Information Security and Computer Fraud, 2017, Vol. 5, No. 1, 1-8
Available online at http://pubs.sciepub.com/iscf/5/1/1

DOI:10.12691/iscf-5-1-1

Detecting and Tracking Pseudo Base Stations in GSM Signal Hijacking and Frauds: a Visualized Approach

Yongxing Li[1], Yang Heng[1], Ankang Hao[1], Tianxing Wang[1], Xiaojie Liu[2], Lan Huang[1,*]

[1]College of Computer Science, Yangtze University, Jingzhou, Hubei, China
[2]Beijing Gehua CATV Network Co. Ltd., Beijing, China
*Corresponding author: lanhuang@yangtzeu.edu.cn

Abstract Pseudo base station (PBS), sometimes called fake base station, refers to cellular base stations that are employed for malicious and usually illegal purposes. Through the pitfalls of the GSM protocol, PSBs can hijack GSM signals of cellphones close by. Most PBSes are portable, for example hidden in vans or even carried in backpacks, and are deployed in densely populated regions. Then they can steal personal information from neighboring smartphones, or send intriguing messages to them that would ultimately lead to telecom frauds. In recent years, there has been a terrifying increase in the number of telecom frauds and the smartphones infected by viruses sent from PBSes. This urgently calls for methods and systems that can effectively identify and track PBSes. In this study, we designed and implemented a PBS detecting and tracking system, by conducting topic analysis of messages received by cellphones and analyzing their temporal and spatial distribution patterns. Using the system, we could perform a variety of exploratory analysis, including categorizing PBSes into either stationary or moving PBSes, discovering and visualizing their behavior patterns, and identifying districts that tend to suffer from a particular type of fraud messages.

Keywords: *Pseudo Base Station, telecom fraud, topic modeling, trajectory clustering, visualization*

Cite This Article: Yongxing Li, Yang Heng, Ankang Hao, Tianxing Wang, Xiaojie Liu, and Lan Huang, "Detecting and Tracking Pseudo Base Stations in GSM Signal Hijacking and Frauds: a Visualized Approach." *Information Security and Computer Fraud*, vol. 5, no. 1 (2017): 1-8. doi: 10.12691/iscf-5-1-1.

1. Introduction

"Congratulations! You have just won 500 dollars!". "You can redeem your credit card points from us!" "Need loans? No mortgage required!" It is increasingly common to receive such messages nowadays. People with good IT and risk awareness usually can make the right choice: delete or simply ignore such messages. Unfortunately, a considerable proportion of people, for example the elderly and university freshmen, who are reasonably new to the smartphone technology and society, and thus lack sufficient knowledge and experiences in telecom frauds, are likely to fall for these intriguing messages. Ultimately, some of them could become victims of this uprising kind of telecom crime and suffer both mentally and financially from great losses.

Where did these messages come from? How did their senders know my cellphone number and my personal details? Victims usually ask such questions afterwards, because the scammer seems to know everything about the victim, and this is usually the critical part that eventually tricks the victims into dispelling all their doubts and falling into traps. Recent investigation revealed that pseudo base stations were the weapon being used to send such scam messages and illegally collect personal information.

Pseudo base station (PBS), sometimes called fake [1] or malicious base station [2], or IMSI catcher [3], refers to cellular base stations that are employed for malicious and usually illegal purposes. Through the pitfalls of the GSM protocol, PSBs can hijack the GSM signals of cellphones in its neighboring area. Then they can steal personal information from neighboring smartphones, or send intriguing messages to neighboring cellphones that would ultimately lead to telecom frauds. To make things worse, most PBSes are portable and moving: for example fraudsters can hide PBSes in vans or carry them in backpacks and drive them around the city or just wander in densely populated regions.

In recent years, the number of telecom frauds and the number of smartphones infected by viruses sent from PBSes have increased terribly. This urgently calls for methods and systems that can effectively identify and track PSBs. In this study, we designed and implemented a PBS detecting and tracking system, by conducting topic analysis of messages received by cellphones together and by analyzing their temporal and spatial patterns. Using the system, we could effectively perform a variety of exploratory analysis, including categorizing PBSes into either stationary or moving PBS, discovering and visualizing their behavior patterns, and identifying districts that tend to suffer from a particular type of fraud messages.

The rest of this paper is organized as following. Next we review related work on PBS detection. Section 3

introduces the architecture and working scheme of PBSes. Section 4 describes the system design and implementation details, and the core components of the system in detail: the topic modeling method and the trajectory clustering algorithm. Section 5 presents experimental setup and discusses empirical experimental results. Section 6 concludes the study.

2. Related Work

Text message scams are also called “smishing” (Short Message Service + phishing) scams [4], because they like phishing emails. Such messages often contain a web site link or a phone number. When the recipient connects to the web site, a Trojan virus will be installed on that smart phone, providing hidden access to third parties. Phishing phone numbers are even worse. When the recipient calls the number, he/she will either be charged seriously for that call, or enter a carefully knit chain of scams. Literally, such scams often prey on people’s panic or sense of urgency. For example, in a recent case, suspects, disguised as court or police officials, claimed that the victims' bank accounts had been breached, inducing them to transfer their money to a so-called "safe account" in panic [5]. Victims usually suffered from huge financial and mental losses, some even lost their lives, causing serious social problems. One gang recently being cracked involved nearly 15 million (in US dollars) illegal gains [6].

IT techniques have been exploited to find solutions to fight text message scams. Hilas developed an expert system for fraud detection in private networks [7]. Borgaonkar et al. [3] developed a framework for evaluating PBS detection applications currently available. Topic modelling is usually used to identify spams in SMS messages as well as in many related scenarios such as emails [8]. More recently, Joo et al. [9] designed a Naïve Bayes classifier to distinguish smishing messages from normal messages, also based on their content.

Recently, it has been revealed that PBSes were the tool behind many of such scams [1]. Instead of focusing on anti-spam apps to intercept scam messages, many researchers start to work on systems that can identify the PBSes and solve the problem from its root causes [2,3,9,10,11,12]. Our study presented in this paper takes this approach.

3. Working Mechanism of PBSes

PBS usually use the GSM900 standard, with an emission power rate between 40-43dBm. Figure 1 shows the general architecture of a GSM PBS [12]. A portable PBS usually consists of an antenna, a diplexer, a power amplifier, a transceiver and a central control unit.

PBS will disguise itself as a legal telecom network by using the same broadcast network number and frequency, and a different location setting. Because PBSes provide far stronger GSM signals than the valid base stations, cellphones will quickly be lured into the fake network to update their location information once they enter the coverage area of PBSes.

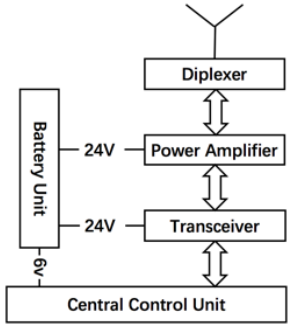


Figure 1. Architecture of a Pseudo Base Station

Unfortunately, cellphones usually cannot tell the difference between a valid base station and a PBS, because the current GSM protocol only requires one-way verification: communication networks need to verify the connected terminals, but none verification is needed at the terminal side. In other words, for a cellphone, a nearby PBS looks exactly the same as a valid one, or even a better network due to its greater signal strength. Upon joining the fake network, a large number of scam messages will be sent to these cellphones, including fraudulent information or links to viruses.

PBSes can be categorized into two types in general: stationary PBS that usually hides in a fixed location (e.g. a hotel room or a rental property), and mobile PBS that usually hides in backpacks, suitcases or vans and travels in cities. Stationary PBSes are usually larger and more powerful than mobile PBSes. Yet they are much easier to detect, as we will see later.

4. System Design and Implementation

Our system mainly consists of two components: topic analysis and trajectory clustering. This section describes each component in detail.

4.1. Topic Analysis

We analyzed the content of each scam message and classified it into one of eleven manually defined topics. Meanwhile, depending on the potential financial loss that it might cause, scam messages were also categorized into three risk levels: severe, moderate and mild. Table 1 lists the eleven topic categories and their corresponding risk levels. Figure 2 demonstrates the distribution of each topic category and risk level in our experimental dataset.

Table 1. Topic Categories and Risk Levels

Risk Level	Topic Category
Severe	Fake Invoices
	Credit Cards
	Loans
	Bank Fraud
Moderate	Telecom Promotion
	Loss and Found
	Stocks
Mild	Real Estates
	Advertisements
	Sex Information
	Others

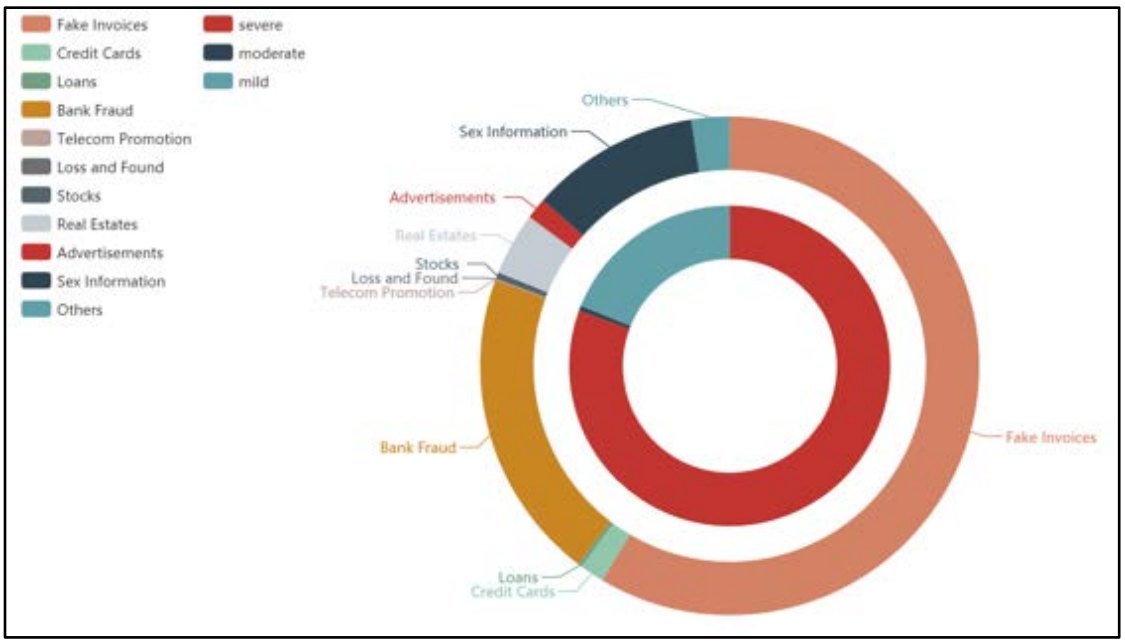


Figure 2. Distribution of Topic Categories and Risk Levels

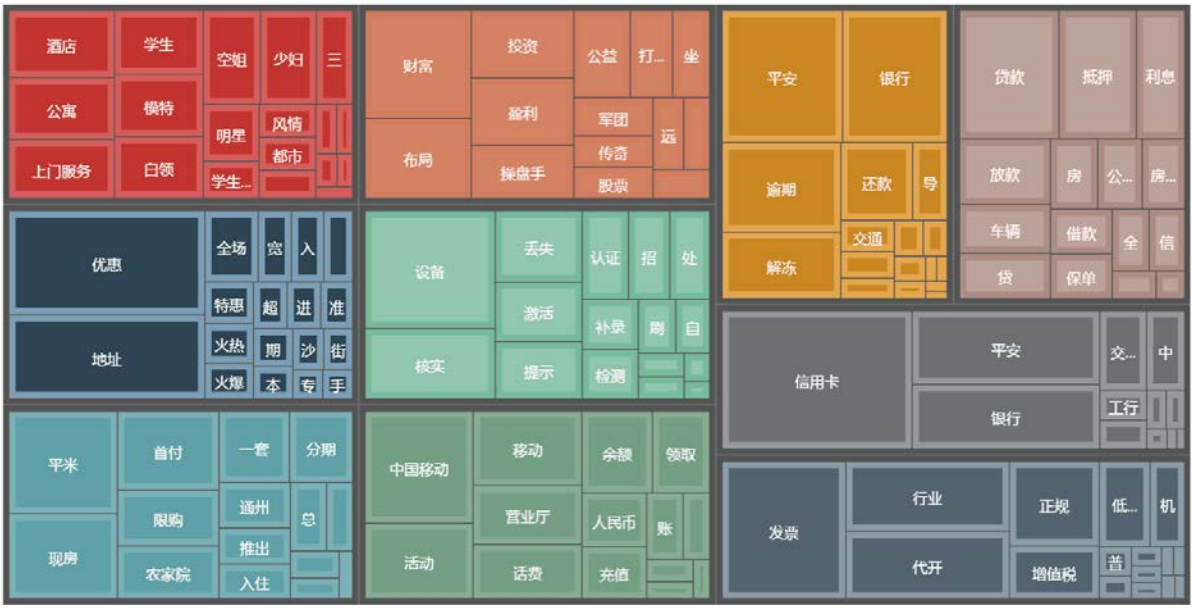


Figure 3. Top Keywords in Topic Categories

Topic modelling is the basis of message categorization. When split into words, our experimental dataset generated a vocabulary of 12K words. Because SMS messages are very short, the resulting word vector space was too sparse for effective topic modelling. To address this problem, we employed the distributed word representation scheme [13] and generated the word vector representation for each word. Then words relating to the same topic can be grouped together based on their word-vector representations. Figure 3 shows the top keywords of the first ten topic category (i.e. except for the Others category).

Specifically, we first trained the word vector model from the entire dataset. Similarity between two words were calculated using the cosine similarity measure of their corresponding word vectors. Then one seed word was selected for each topic category, and was expanded with the top 20 words that yield in the greatest similarity with the seed word, resulting the keyword set shown in Figure 3.

4.2. Trajectory Clustering

The spatial and the temporal information associated with each message, i.e. the latitude, the longitude and the time when the message was received, provides valuable information for categorizing and locating PBSes. We noticed that many of the scam messages share the same content and thus have the same MD5 value. We then grouped all messages by their MD5 values, and clustered them based on their spatial and temporal features.

Figure 4 illustrates this process. Basically, every message collection consists of messages with the same content (i.e. the same MD5 value) and sorted by time in ascending order. For each such collection, we identified tracks by combining messages that were received within 30 minutes and 1000 meters apart at most.

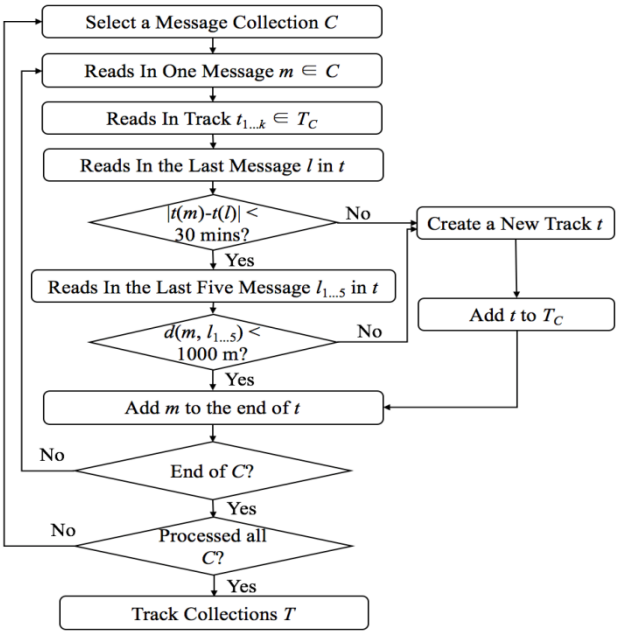


Figure 4. Trajectory Clustering Process

5. System Evaluation and Discussion

5.1. Experimental Dataset

The experimental dataset used in this study came from the first challenge of the fourth Visualization Contest of China (ChinaVis) [14]. The dataset contains over three million (3,358,952) messages collected in Beijing from February 23, 2017 to April 26, 2017. These messages were spam messages reported to a particular smartphone security application. Each data record consists of seven fields, as explained below.

- Phone: sender's (the PBS) phone number;
- Content: message;
- MD5: MD5 value of the message;
- ReciTime: timestamp of receiving the message;
- ConnTime: timestamp of establishing the connection with the PBS;
- Lng: (approximate) longitude of the location where the message is received;
- Lat: (approximate) latitude of the location where the message is received.

Based on data in these fields and our system, we performed exploratory analysis from several aspects. The remaining of this section explains the details and discusses their results.

5.2. General Temporal and Spatial Patterns

Figure 5 shows the general active days and time periods of PBSes. The general pattern of PBSes obeys the commute patterns of most people: most active around 10am and 6pm-8pm. Yet, there is no obvious differences between days of a week.

Figure 6 shows a heat map generated by plotting messages' geolocations onto the map of Beijing. When combined with Beijing's administrative divisions, it becomes obvious that city east (e.g. the Chaoyang and the Chongwen districts) and city north (e.g. the Haidian district) attract more PBSes than other districts. East is the central business district and the entertainment center of Beijing where theaters, restaurants, hotels and pubs locate, whereas the Haidian district is a gathering place of educational institutes and IT companies.

Figure 7 combines the temporal and the spatial information together, showing the amount of messages in different geographical divisions at different times. It is quite clear that around midnight time, only one district still has a decent traffic and thus still suffers from PBSes (i.e. the Chaoyang district).

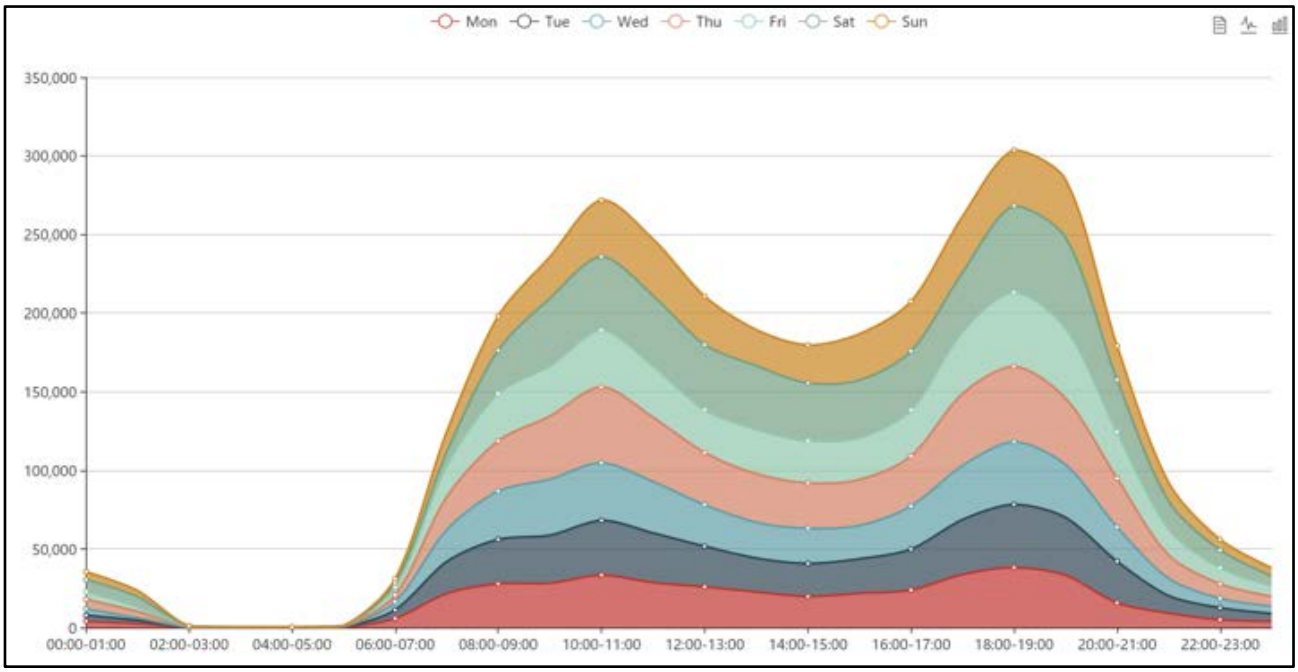


Figure 5. Active Days and Time Periods of PBS

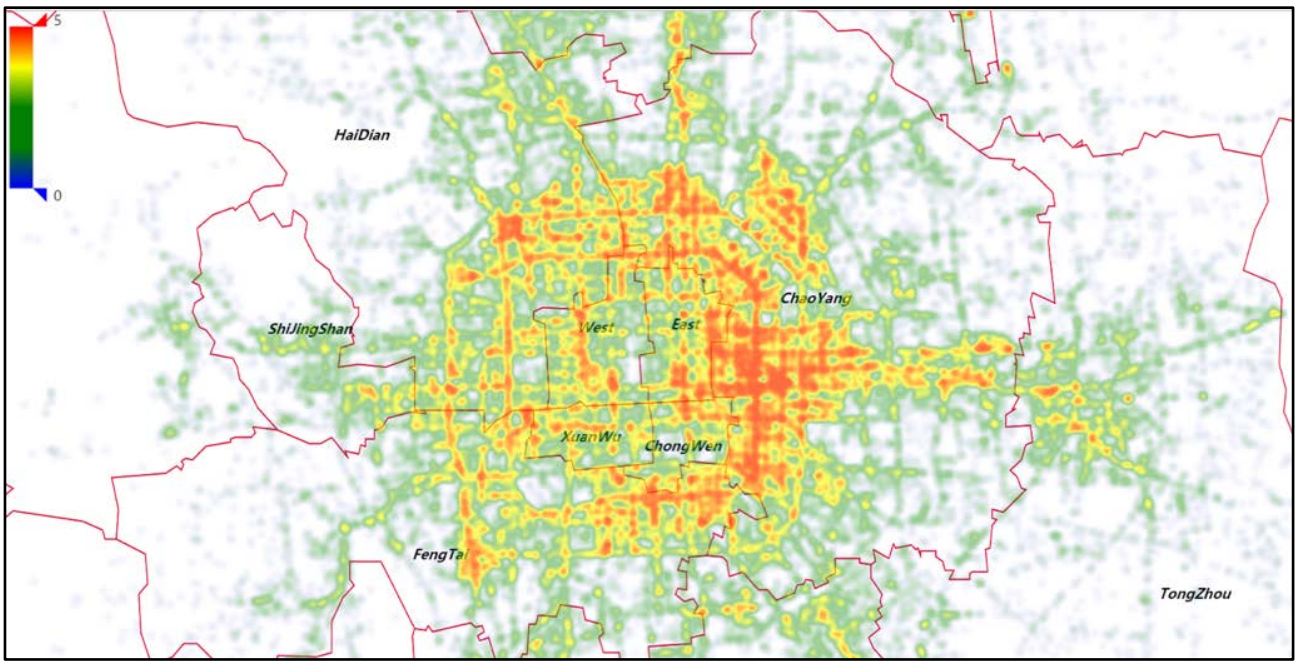


Figure 6. Heat Map of Scam Messages in Different Administrative Divisions of Beijing

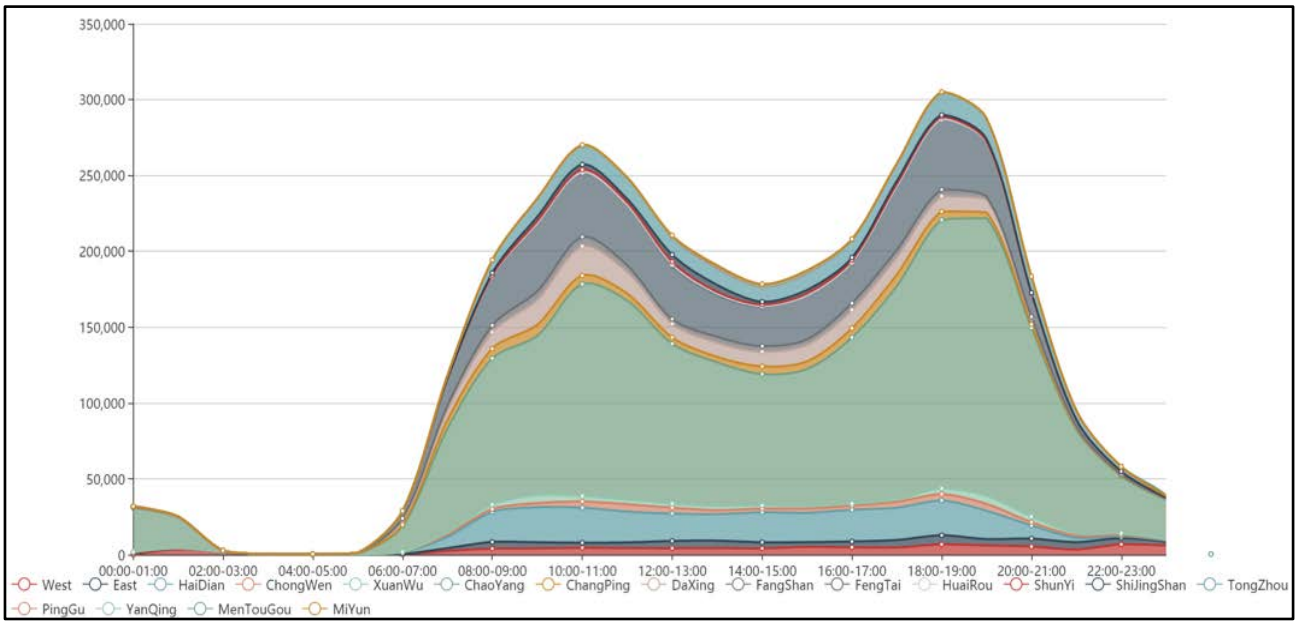
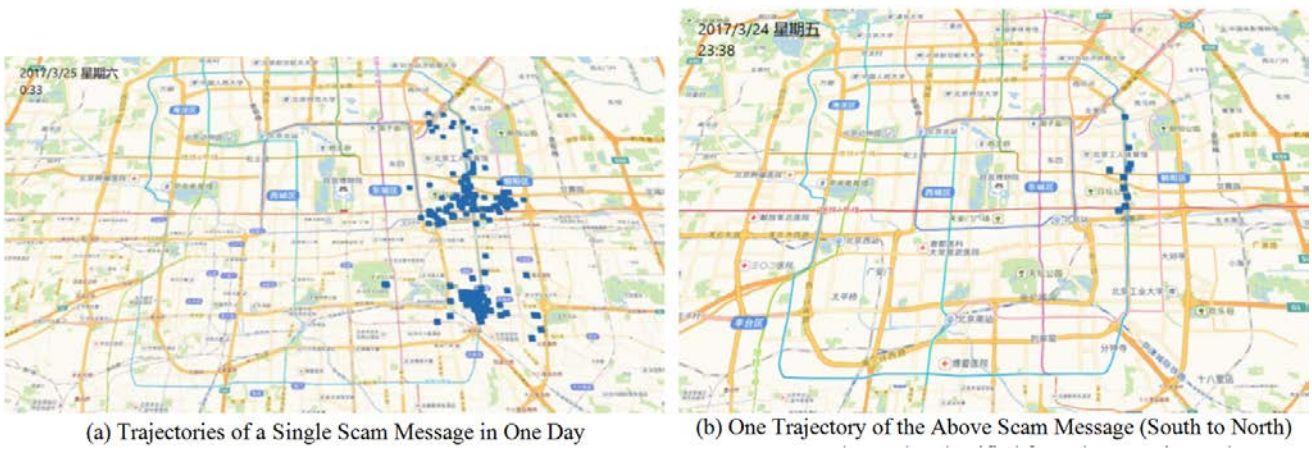


Figure 7. Distribution of Scam Message in Different Geographical Divisions at Different Time Periods



(a) Trajectories of a Single Scam Message in One Day

(b) One Trajectory of the Above Scam Message (South to North)

Figure 8. Example Tracks Identified from the Experimental Dataset

5.3. Tracking Mobile PBSes

As explained in Section 4.2, we clustered scam messages based on their content, longitude, latitude and time. Then message clusters close in both time and space were grouped together to form tracks. Figure 8 shows the tracks of a single message in one day and one specific track identified during midnight. Each blue square represents one message, and they can overlap.

In total, we extracted 359 tracks from the top ten popular scam messages. When combined with detailed city maps, it showed that PBSes tend to travel along major streets or along subways. This provides useful information for tracking down and finally cracking these PBSes.

5.4. Locating Fixed PBSes

When generating PBS tracks, we found that some PBSes seemed to be fixed at one location. Empirically, we found a considerable proportion of certain messages with quite concentrated locations. Some messages even occurred at the same location yet on different dates. We then applied two clustering algorithms DBScan [15] and K-means, and found 95 suspected fixed PBSes, as shown in Figure 9. Specifically, we first clustered messages that were received within 1000 meters. Then for each cluster, we applied the DBScan algorithm to find the dense regions. Finally, K-means was applied to locate the cluster center, i.e. the location of the suspected fixed PBS.

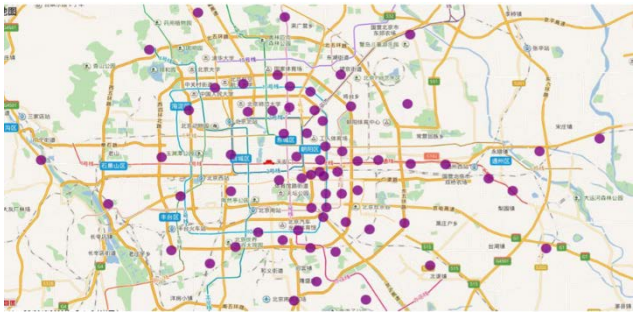


Figure 9. Fixed PBSes Identified from the Experimental Dataset

5.5. Identifying Gangs of PBSes

We also identified gangs of PBSes: PBSes that collectively travelled in different parts of the city at the same time and broadcast the same message. For example, we found such a gang when analyzing a particular message (MD5 d4acefe46710e6fcbf38a6ce6c82a1bf, about fake invoices) that occurred 48,954 times in merely two days (March 3-4, 2017). This message accounted for about 17% of the total number of scam messages reported in those two days.

Figure 10 shows the amount of this message reported from different locations within one minute. The minimum distance between two message clusters was 3 km. Some even belonged to different administrative divisions. This suggested that different PBSes were hired to propagate the same message, probably through the same organization.

Figure 10. Gangs of PBSes Simultaneously Broadcast the Same Message in Different Regions

5.6. Temporal and Spatial Patterns of Message Topics

Figure 11 shows the distribution of different topic categories in different times and days. It is quite clear that sex-related information (the grey section) was the most active category at night. Messages about fake invoices (the red section) usually became the most active during morning and afternoon peak hours, probably targeting people in the workplace. This is also the reason why this category is more active during working days. In contrast, bank fraud (the orange section) was more active during weekends, for example to target the elderly and the students.

Spatial-wise, Figure 12 shows a similar distribution of topic categories as in Figure 6: the worst affected areas were Chaoyang, Fengtai and Haidian. The real-estate-related information (the light yellow section) and the sex-related information (the grey section) tend to concentrate in the Chaoyang district. This is probably because many entertainment facilities and financial institutes are located in this district.

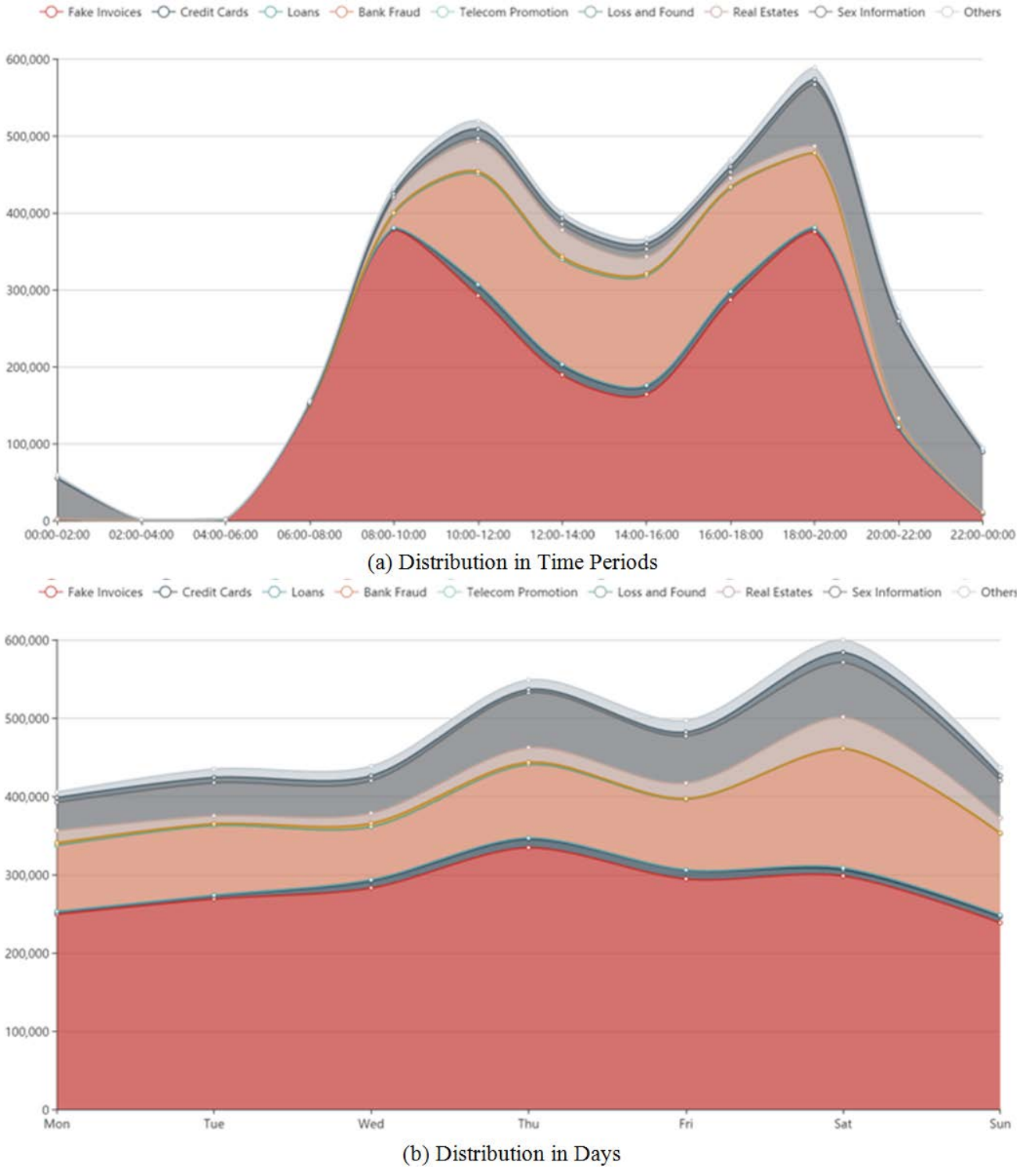


Figure 11. Distribution of Topic Categories in Time and Days

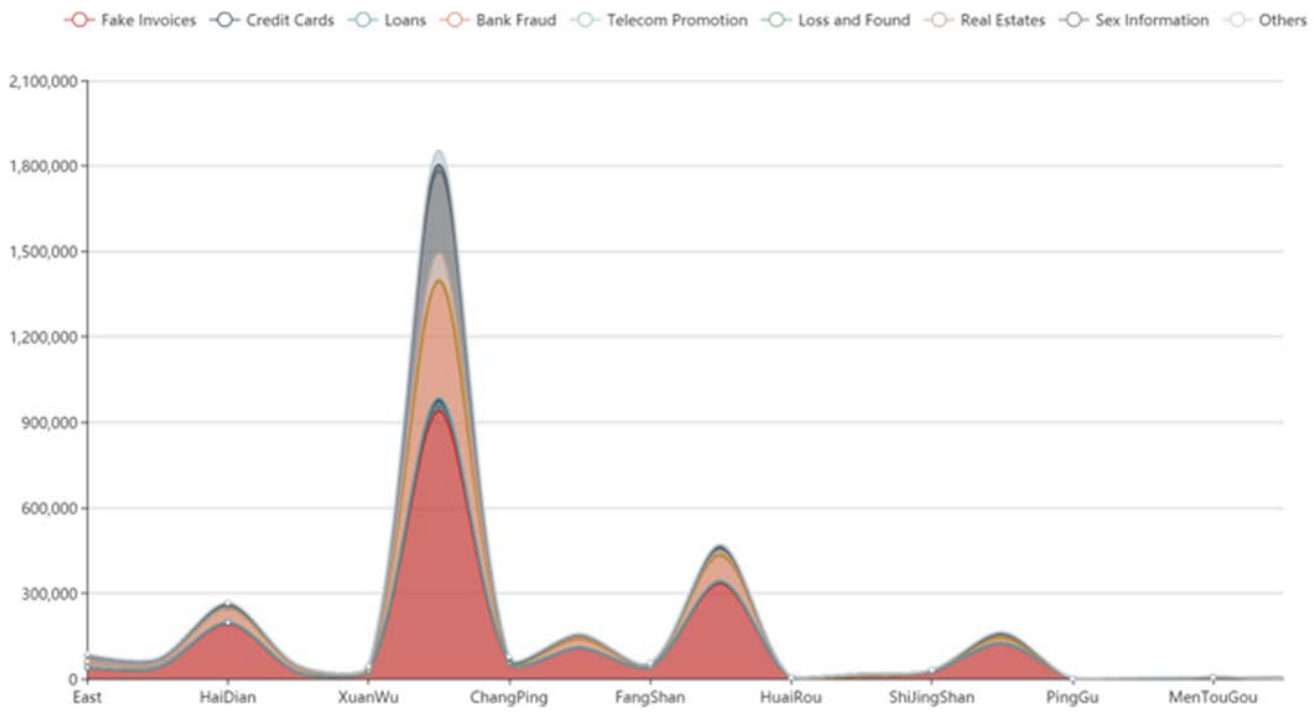


Figure 12. Distribution of Topic Categories in Geographical Divisions

6. Conclusions

Further analysis also showed a strong positive association between GDP, population size and the number of scam messages. Our analysis results can provide valuable insight into the behavioral patterns of PBSes. By categorizing PBSes into mobile and fixed PBSes and by discovering the suspected gangs of PBSes, our findings could be used to assist further detection and cracking of PBSes.

Acknowledgements

This study was funded by Yangtze University (grant no. JY2014032 and 2015cqn52).

References

[1] Cappella, N.. Fake Mobile Base Stations Used to Spread Malware in China. [Online]. Available: https://thestack.com/security/2017/03/23/fake-mobile-base-stations-used-to-spread-malware-in-china/ [Accessed Aug. 20, 2017].

[2] Zhang, C.. Malicious Base Station and Detecting Malicious Base Station Signal. Communications System Design, 2014, pp. 59-64.

[3] Borgaonkar, R., Martin, A., Park, S., Shaik, A., Seifert, J.-P.. White-Stingray: Evaluating IMSI Catechers Detection Applications. In: 11th USENIX Workshop on Offensive Technologies, 2017.

[4] Kang, A., Lee, J. D., Kang, W. M., Barolli, L., Park, J. H.. Security Considerations for Smart Phone Smishing Attacks. In: Jeong H., S. Obaidat M., Yen N., Park J. (eds) Advances in Computer Science and its Applications. Lecture Notes in Electrical Engineering, vol 279. Springer, Berlin, Heidelberg.

[5] ChinaDaily. 124 Arrested for Cross-Border Telecom Fraud [Online]. Available: http://www.chinadaily.com.cn/china/2017-07/27/content_30267342.htm [Accessed Aug. 20, 2017].

[6] ChinaDaily. 77 Telecom Fraud Suspects Returned to China from Fiji [Online]. Available: http://www.chinadaily.com.cn/china/2017-08/05/content_30349375.htm. [Accessed Aug. 20, 2017].

[7] Hilas, C. S.. Designing an Expert System for Fraud Detection in Private Telecommunications Networks. Expert Systems with Applications, 2009, 36 (9): pp. 11559-11569.

[8] Bergholz, A., Paaß, G., Reichartz, F., Strobel, S., Chang, J. H.. Improved Phishing Detection using Model-Based Features. In: Fifth Conference on Email and Anti-Spam, 2008.

[9] Joo, J. W., Moon, S. Y., Singh, S., Park, J. H.. S-Detector: an Enhanced Security Model for Detecting Smishing Attack for Mobile Computing. Telecommunication Systems, 2017, 66(1): pp. 29-38.

[10] Ney, P., Smith, I., Cadamuro, G., Kohno, T.. SeaGlass: Enabling City-Wide IMSI-Catcher Detection. In: Proceedings on Privacy Enhancing Technologies, 2017(3).

[11] Li, Z., Wang, W., Wilson, C., Chen, J., Qian, C., Jung, T., Zhang, L., Liu, K., Li, X., Liu, Y.. FBS-Radar: Uncovering Fake Base Stations at Scale in the Wild. In: 23rd Network and Distributed System Security Symposium, 2017.

[12] Meng, J.. Detection and Location of GSM Pseudo Base Station Based on Software Defined Radio. Master Thesis, 2016, Lanzhou Jiaotong University.

[13] Goldberg, Y., Levy, O.. word2vec Explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint arXiv:1402.3722, 2014.

[14] ChinaVis 2017. [Online]. Available: http://chinavis.org/2017/challenge.html. [Accessed Aug. 20, 2017].

[15] Ester, M., Kriegel, H. P., Sander, J., Xu, X.. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: 2nd International Conference on Knowledge Discovery and Data Mining, 1996, pp. 226-231.