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

TSINGHUA UNIVERSITY 1911

UCI University of California, Irvine

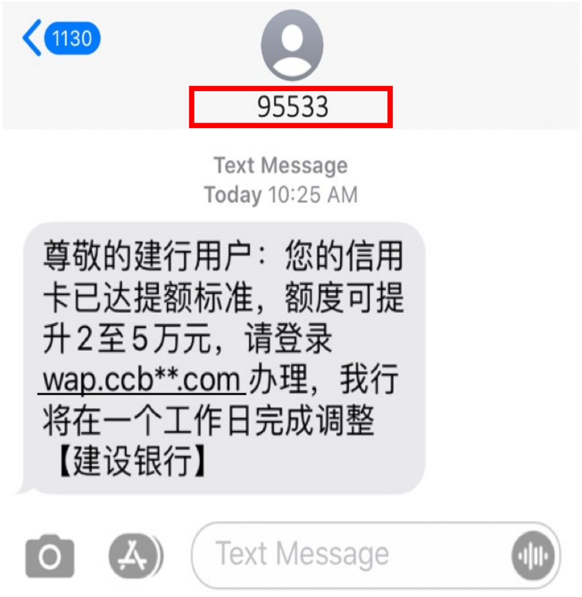
UT DALLAS The University of Texas at Dallas

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

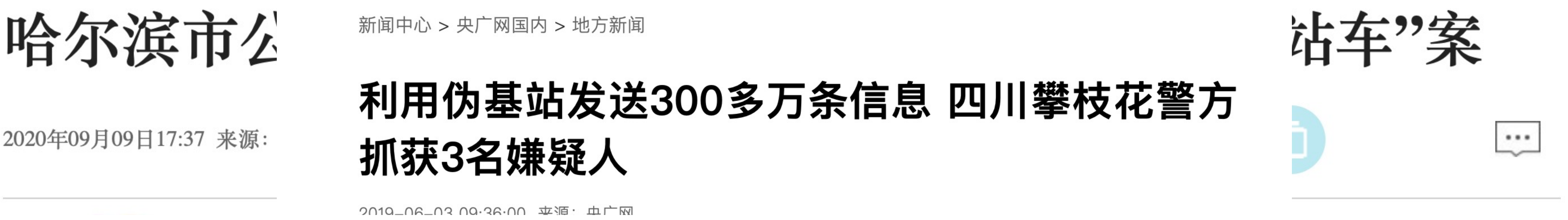


Text Message
Today 10:25 AM

尊敬的建行用户：您的信用卡已达提额标准，额度可提升2至5万元，请登录wap.ccb**.com办理，我行将在一个工作日完成调整【建设银行】

FBS Spam Message

FBS Devices



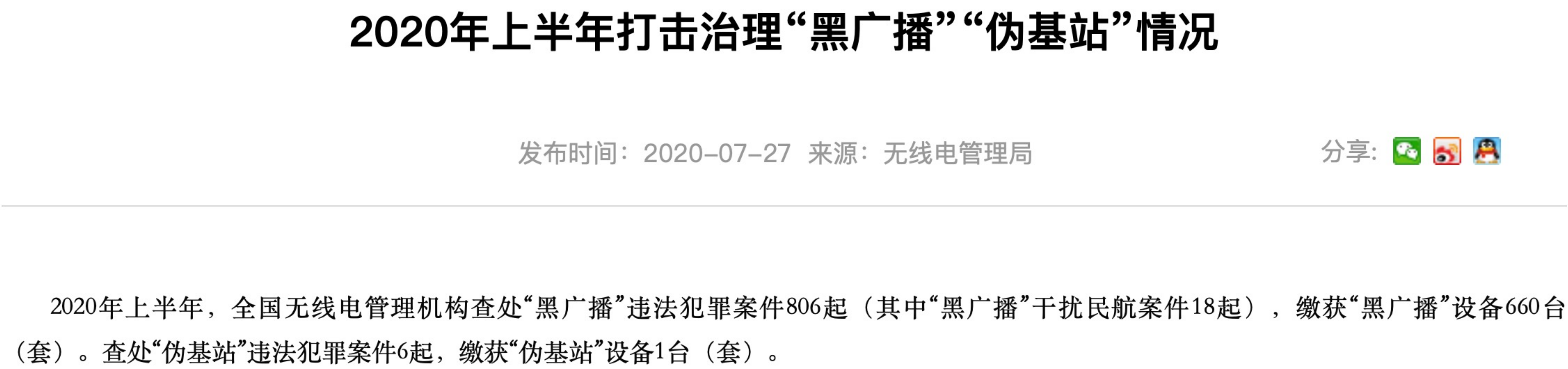
哈尔滨市公 站车”案

2020年09月09日17:37 来源:

新闻中心 > 央广网国内 > 地方新闻

利用伪基站发送300多万条信息 四川攀枝花警方抓获3名嫌疑人

2019-06-03 09:36:00 来源：央广网



2020年上半年打击治理“黑广播”“伪基站”情况

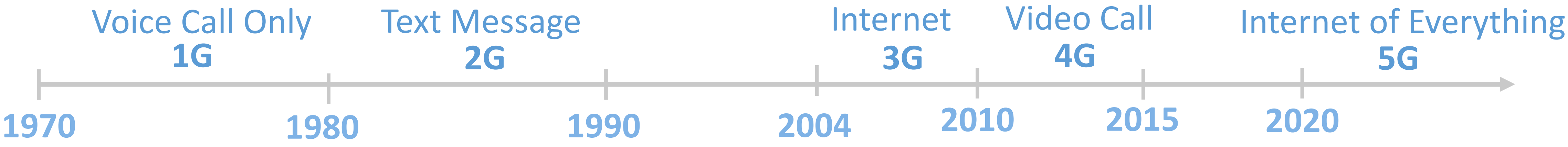
发布时间：2020-07-27 来源：无线电管理局 分享:

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）。

以到全国各地“旅行”赚

入诈骗“黑灰产业链”中

Fake-base-station: A Long-standing Problem



Voice Call Only 1G

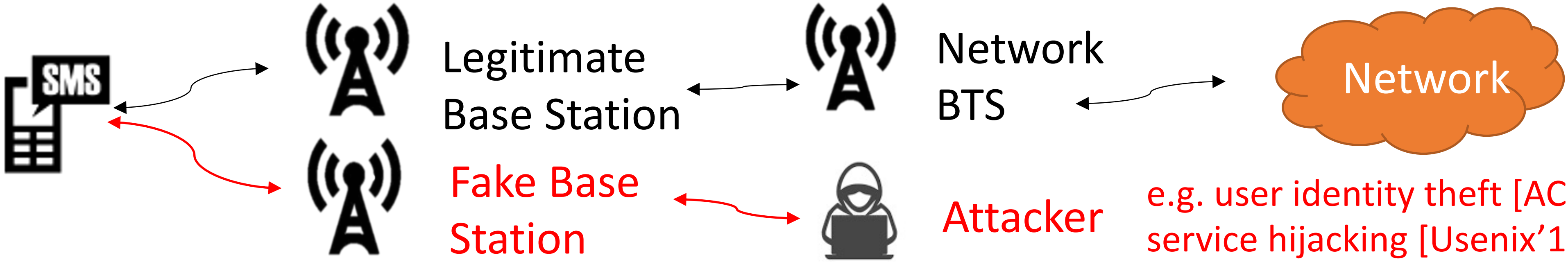
Text Message 2G

Internet 3G

Video Call 4G

Internet of Everything 5G

1970 1980 1990 2004 2010 2015 2020



Legitimate Base Station

Network BTS

Network

Fake Base Station

Attacker

e.g. user identity theft [ACSAC'14], service hijacking [Usenix'13]

- 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.

VS. Other Spam Channels

e.g., SMS Gateway

Deceptiveness

Less Restrictions

Low Cost, High Profit

SDR

Focus on detecting FBSes

Strong Signal Strength

Significant Change of BS ID

Similar features were used to collect data in this work*

* 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/Cypher-Z/FBS_SMS_Dataset.

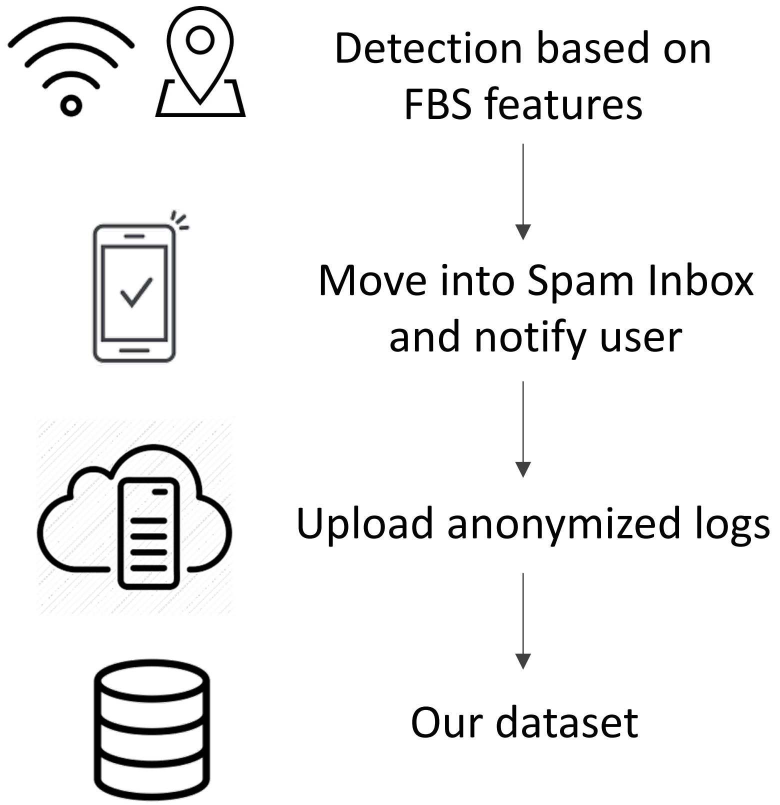
Data Collection

Data Collection

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

360
Mobile Guard

Example of Collected Data Logs	
2018-12-03 18:43:07	Timestamp (logged time)
95588 (ICBC)	Sender Phone Number
HASH_1	IMEI (hashed for anonymity)
HASH_2	IMSI (hashed for anonymity)
Cellinfo: lac:9418&cellid:3133 2018-12-03 18:43:08,…,…	Information of Recently connected Base Station
157.xxx.xxx.132	IP address of mobile client
您的综合评分良好，可申请提升信用卡额度2万元，www.lcbl95588.com 【工商银行】	Message Content



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

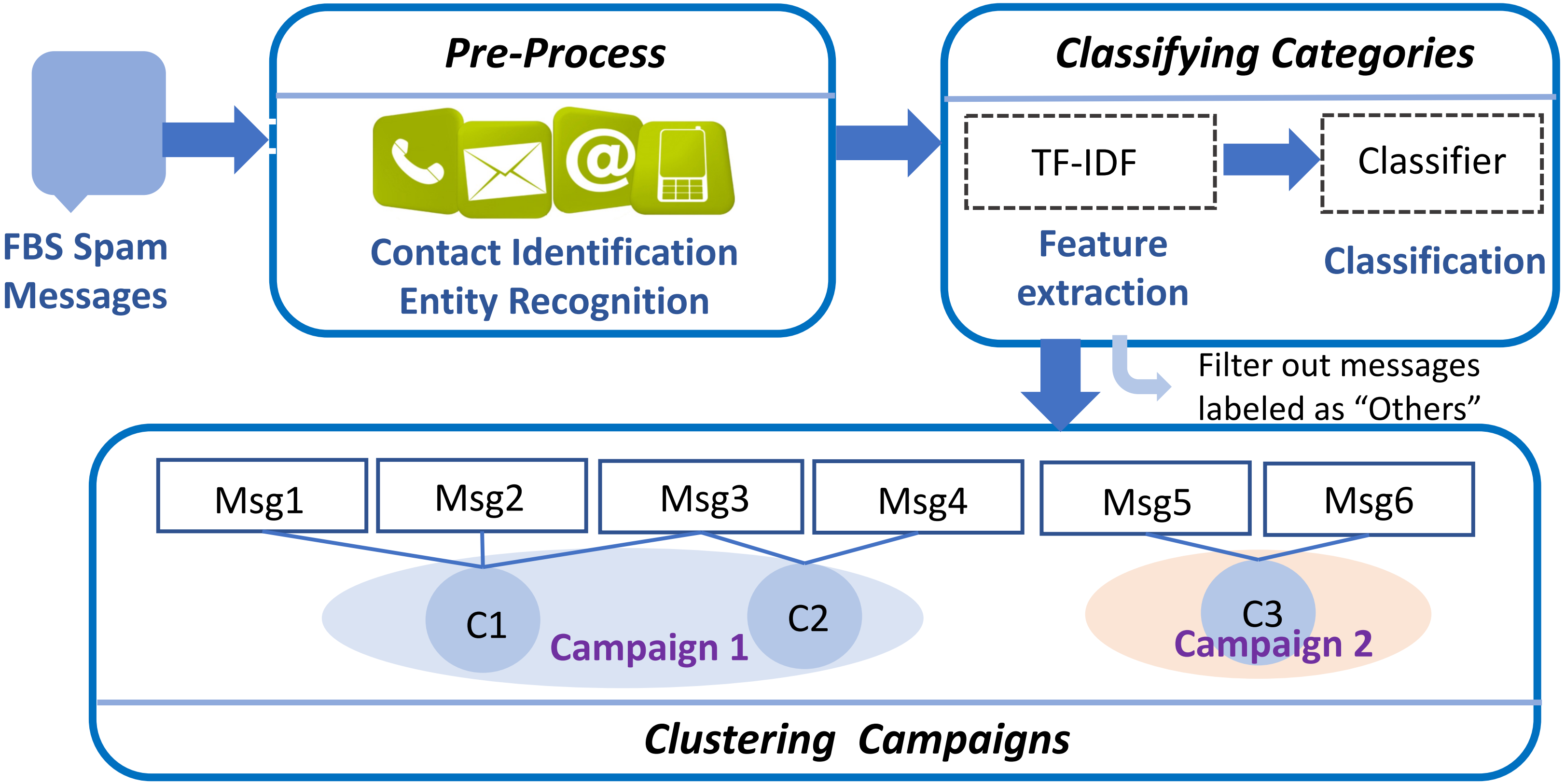
BS Geo-location Database

Geographic distributions of FBS spammers

Categorizing FBS Spam Messages

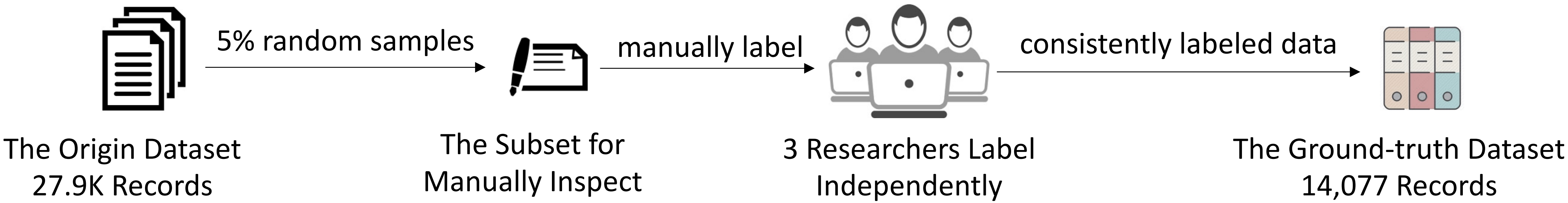
Overview of Data Processing

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



Overview of the data processing flow

Step 0: Collect Ground-truth



The Origin Dataset
27.9K Records

5% random samples

The Subset for
Manually Inspect

manually label

3 Researchers Label
Independently

consistently labeled data

The Ground-truth Dataset
14,077 Records

14 Categories of FBS Messages

Category	**Illegal Promotion**	**Fraud**	**Advertisement**	**Other**
Sub-category	Gambling Fake ID and invoice Political propaganda Escort service	Phishing (Bank) Phishing (Other) Financial fraud Others	Retail Loan service Real estate Network service Others	-

Illegal in China

Removed from dataset

Step I: Classify Business Categories

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

Classification ☹

Spam Campaign analysis ☺

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

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】

Original 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

Pre-processed message

Classification

TF-IDF Scikit-learn -> Evaluate on labeled set -> Choose SVM

Five-fold cross validation results

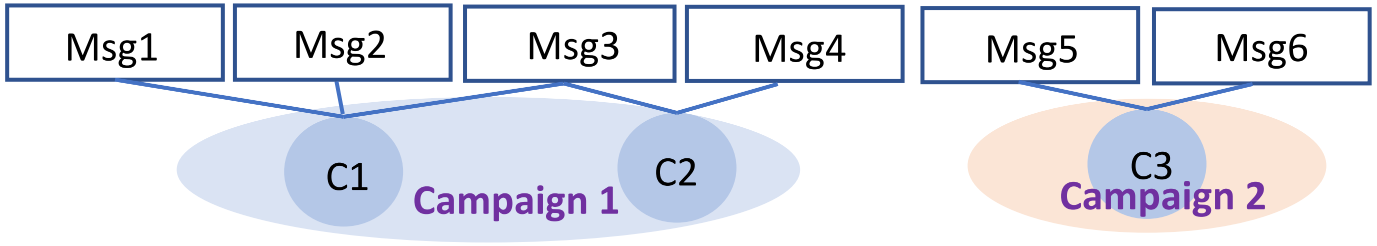
Classifier	Precision	Recall	F1-score
SVM	96.90%	96.96%	96.87%
NB	95.23%	95.16%	95.06%
LR	94.90%	94.64%	94.32%
RF	75.63%	71.88%	72.89%

Step II: Cluster Spam Campaigns

Previous Work	**URLs**	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.

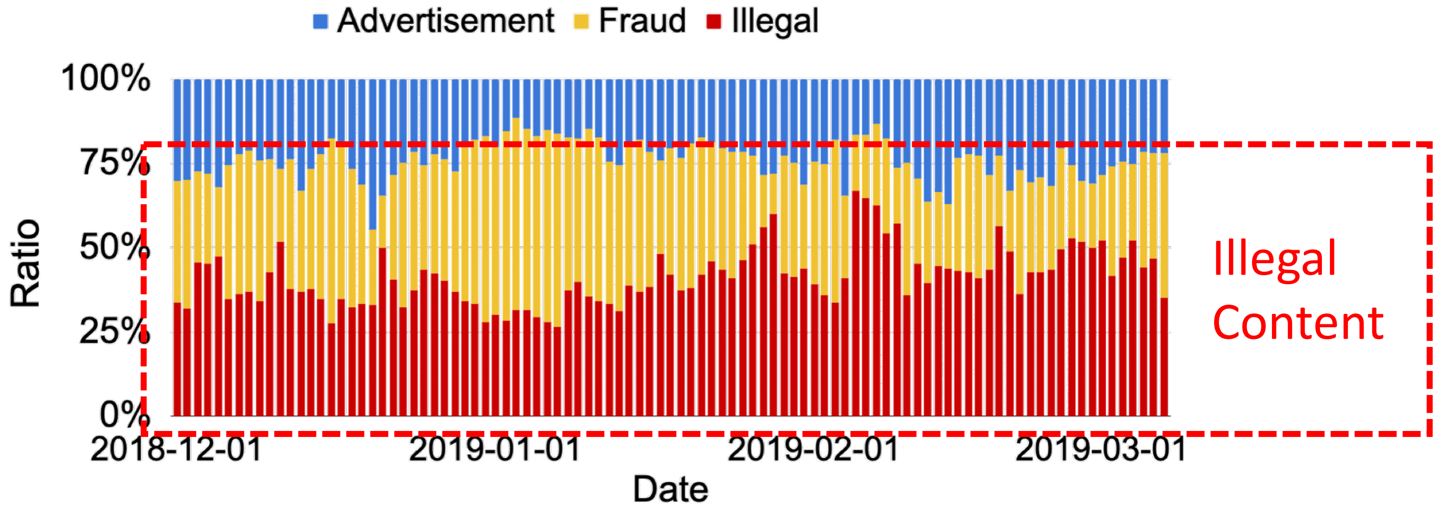


Iteratively clustering spam campaigns

Discover 7,884 spam campaigns

Measuring the Patterns of FBS Spammers

Business: Profit-driven, Mostly Illegal



Illegal Content

Ratio of three major spam categories

FBS Messages are mostly used to advertise illegal business

Rank	Sub-Category
1	Fake ID and invoice
2	Bank Phishing
3	Gambling
4	Escort Service

FBS Spam

Rank	Sub-Category
1	Payday Loan
2	Job Advertisement
3	Contest
4	Telephone Service

SMS Gateway Spam[Wisec'16]

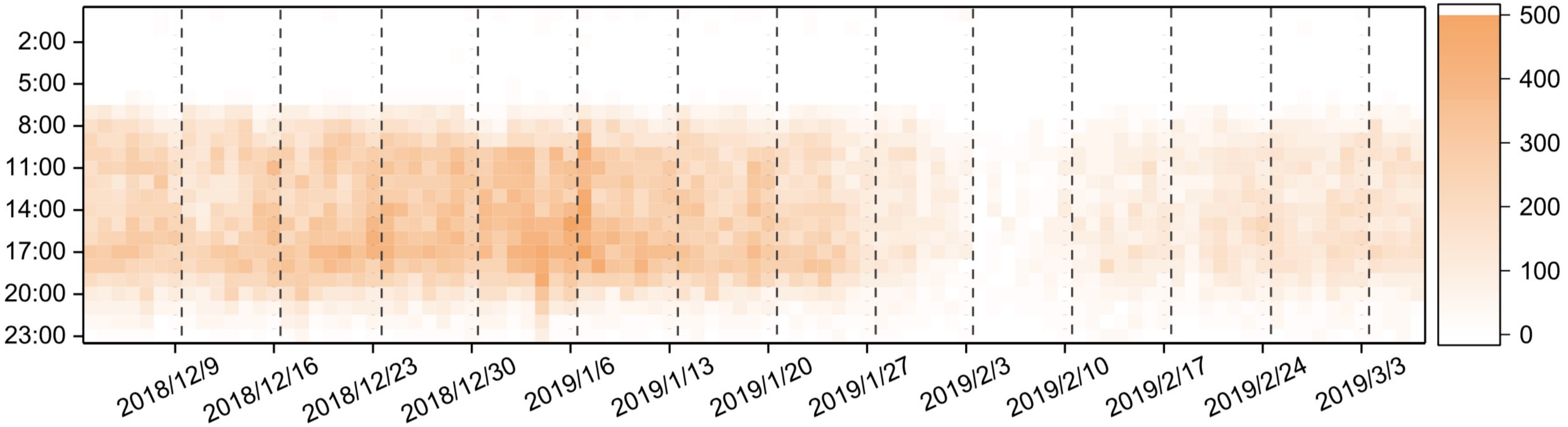
Different business type compared with other spam

- **Social-economic diversity**
- **Profit margin**

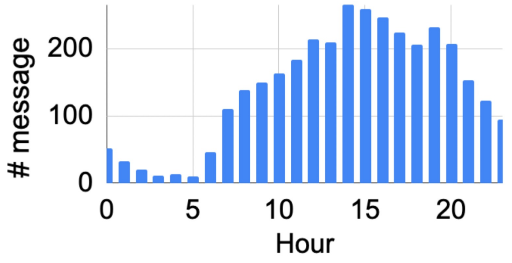
Temporal: FBS Spammers are Hard Working

Do spammers keep working on weekends?

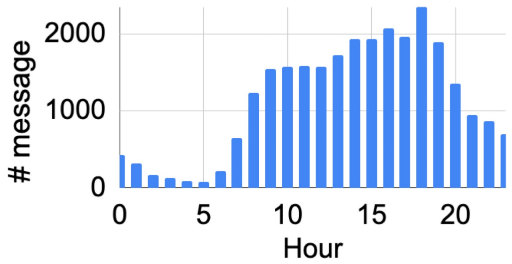
- **Other types of spam**, e.g., domain squatting[ISRAID'17], spam calls[S&P'18]: **No! Take a break!**
- **FBS spam in China : Yes! Rarely rest.**



Heatmap of spam activities



Escort service activities



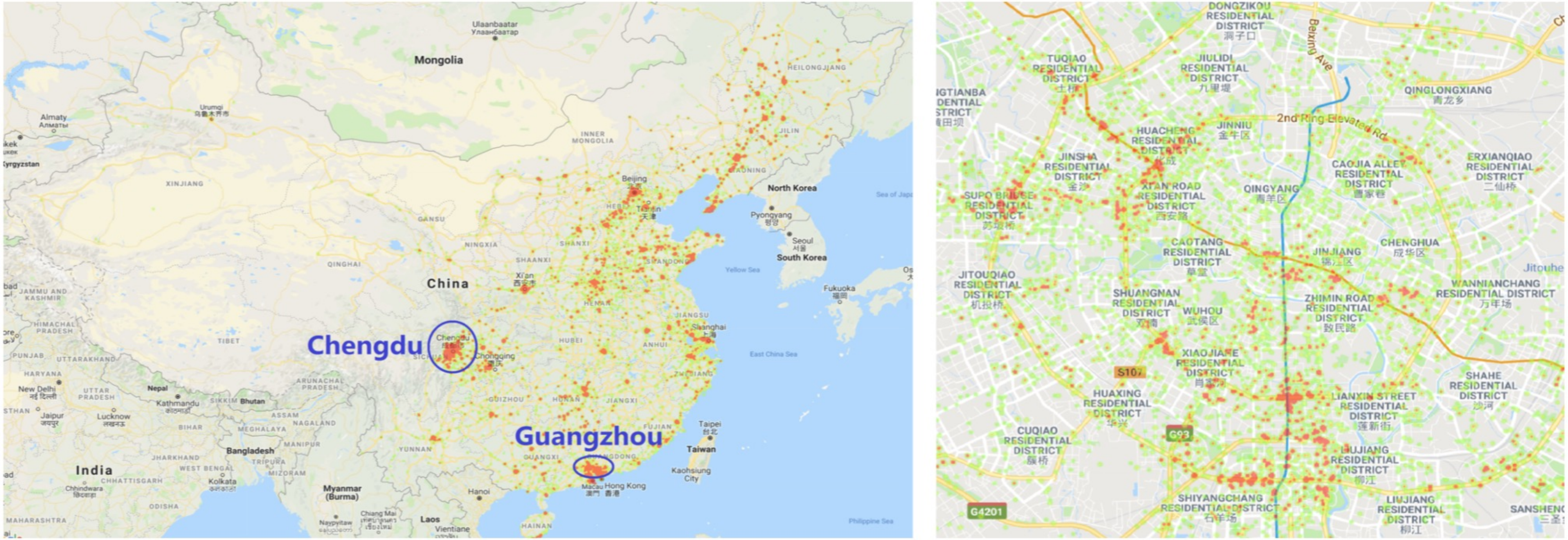
Gambling activities

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

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

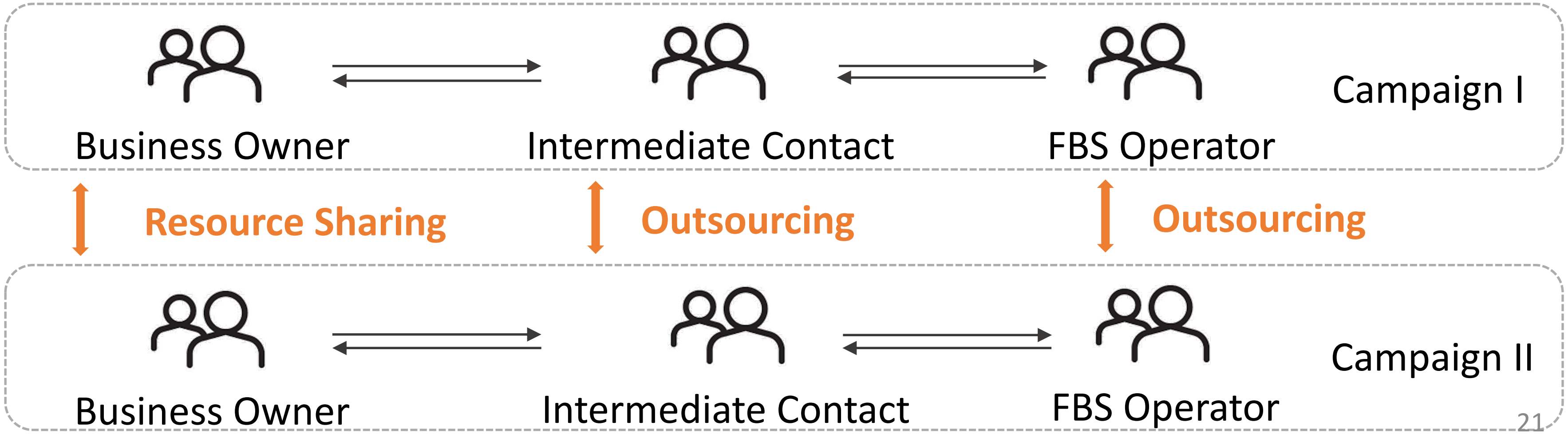
Life time

Mostly short-lived

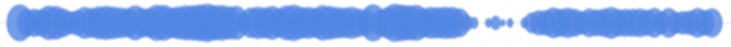
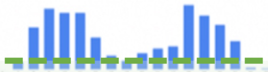
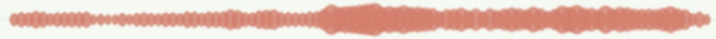
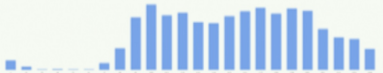
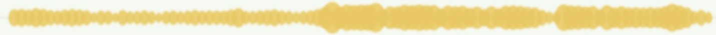
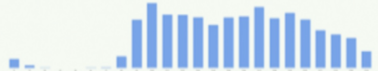
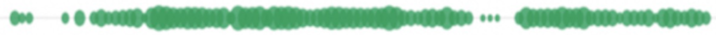
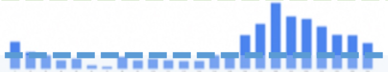
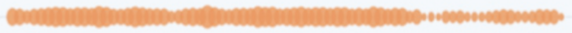
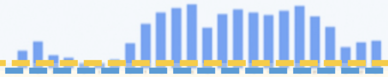
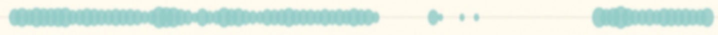
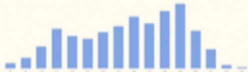
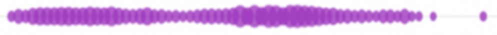
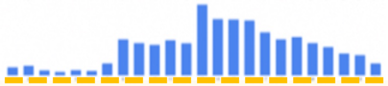
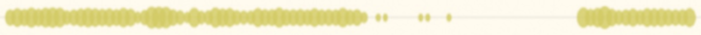
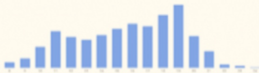
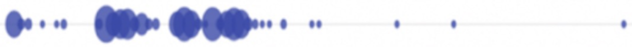
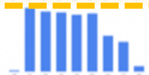
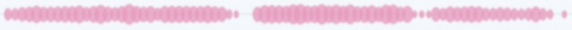
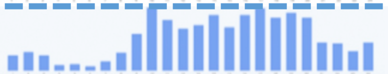
- 92% active for less than 10 days
- Top 20 **long-lived**: mostly Fake ID and invoice, **“light crime”, low risk**
- Top 50 **least-active**: mostly Phishing messages , **“illegal business”, high penalty**

Organization

Hierarchical architecture
Outsourcing models



Top 10 spam campaigns sending most messages

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

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

Message Content	Category	Active Days
提供贷款，请联系陈经理**微信132****1290**	Ad-Loan	Dec.30, 2018
新鲜乡村花生油，自然无添加，**微信132****1290**	AD-Other	Jan 23, 2019
皇冠娱乐十周年，惊人优惠、返现，联系**微信132****1290**获取更多惊喜！	IL-Gambling	Jan 23, 2019 – Jan 27, 2019

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.

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.

Category	# Template	# Msg	% Templated
Phishing (Bank)	168	22,737	32.35%
Fake ID and invoice	19	4,427	5.77%
Gambling	20	10,808	39.62%
Loan Service	18	11,135	49.48%
Retail	11	3,385	28.11%
Ad (Other)	10	2,314	17.19%
Network Service	5	929	6.46%
Escort Service	6	504	14.97%
Financial Fraud	3	57	13.23%
Real Estate	2	32	1.35%

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]

Examples of Top Spoofed Senders

Type	Sender	# Msg
Bank	95588	23,444
ISP	10086	12,161
Payment	95107	5,039
Insurance	95518	149

- Message Wording

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

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

User education would be necessary

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

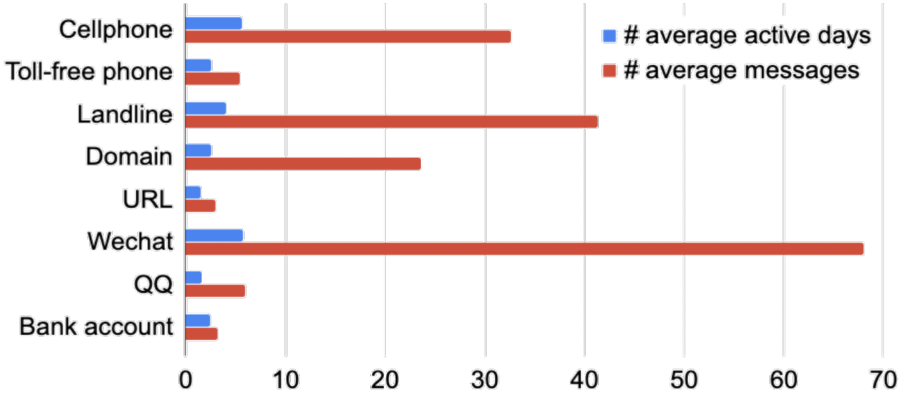
Loose Authentication

Free Secondary Card

-> Avoid Bank Blocking

Spammer Contacts

Social platform accounts for the most

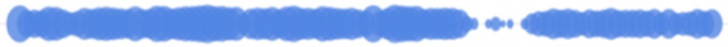
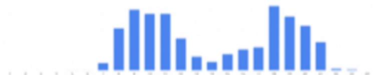
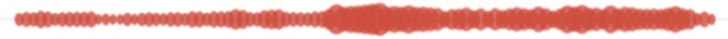
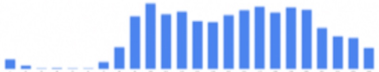
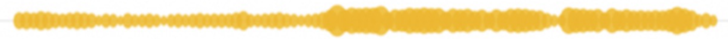
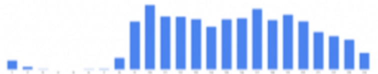
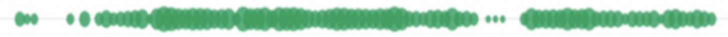
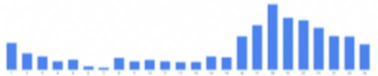
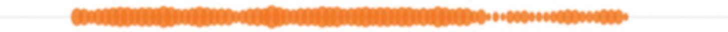
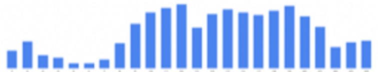
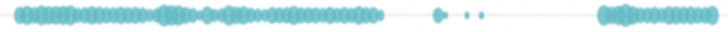
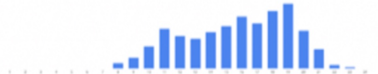
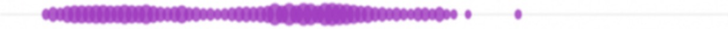
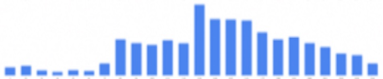
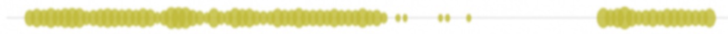
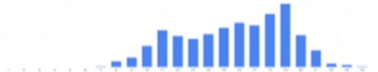
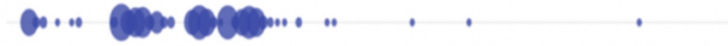
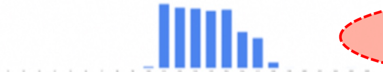
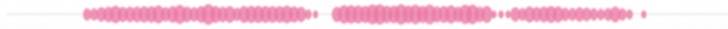
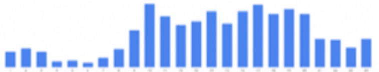


Low blocking rate, long live-time

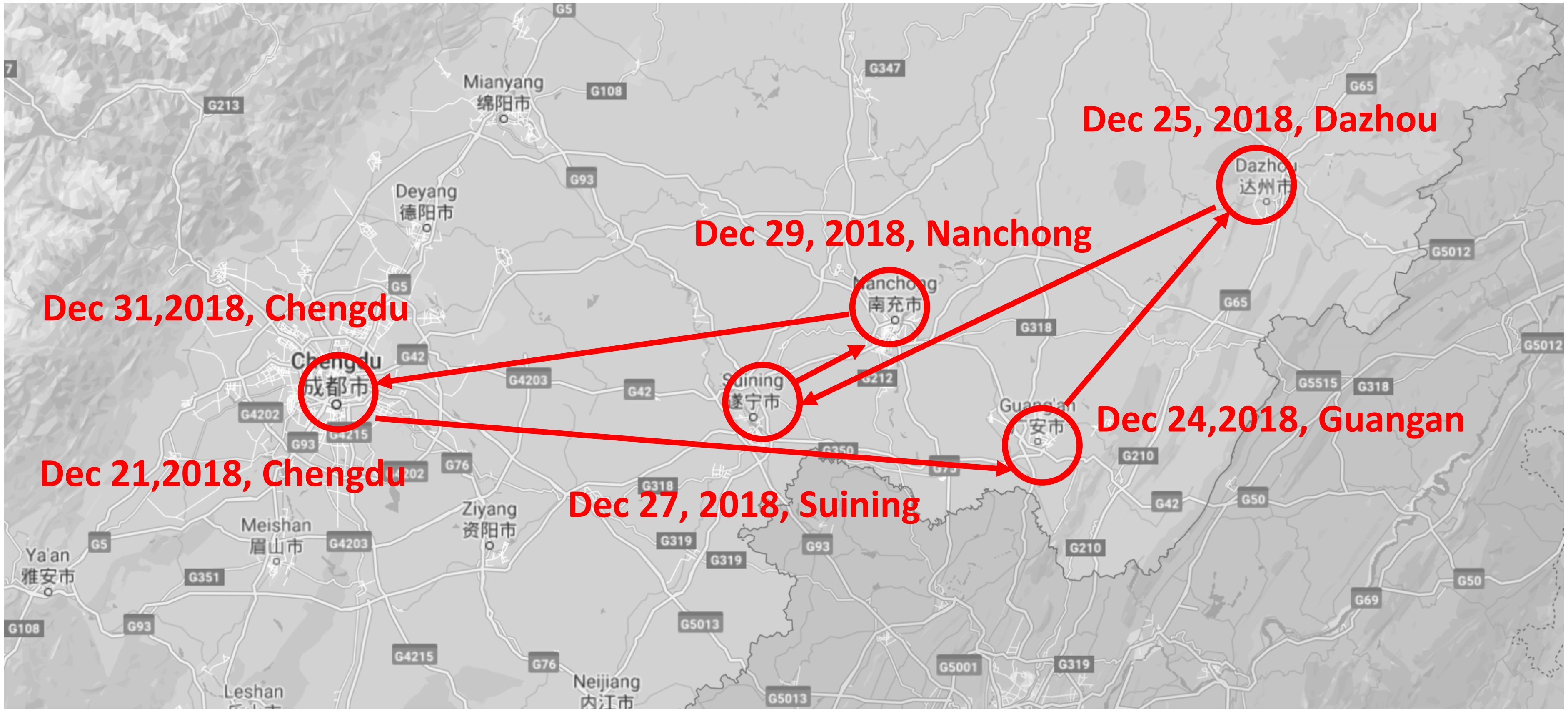
-> Avoid Account Blocking

Interesting Case: Moving Spam Campaign

Top 10 spam campaigns sending most messages

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

Interesting Case: Moving Spam Campaign

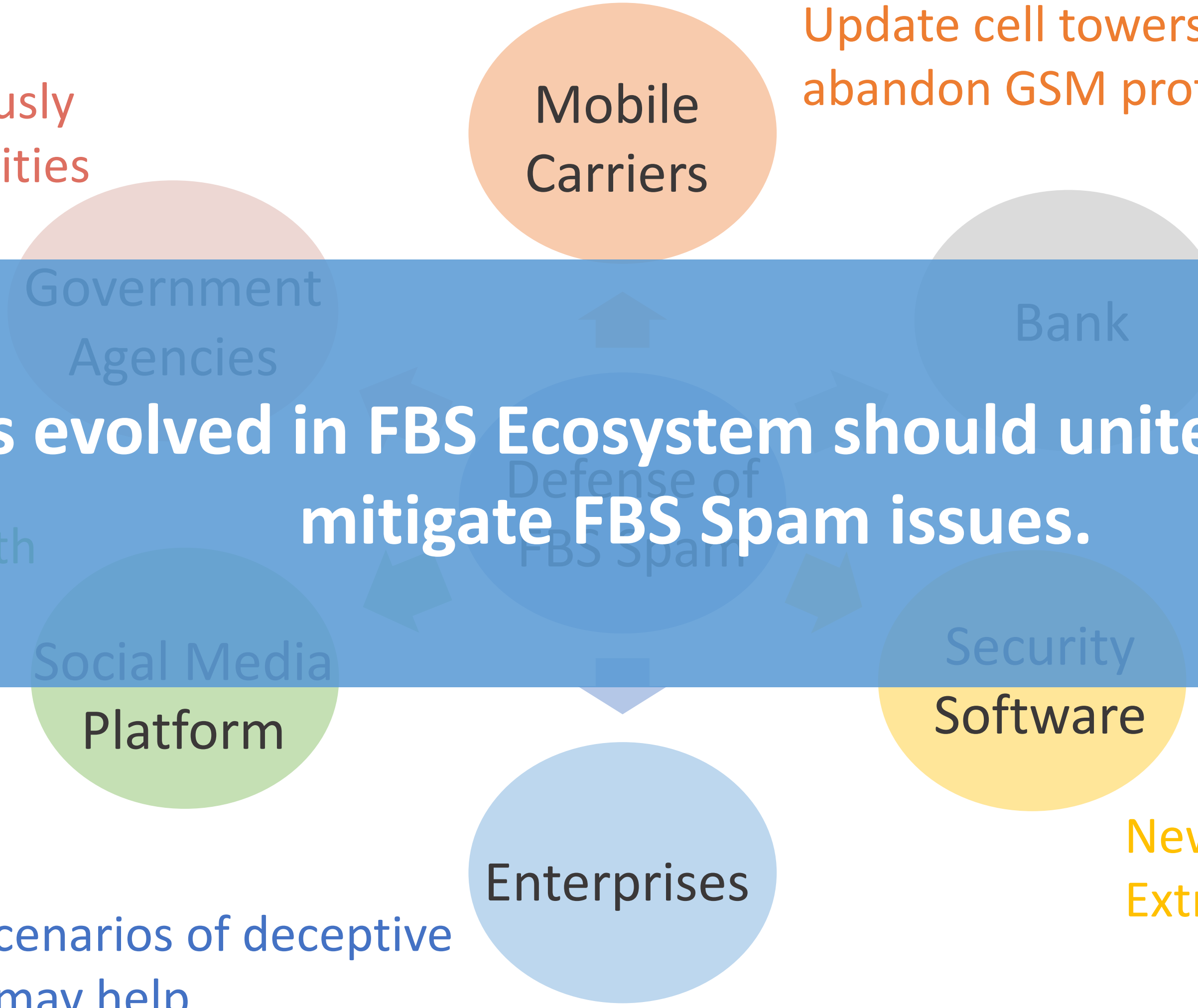


Cooperation of government departments across cities is necessary to combat campaign migration

Recommendations for the Community

More efforts in seriously effected places and cities

Update cell towers, abandon GSM protocol



Re-evaluate their bank card policies to avoid being abused

Checking accounts with fraudulent activities

New UI system of application
Extracted templates as new features

User education, new scenarios of deceptive messages in our work may help

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

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