>Telephone Service</td>
</tr>
</tbody>
</table>

Different business type compared with other spam
- Social-economic diversity
- Profit margin

FBS Messages are mostly used to advertise illegal business
Temporal: FBS Spammers are Hard Working

Do spammers keep working on weekends?

- Other types of spam, e.g., domain squatting[ISRAID’17], spam calls[S&P’18]: No! Take a break!
- FBS spam in China: Yes! Rarely rest.

Also keep working after midnight / on New Year’s Day.

Only rest around Spring Festival (Gambling spammers remain active then).
Spatial: Crowd Targeted, Regional Customized

Largely active near main roads and highly-populated regions for increasing influence

Geo-distribution of FBS spam victims: China-wide (Left), Chengdu city (Right)
Impact: Severely Adverse Impact in China

IMEI/IMSI
Estimate the affected client population

Over **100,000 mobile devices** still receive FBS messages with in study period.

*e.g.,* **38 IMEIs and 34 IMSIs receive more than 100 FBS spam messages during 97-day collection period**

PDNS data
Learn the “successful click rate” (follow-up visit)

**Domains** in the messages still receive **considerable visits** even after marked as spam.

*e.g.,* **3,197 (83.4%) domains are labeled as malicious**

**403 (12.6%) domains were queried for over 5,000 times**

**34 Gambling domains have been visited more than 100K times**

**Our estimation suggests FBS does have considerable impact on people in China**
Measuring the Strategies of Spam Campaigns
Overview of FBS Spam Campaign

7,884 spam campaigns are identified associated with 8,316 unique spam contacts

- **Scale**
  - Unevenly distributed
  - Mostly short-lived

- **Life time**
  - 92% active for less than 10 days
  - Top 20 long-lived: mostly Fake ID and invoice, “light crime”, low risk
  - Top 50 least-active: mostly Phishing messages, “illegal business”, high penalty

- **Organization**
  - Hierarchical architecture
  - Outsourcing models

- **Details**
  - The first 100 campaigns (1.3%) account for 35% messages
  - The largest campaign, with over 11,120 messages (4.55%) in 97 days
  - The largest campaign, with over 11,120 messages (4.55%) in 97 days

---

**Campaign I**

- Business Owner ➔ Intermediate Contact ➔ FBS Operator
- Resource Sharing ➔ Outsourcing ➔ Outsourcing

**Campaign II**

- Business Owner ➔ Intermediate Contact ➔ FBS Operator
- Outsourcing ➔ Outsourcing

### Top 10 spam campaigns sending most messages

<table>
<thead>
<tr>
<th>No.</th>
<th>Category</th>
<th>#Msg</th>
<th>#IMEI</th>
<th>Days</th>
<th>Active Time (Dec 1, 2018 – Mar 7, 2019)</th>
<th>Hourly Distribution</th>
<th>Locality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Loan</td>
<td>11,120</td>
<td>1,646</td>
<td>95</td>
<td></td>
<td></td>
<td>Dalian</td>
</tr>
<tr>
<td>2</td>
<td>Gambling</td>
<td>3,623</td>
<td>2,080</td>
<td>97</td>
<td></td>
<td></td>
<td>Macau</td>
</tr>
<tr>
<td>3</td>
<td>Gambling</td>
<td>2,971</td>
<td>1,904</td>
<td>97</td>
<td></td>
<td></td>
<td>Macau</td>
</tr>
<tr>
<td>4</td>
<td>Loan</td>
<td>2,327</td>
<td>687</td>
<td>88</td>
<td></td>
<td></td>
<td>Dalian</td>
</tr>
<tr>
<td>5</td>
<td>Gambling</td>
<td>1,416</td>
<td>580</td>
<td>77</td>
<td></td>
<td></td>
<td>Macau, Zhuhai</td>
</tr>
<tr>
<td>6</td>
<td>Fake ID</td>
<td>1,318</td>
<td>940</td>
<td>71</td>
<td></td>
<td></td>
<td>Chengdu</td>
</tr>
<tr>
<td>7</td>
<td>Gambling, Loan, Escort</td>
<td>1,283</td>
<td>460</td>
<td>60</td>
<td></td>
<td></td>
<td>Macau, Zhuhai</td>
</tr>
<tr>
<td>8</td>
<td>Ad-Other</td>
<td>1,249</td>
<td>889</td>
<td>72</td>
<td></td>
<td></td>
<td>Chengdu</td>
</tr>
<tr>
<td>9</td>
<td>Bank Phishing</td>
<td>1,206</td>
<td>903</td>
<td>35</td>
<td></td>
<td></td>
<td>Cities of Sichuan</td>
</tr>
<tr>
<td>10</td>
<td>Gambling</td>
<td>1,127</td>
<td>486</td>
<td>76</td>
<td></td>
<td></td>
<td>Macau, Zhuhai</td>
</tr>
</tbody>
</table>

**Outsourcing of FBS Operator**

Multiple campaigns could be undertaken of the same FBS operator at the same time

- Campaign 2&3, 5&10, 6&8 are similar both in active time and active location, with at least 54% overlap of affected IMEIs
Resource Sharing Between Spam Campaigns

**Outsourcing of Victim Interaction**

<table>
<thead>
<tr>
<th>Message Content</th>
<th>Category</th>
<th>Active Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>提供贷款，请联系陈经理 微信132****1290</td>
<td>Ad-Loan</td>
<td>Dec.30, 2018</td>
</tr>
<tr>
<td>新鲜乡村花生油，自然无添加，微信132****1290</td>
<td>AD-Other</td>
<td>Jan 23, 2019</td>
</tr>
<tr>
<td>皇冠娱乐十周年，惊人优惠返现，联系微信132****1290</td>
<td>IL-Gambling</td>
<td>Jan 23, 2019 – Jan 27, 2019</td>
</tr>
</tbody>
</table>

**Template Sharing**

- **262 templates** are identified among 994 campaigns.
- **83 templates (31.68%)** were shared while 858 campaigns (86%) share templates with others.
- **Phishing (Bank)** has the highest sharing rate.

**Find shared contacts: category entropy**

\[ H = - \sum_i P_i \log P_i \]

185 contacts (2.22%) are shared among multiple categories in all campaigns.

<table>
<thead>
<tr>
<th>Category</th>
<th># Template</th>
<th># Msg</th>
<th>% Templated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phishing (Bank)</td>
<td>168</td>
<td>22,737</td>
<td>32.35%</td>
</tr>
<tr>
<td>Fake ID and invoice</td>
<td>19</td>
<td>4,427</td>
<td>5.77%</td>
</tr>
<tr>
<td>Gambling</td>
<td>20</td>
<td>10,808</td>
<td>39.62%</td>
</tr>
<tr>
<td>Loan Service</td>
<td>18</td>
<td>11,135</td>
<td>49.48%</td>
</tr>
<tr>
<td>Retail</td>
<td>11</td>
<td>3,385</td>
<td>28.11%</td>
</tr>
<tr>
<td>Ad (Other)</td>
<td>10</td>
<td>2,314</td>
<td>17.19%</td>
</tr>
<tr>
<td>Network Service</td>
<td>5</td>
<td>929</td>
<td>6.46%</td>
</tr>
<tr>
<td>Escort Service</td>
<td>6</td>
<td>504</td>
<td>14.97%</td>
</tr>
<tr>
<td>Financial Fraud</td>
<td>3</td>
<td>57</td>
<td>13.23%</td>
</tr>
<tr>
<td>Real Estate</td>
<td>2</td>
<td>32</td>
<td>1.35%</td>
</tr>
</tbody>
</table>
Tricking Strategies of FBS Spam Campaigns

• Sender spoofing

FBS Spammers use spoofed sender numbers of well-known companies to make spam messages more deceptive.

Sender ID (Caller ID) Spoofing is very effective in telephone scams. [Usenix’ 19]

• Message Wording

The language of FBS messages is usually captivating (with scares and monetary lures) to engage users.

User education would be necessary

Examples of Top Spoofed Senders

<table>
<thead>
<tr>
<th>Type</th>
<th>Sender</th>
<th># Msg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank</td>
<td>95588</td>
<td>23,444</td>
</tr>
<tr>
<td>ISP</td>
<td>10086</td>
<td>12,161</td>
</tr>
<tr>
<td>Payment</td>
<td>95107</td>
<td>5,039</td>
</tr>
<tr>
<td>Insurance</td>
<td>95518</td>
<td>149</td>
</tr>
</tbody>
</table>

114 templates (43.5%)

Scare users
76 frozen credit cards, 16 blocked accounts, 14 stolen devices

104 templates (39.7%)

Attract users by money lures
58 credit card limit increasing
15 ISP discounts
Evasion Strategies of FBS Spam Campaigns

**Domain Infrastructure**
- Newly registered domains
  - 1,155 (38.4%) domains are registered after 2019
- Domain-squatting services
  - 278 are over 3 years old registered early, leveraged in batches
- URL-shorten
  - 397 (69%) URLs use URL shorteners
    - http://t.cn/xxxxxx
    - http://dwz.cn/xxxxx

- Avoid Domain Blacklisting

**Bank Account**
- Abusing flawed bank policy
  - Registered in mid-west China with flawed bank policy
- Loose Authentication
- Free Secondary Card

- Avoid Bank Blocking

**Spammer Contacts**
- Social platform accounts for the most
- Low blocking rate, long live-time

- Avoid Account Blocking
## Interesting Case: Moving Spam Campaign

### Top 10 spam campaigns sending most messages

<table>
<thead>
<tr>
<th>No.</th>
<th>Category</th>
<th>#Msg</th>
<th>#IMEI</th>
<th>Days</th>
<th>Active Time (Dec 1, 2018 – Mar 7, 2019)</th>
<th>Hourly Distribution</th>
<th>Locality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Loan</td>
<td>11,120</td>
<td>1,646</td>
<td>95</td>
<td></td>
<td></td>
<td>Dalian</td>
</tr>
<tr>
<td>2</td>
<td>Gambling</td>
<td>3,623</td>
<td>2,080</td>
<td>97</td>
<td></td>
<td></td>
<td>Macau</td>
</tr>
<tr>
<td>3</td>
<td>Gambling</td>
<td>2,971</td>
<td>1,904</td>
<td>97</td>
<td></td>
<td></td>
<td>Macau</td>
</tr>
<tr>
<td>4</td>
<td>Loan</td>
<td>2,327</td>
<td>687</td>
<td>88</td>
<td></td>
<td></td>
<td>Dalian</td>
</tr>
<tr>
<td>5</td>
<td>Gambling</td>
<td>1,416</td>
<td>580</td>
<td>77</td>
<td></td>
<td></td>
<td>Macau, Zhuhai</td>
</tr>
<tr>
<td>6</td>
<td>Fake ID</td>
<td>1,318</td>
<td>940</td>
<td>71</td>
<td></td>
<td></td>
<td>Chengdu</td>
</tr>
<tr>
<td>7</td>
<td>Gambling, Loan, Escort</td>
<td>1,283</td>
<td>460</td>
<td>60</td>
<td></td>
<td></td>
<td>Macau, Zhuhai</td>
</tr>
<tr>
<td>8</td>
<td>Ad-Other</td>
<td>1,249</td>
<td>889</td>
<td>72</td>
<td></td>
<td></td>
<td>Chengdu</td>
</tr>
<tr>
<td>9</td>
<td>Bank Phishing</td>
<td>1,206</td>
<td>903</td>
<td>35</td>
<td></td>
<td></td>
<td>Cities of Sichuan</td>
</tr>
<tr>
<td>10</td>
<td>Gambling</td>
<td>1,127</td>
<td>486</td>
<td>76</td>
<td></td>
<td></td>
<td>Macau, Zhuhai</td>
</tr>
</tbody>
</table>
Interesting Case: Moving Spam Campaign

Cooperation of government departments across cities is necessary to combat campaign migration.
Recommendations for the Community

All of the parties evolved in FBS Ecosystem should unite and work together to mitigate FBS Spam issues.

- **Government Agencies**: Re-evaluate their bankcard policies to avoid being abused.
- **Bank**: Better fraud detection mechanisms.
- **Social Media Platform**: Checking accounts with fraudulent activities.
- **Security Software**: New UI system of application.
- **Enterprises**: Extracted templates as new features.
- **Mobile Carriers**: Update cell towers, abandon GSM protocol.
- **Update cell towers, abandon GSM protocol**.
- **More efforts in seriously effected places and cities**.
- **User education, new scenarios of deceptive messages in our work may help**.
Summary

A first comprehensive measurement study on FBS spam ecosystem
Country-level perspective, 27K real-world data
Classification of FBS business, identify spam campaigns

Understand how FBS ecosystem is organized, how spammers behave
“Micro-level” and “Macro-level
Still active and evolving, severe real-world impact

Recommendations for better solutions against FBS spam
Cooperation of multiple parties
Released dataset and extracted FBS templates
Lies in the Air: Characterizing Fake-base-station Spam Ecosystem in China

Yiming Zhang, Baojun Liu, Chaoyi Lu, Zhou Li, Haixin Duan, Shuang Hao, Mingxuan Li, Ying Liu, Dong Wang and Qiang Li

Email: zhangyim17@mails.tsinghua.edu.cn