Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Temporal and Spatial Analysis of Fake Base Stations

To cite this article: Yue Jia *et al* 2018 *J. Phys.: Conf. Ser.* **1060** 012019

View the article online for updates and enhancements.

Related content

- Rotation, Reflection, and Frame Changes: Computer graphics visualization
R M Brannon
- Interactive visualization system to analyze corrugated millimeter-waveguide component of ECH in nuclear fusion with FDTD simulation
N Kashima, H Nakamura, Y Tamura et al.
- Development of Visualization System of Neutral Particles Generated from Laser-Produced Plasma for an EUV Light Source
Hiroki Tanaka, Atsushi Matsumoto, Akihiko Takahashi et al.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 185.101.69.197 on 21/11/2018 at 09:43

IOP Conf. Series: Journal of Physics: Conf. Series **1060** (2018) 012019 doi:10.1088/1742-6596/1060/1/012019

Temporal and Spatial Analysis of Fake Base Stations

Yue Jia, Bo Lyu* and Yangyang Li

Innovation Center China Academy of Electronics and Information Technology

*blv@csdslab.net

Abstract. With the development of information technology, send spam messages by using the fake base station has become a new means of criminal and marketing advertisement. Fake base station usually dissembles itself as numbers which stand for banks, operators, state organs and forcible and sends fraud, pornography, gambling, advertising and other short messages to users. This action affects the normal communication and also has a serious violation of public security. Based on sample data sets of junk text messages opened by ChinaVis2017 Challenge, this paper develops a temporal and spatial trajectory analysis and visualization system for stopping fake base station. Moreover, this paper deeply excavates the temporal and spatial laws of the location and movement of fake base stations. Based on the analysis results of fake base station behaviour patterns, this paper provides several recommendations of tracking and arresting fake base station owners. This paper firstly introduces the realization scheme and key technology of spatial-temporal trajectory analysis and visualization system of fake base stations. Then, it analyses the temporal and spatial trajectory distribution, the content classification and the degree of harmful of fake base stations, and finally analyses the actual demand of fake base stations supervision and management with several practical solutions.

1. Introduction

Currently, fake base station is a means of telecom fraud high-tech equipment and it is mainly composed of hosts and notebook computers. It searches for a certain radius of the phone card information, and fraudulently imitates any other phone number sending fraud, advertising and other short messages to compromised users. When such device is running, the compromised users' mobile phone signal is forced to connect to the fake base station and they cannot connect to the public telecommunications network. Under this circumstance the normal mobile phone services cannot be performed [1].

As long as users are forced to connect to a fake base station, the fake base station can unlimitedly send any content of the message to the users. Those advertising short messages, fraudulent short messages and sensitive content messages could do great harm to security. Even if the spam message does not successfully lead to a crime, fake base station also affects the users' normal operator service as the user's mobile phone signal is forced to connect to the device of fake base station. Mobile phone users will temporarily off the public network in about 8 to 12 seconds and after that they may return to normal. However, some mobile phone users must restart their devices to re-enter the network. In addition, the fake base station will lead to mobile phone users frequently update the location, occupying the region's wireless network resources and the emergence of network congestion, affecting the normal communication of users. Harms as the fake base station may arise show as following:

a) Counterfeit bank customer service phone. For example, counterfeiting Industrial and Commercial Bank of China, fake base station will pretend as 95588 and sign the number at the end of

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

the message. Short messages content might include prompting user password failure and a modify is needed by user to log in the site shown in the message, or indicating user to log on the site to exchange cash and other tricks.

b) Guide the user to log on to a phishing site. The fake base station can imitate the bank customer service number which is exactly the same number that the real bank customer service used. Many users will trust and click on the relevant link. At this time, users will be led to a fishing sit which is really similar to the bank's website.

c) Get users' confidential information. When logging on the phishing site, the fraud may guide users step by step to enter their information. Those information including bank card account number, mobile phone number and login password. If the fraudster wants to get more information about the victim, they will be on-line and offer more guidance to the users.

2. Theory and challenges

Fake base station occupies the public radio communication system network number, frequency resources, and so on, to disguise itself as the public radio communication network base stations neighbourhood. It uses public radio communication network to monitor signalling. In the normal system monitoring radio communication process, the phone searches for a base station signalling every five seconds, and interacts with a number of nearby base stations at the same time. According to the base station signal strength, the user selects the nearest and strongest base station. However, one of the easiest disguise methods is through the instantaneous signal power transmission, then the fake base station can take place of normal base station. After that, it can and automatically collect coverage within the public radio communication network mobile phone users mobile phone number and the current location information. Then, fake base station will earn trust from user and fraudulent use of other mobile phone numbers to these off-line mobile phone to send pre-set business customized advertising messages [2, 3].

Fake base station can be carried by a person or a vertical. So, it is hard to trace or arrest owners of fake base stations. Moreover, one fake base station can send as many as 50000 short messages per hour. Not only GSM users, but also 3G network and LTE network users may be compromised by fake base station. Moreover, there are two problems that make it impossible to use it as a sign to identify the fake base station trajectory. One problem is that a number of fake base stations may masquerade as the same number (for example, 95580 which is the service number for Postal Savings Bank of China, 10086 which is the service number of China Mobile, etc.), and another problem is that the same fake base station may use multiple camouflage numbers at the same time. Finding an efficiency searching system to tracing and locating fake base station becomes a crucial task for public security.

3. Data set analysis

The sample data set, with which this paper processed, has 63 days of junk text messages for Beijing, the capital city of China, from the data February 23th to April 26th of the year 2017. All messages are collected through user reporting mechanism. In total, there are 3358952 lines of data and each of them including message content, fake phone number, message receiving time, user reporting time and the longitude and latitude of the reporting location. The total number of different fake phone number is 11371.

In the part of the data preparation, in addition to match the data provided, we also use the following additional data sources: (a) Beijing city map data that provided by OpenStreetMap, to offset the existence of competition between open data and OSM data of manual correction; (b) in order to study the regularities of distribution of content of short messages, we use LBS API provided by Baidu to map the longitude and latitude coordinates in the data set. In the data cleaning section, the exception points of latitude and longitude beyond the geographical boundaries are eliminated, and data items beyond the scope of the study are eliminated [1].

In the cleaned data, the sender's telephone number disguised by the pseudo base station is the identification of a spatial-temporal trajectory, but there are two problems that make it impossible to

use it as a sign to identify the pseudo base station trajectory. The problem is that a number of pseudo base stations may masquerade as the same number (for example, 95580, 10086, etc.), and the problem two is that the same pseudo base station may use multiple camouflage numbers simultaneously. In order to pseudo base station trajectory visualization, each with the phone number labelled according to the spatial-temporal trajectory space of Euclidean distance for clustering, multi path differential behind the same period with the same numbers, so as to solve the problem. Then, by using the similarity of trajectories, the trajectories which are very close to the laws of time and space are aggregated into the same pseudo base station, so as to solve the second problem [4].

3.1. Content Categories

In order to analyse the spatial-temporal regularities of fake base station moving patterns and messages' occurring regulars, this paper analyses the word segmentation and the semantic of the short message and extract the key words of the content of the message. After that, the content of the message is classified and clustered according to the distance of the keyword. By applying above classification methods, the short messages will be divided into fake invoices, posing identity, pornographic, real estate advertising, contraband promotion, employment advertisements, illegal gambling, sales advertisements, education immigration and other types of ten categories.

3.2. Spatial Distribution Regularities

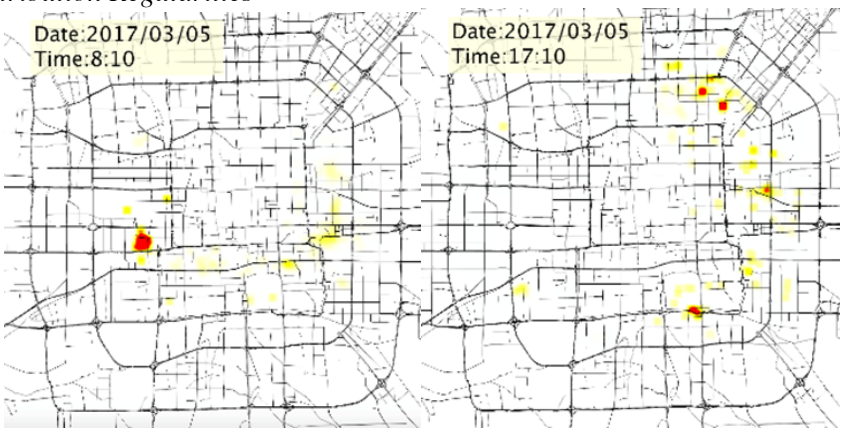

Figure 1: Samples of fake base station activity map.

From the Figure 1, we can see the temporal and spatial trends of the distribution of fake base stations. For example, on the day of March 5th, much more short messages were send by fake base stations, and more concentrated in a few hot spots (such as office buildings, shopping district, etc.). Vice the versa, one the day of April 5th, which is the tomb sweeping holidays, as most people rested at home or went out of town for tourism, the total number of messages has a significant reducing. Moreover, the distribution is more uniform. From the laws of the thermos gram, it can be concluded that the fake base station is very cautious and targeted when selecting the location and the sending time, especially in relation to the daily behaviour of the public. The thermal map is intuitive to see the distribution of fake base stations, but for the fake base station trajectory patterns, this paper provides more fine granular analysis means.

In raw data, the sender's telephone number disguised by the fake base station is the identification of a spatial-temporal trajectory. However, there are two issues that prevent us from using it as a marker to identify the fake base station trajectory. The first problem is that there are multiple fake base stations may disguise the same number (such as 95580,10086, etc.), the other problem is that the same fake base station may also use a number of camouflage numbers.

In order to visualize the trajectories of the fake base stations, each time-space trajectory marked with the telephone number is clustered according to the Euclidean distance in space, and the multiple trajectories behind the same number are identified in the same period to solve the problem. And then

use the temporal and spatial similarity of the trajectory to gather the trajectories of the spatially close to the same fake base station to solve the problem.

Through the above spatiotemporal data analysis method, the action trajectory of the fake base station can be effectively extracted. On this basis, the team uses Processing, which is a flexible software sketchbook and a language for learning how to code within the context of the visual arts [5], to construct the space-time trajectory query and visualization system. The system supports the basic functions such as zoom out, track searching and tracking visualization. While supporting the map style selection, mouse latitude and longitude and other auxiliary functions. The user can enter the phone number in the interactive interface to retrieve the track and specify the visual colour of the track. In the visual trajectory, the traces of the thickness and saturation of the fake base station to send the number and frequency of text messages.

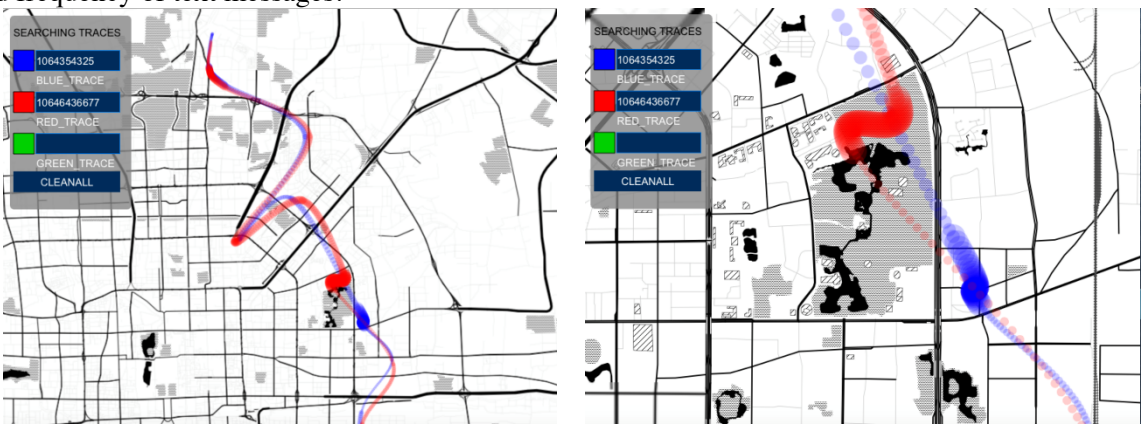

Figure 2:One fake base station forges different phone numbers

This paper founds that many trajectories have a high spatial and temporal similarity. As can be seen from the Figure 2, the same fake base station operator uses different cell phone numbers. To enlarge the map can be found (as shown on the right), the fake base station operator in parking layout of the north gate of the park and the South Gate of the fake base station, after a period of time by sending two mobile phone number, quickly leave the place.

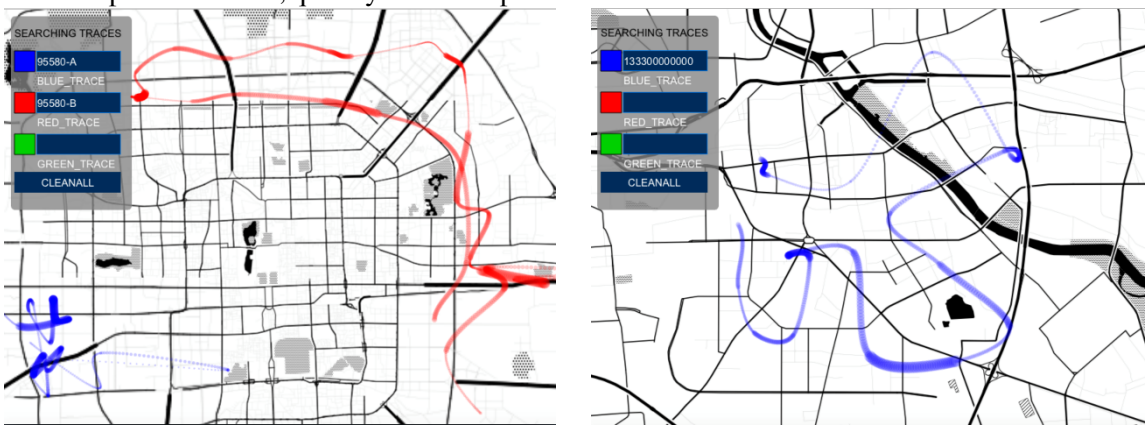

Figure 3:One phone number forged by different fake base stations

In data analysis, the authors found that many fake base stations forged the same number to send text messages. In particular, the identity of the class like text messages like to use 10086, 95580, 95560 and other service class number. By means of spatial-temporal clustering can separate the trajectory of the same time distance, as shown on the left, the distinction between the two 95580 track of Chaoyang District and Haidian District.

Different fake base station hidden traces of the means are also different. The trajectory can be classified by calculating the distribution of the upper coordinate points of a trajectory. For example, some tracks are used to move around a place, or even stationary, and focus on sending messages at

specific times (for example, rush hour), while the rest of the time hides the movement. While some fake base is to keep moving, usually along the fourth ring road along or send text messages as shown in Figure 3.

3.3. Temporal Distribution Regularities

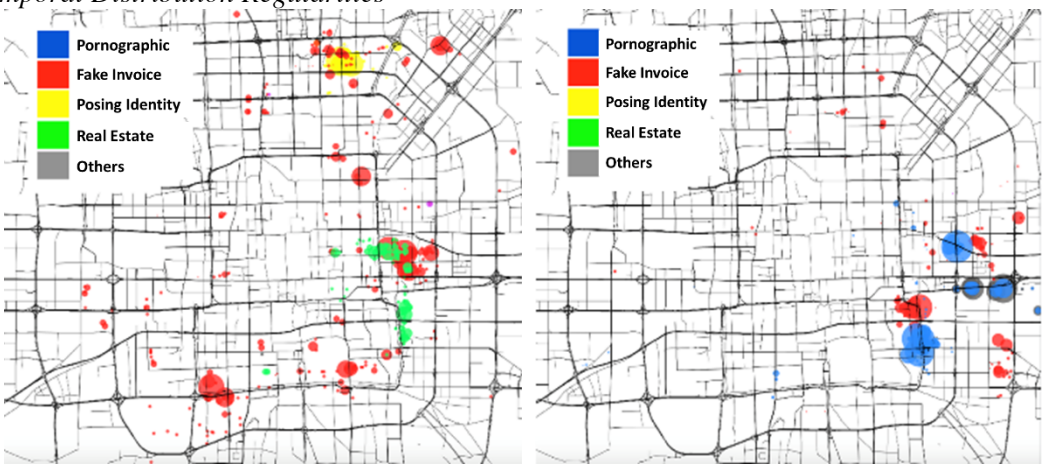

Figure 4:Samples of the temporal and spatial distribution of all kinds of short messages

In this paper, the macro situation of all kinds of text messages is displayed intuitively. As shown in Figure 4, the colour of scatter pictures represents the type of SMS, and the size of scatter points represents the number of messages. In February 23rd, for example, will be divided into four periods, each period from each of a representative point in time, as shown below, the morning pornographic messages as the main messages sent by the fake base, but mainly concentrated in the Sanhuan nearby bustling business district. The invoice information of the invoice is more evenly distributed, and exists from morning till night. February 23rd for the working day, the rush hour is also the fake base station to send short messages peak, real estate advertising and processing invoices, short messages concentrated in the vicinity of the financial district to send, sending time is short, but the number of sending large. In the evening, lurking a whole day porn Service short messages resurgence.

The X axis represents 24 hours a day, and the Y axis represents the number of days per month. Figure 5, Figure 6, Figure 7, Figure 8 are all under the same scale and use a tool named Echarts to generate [6].

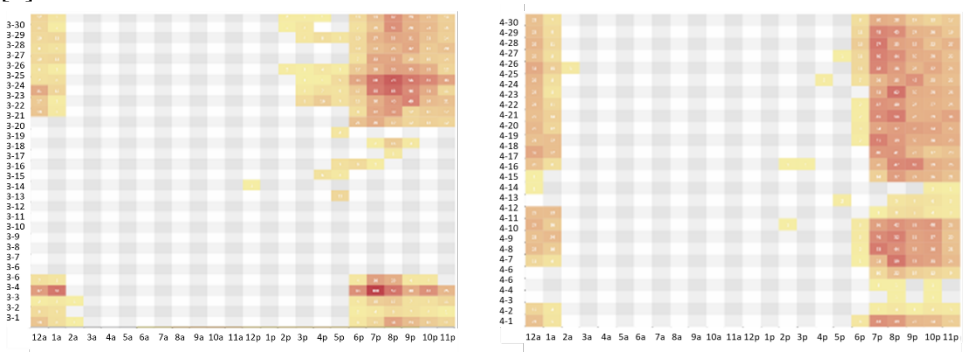

Figure 5:Temporal distribution of pornographic service message

From Figure 5, pornographic information is mainly active in the evening from 18 to 3 a.m.. April 3rd, 4 and 5 are tomb sweeping holidays. During this period, only a small amount of erotic information is active and distributed throughout the day. Another anomalous period was from March 6th to March 20th, when there was almost no information about sexual services. The analysis from the State Council Decree No. 666th "the State Council on amending some administrative regulations" published in March 1st and after the implementation of the national cultural market law enforcement departments to strengthen law enforcement, the Beijing section cracked several pornography cases,

erotic services relevant information is greatly reduced. But after that time, porn messages have a clear trend of rebound.

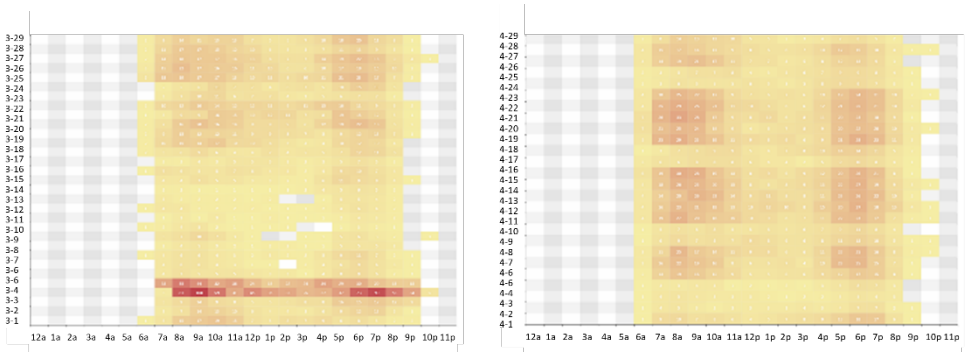

Figure 6:Temporal distribution of fake invoice message

Through the analysis chart of Figure 6, we can clearly find that the invoice information is mainly active from 6 to 21. There is a strong "cyclical", active working days, weekends more convergence. In addition, most dates in March were less active than in April. Abnormal 4-5 March rush invoice message, the paper believes that the overall amount of information in March 4-5 on the two day of the great, various types of message data grew, far more than other days of data for this abnormal surge in the amount of data, when analysing the behaviour is not too much to consider.

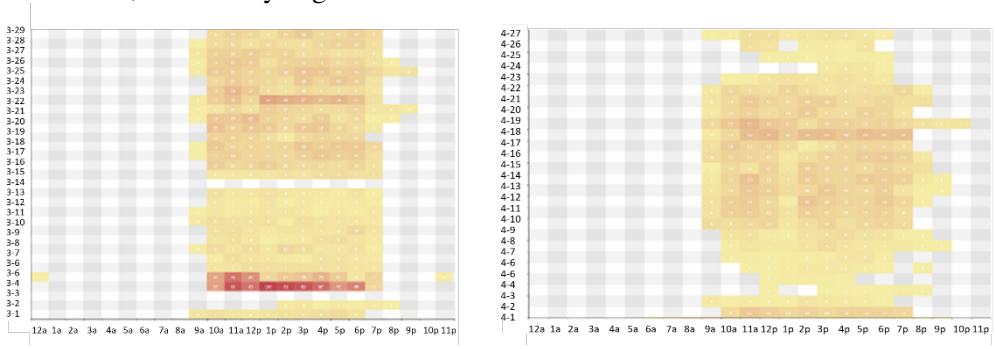

Figure 7:Temporal distribution of posing identity message

From Figure 7, pretending to be an identity class is mostly active during the day and more active. Interestingly, these messages did not appear in the two days of March 3rd and March 15th. March 15th is the consumer protection day, March 3rd is the fifth session of the twelfth CPPCC National Committee, the control is more stringent. On the tomb sweeping holidays, posing as identity class short messages active less.

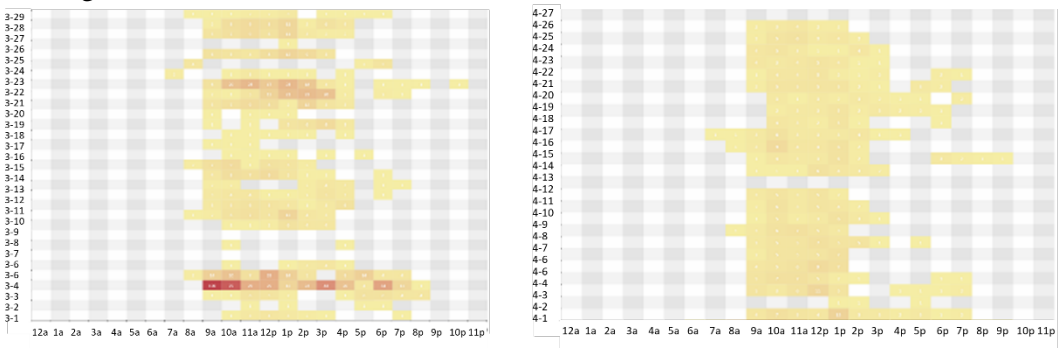

Figure 8:Temporal distribution of real estate advertisement

Through the analysis of Figure 8, it can clearly observe that the real estate advertising is mainly active in 9 to 13, during the tomb sweeping holidays, the active period is relatively small. Based on the above analysis can reveal various types of messages are distribution of their time, the whole tomb sweeping holidays. 3.15 is consumer protection day, holidays and important activities and policy

release date, all kinds of messages that appear less, some types of information are not even all day long, have a great relationship crackdown should be with a police officer, but once the crackdown fell, all kinds of fake base short messages has rebounded, pornographic information is obvious.

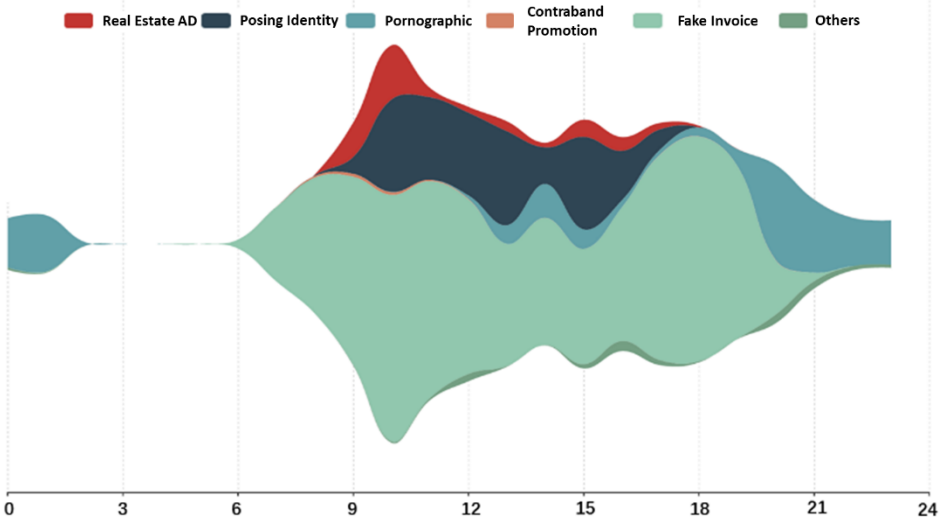

Figure 9:The proportion of different kinds of messages

In order to show the approximate time distribution and total amount of all SMS messages within a day, Figure 9 provides a view. The X axis represents 24 hours a day, and the Y axis represents the total number of each kind of message. From the Figure 9, the overall number of fake invoice message is the most, mainly distributed in the daytime; followed by a large number of pornographic messages are active in the night and noon; and after posing as a real estate advertising message is active in the work period; the remaining amount of other types of messages are scattered in all day long.

To more fully grasp the regularity and influence of the fake base short messages relation, the team used the Baidu LBS API will provide the original data is mapped to the latitude and longitude of the administrative district of Beijing City, according to the distribution of the fake base short messages to the administrative division of Beijing city as a unit of statistics, and the 10 class message division of harm in 5 degree, the degree of hazard level "as the worst impact degree and so on until five, according to the different factors of the content of messages, camouflage number, time distribution and geographical distribution, ten kinds of different text divided into 5 grade level of hazard system, the specific content is shown below.

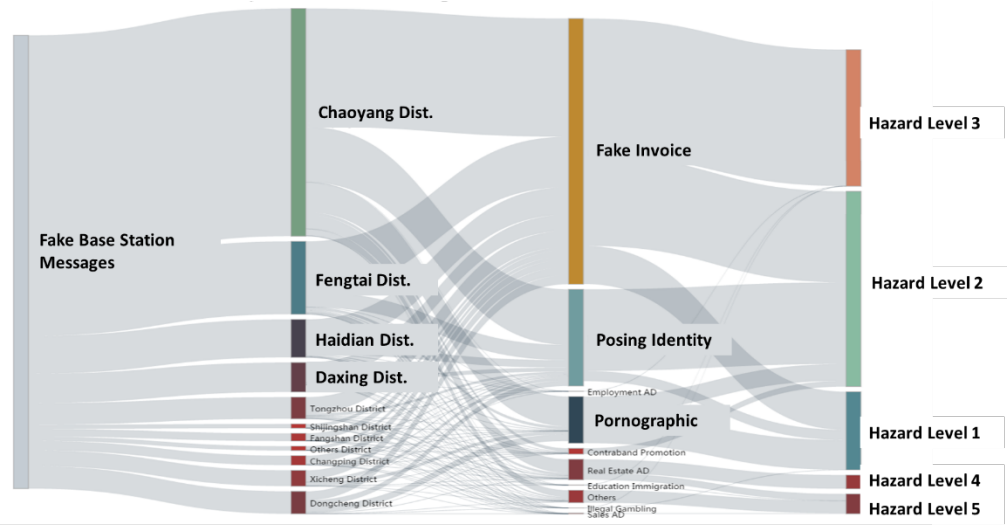

Figure 10: Distribution and hazard levels of pseudo base station messages

It can be obtained from Figure10 that fake base short messages are mainly distributed in the developed economy, office intensive six districts of the city, led by Chaoyang District, followed by Fengtai District, Haidian District, Dongcheng District, Xicheng District and Shijingshan District, but not with the population and land area is. From the Sankey diagram can be seen intuitively fake base message flows, such as short-term erotic services mainly from Chaoyang District, but the Dongcheng District erotic services short messages proportion also cannot be ignored, while the Haidian District erotic services short messages minimum proportion; Shijingshan District has no education immigration short messages.

4. Suggestions

In reality, fake base stations are usually in the form of vehicles, vehicles along the road, on the way to send text messages at fixed points. However, there is no license plate number in the existing fake base station system, which brings challenges to the fake base station traceability and the implementation of the arrest. The problem can be solved effectively by association analysis of fake base station data with traffic camera data. The existing traffic camera can recognize the license plate number accurately, and the moving track of a license plate number can be obtained by retrieving the license plate number in the system database. By using the method of correlation between time and space trajectories proposed in this paper, the license plate number of fake base station vehicles can be accurately locked. For example, as shown in Figure 3, in the 95580-B path, the vehicle from the starting from the East Fourth Ring Road, parking time in the north of the city of Victoria to normal running, turn from the North Sanhuan back through the road, location for the camera coverage area and easily from the vehicle monitoring a license plate number and running track.

According to the previous analysis, this paper can initially determine the highest probability of occurrence of different types of pseudo base stations in different periods, which can serve as a basis for priority attacks. For example, analysis found that after a period of hard strikes, pseudo base station activity will rebound substantially. According to previous data summary, in the size of holidays, important activities, important policies after the implementation of the rebound in the timing, sentinel arrest, you can raise the probability of capturing pseudo base station. Pornographic texts, for example, are expected to rebound after a period of hard strikes. According to the temporal and spatial patterns analysed from historical records, we can provide law enforcement personnel with a number of designated capture areas, and simultaneously with road monitoring linkage to improve the success rate of arresting.

5. Conclusion

Fake base station not only compromises the normal telecommunications order, endangering public safety, disrupt the market order, but also seriously damages the rights and interests of the masses of property, violations of personal privacy of citizens, which lead to serious social harm. According to the "People's Network" statistics, in each year there are nearly 100 billion frauds, gambling, marketing, winning and other text messages has been sent through the fake base station equipment, and the fake base station has become a major public hazard.

In the process of visual analysis, the team found that the time, space and content of the fake base station were highly reproducible. The fake base station movement route, the short message time and the text message content had a fixed behaviour pattern, which means that most of the fake The base station is controlled by a small number of operators. Using the means of trajectory correlation analysis and pattern analysis, the fake base station can be clustered into several operators, thus revealing the operating family behind the fake base station, providing the basis for the control of the fake base station.

6. Acknowledgement

This work is supported by National Key R&D Program of China (No. 2017YFC0803300), the Innovation Funds of CETC (Grant No.16105501) and the Joint Funds of CETC (Grant No. 20166141B08020101).

References

[1] Li X, Cao B, Criminal Regulation on a Case of Fake-base-station[J]. Journal of Guizhou Police Officer Vocational College, 2014.
[2] Zheng Y, Urban Computing: Tackling Urban Challenges Using Big Data[C]// Requirements Engineering Conference. IEEE, 2016:3-3.
[3] Javed M, Siddiqui A T, Transformation of Mobile Communication Network from 1G to 4G and 5G[J]. International Journal of Advanced Research in Computer Science, 2017, 8(3).
[4] Al-Dohuki S, Wu Y, Kamw F, et al. SemanticTraj: A New Approach to Interacting with Massive Taxi Trajectories[J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 23(1):11-20.
[5] https://processing.org/, official website for processing.
[6] http://echarts.baidu.com/, official website for Echarts.