An Interactive Visual System for Generating Striking Pseudo Base Stations Decisions

Fujia Yu
School of Software Engineering, TianJin University, TianJin 300350, China
yufujia@tju.edu.cn

Abstract. The Pseudo base stations cause a great number of social and economic security problems with spreading spam messages. However, it is a serious challenge for security department to recognize the behavior patterns of pseudo base stations, then make use of limited police resources to prevent their activities and even to arrest related criminals. In this paper, a novel visual analytic approach to reveal the spatial-temporal behavior patterns of pseudo base stations is presented. In this approach, it takes advantage of density clustering over high frequency positions, and curve fitting on each cluster. This system supports decision makers to formulate patrol routes on the basis of above approximate trajectory. This approach has been demonstrated effective by means of case studies of authority data and has been evaluated novel and valuable in the view of domain experts.

1. Introduction
Telecom fraud threatens property security of citizens in all time. A wide variety of illegal services are spread here without the trepidation of being discovered. A great number of useless information harasses mobile phone users day and night. One of the top culprits of all the above behaviors is pseudo base stations. The problem of formulating solutions striking pseudo base stations efficiently has plagued decision makers for a long time, owing to the pseudo base stations' strong mobility and high concealment, besides, the lack of police resources. Therefore, it is of great significance to find out pseudo base stations and combat them precisely according to their behavior regularity.

This work applies visual analytics techniques[1] on pseudo base stations behavior patterns mining problem. It at last allows decision makers to propose patrol routes also in visualized approach. A complete procedure which decision makers should follow in one analysis and decision process is combined by three series:

(1) the analysis of spatial-temporal behavior patterns of pseudo base stations.
(2) the observation of the spatial-temporal distribution of spam messages in various themes.
(3) formulating one or several available patrols routes with police resources taken into account.

In addition, the effectiveness of our system is demonstrated in case studies and valuable evaluation is approved by domain experts. After going through our approach, a security department executive commended that, it is a novel idea that is worth to be attempt.

2. Related Work

2.1. Pseudo Base Station Working Principle
"Pseudo base station" generally consists of a host, a laptop computer and an antenna. It can search the mobile phone card information within a certain radius by means of a short message group sender, a
short message transmitter, and the like, and masquerade as a base station of the operator. It pretends to be other people's phone numbers and forces SMS, advertising and other text messages to the user's mobile phone. Pseudo base stations utilize the flaws in the mobile network location registration process, that a terminal could not authenticate whether the base stations belong to its own contracted operator. Pseudo base stations can be placed in a moving motor vehicle, or even a pedestrian's backpack. The strong fluidity of them undoubtedly leads to an increase in difficulty of striking them.

2.2. Pseudo Base Station Detection
Detecting and locating pseudo base stations algorithms and mechanisms[2,3] have been researched in the field of communication. HUANG et al.[4] come up with a method of identifying short messages from pseudo base stations by means of identifying time of cell-site-switch operation that occurred the closest in time to receipt of the message. Shao et al.[5] propose a joint detection method that identifies pseudo base station based on multi-mode intelligent terminals. However, identifying pseudo base station locations precisely with spam messages locations provided by mobile users is basically impossible in real world because of the user locations information privacy. Furthermore, there is no need to obtain the precise location of the pseudo base station but a tendency is enough. Hence, the visualization method on analysing pseudo base stations comes into being.

2.3. Pseudo Base Station Trajectory Visualization
A few of systems have been exploited for visualizing trajectory data. They are widely used in traffic monitoring[6], position selection and so on. Visualization with big data method has also been applied to pseudo base stations monitoring. The "Sky Eye" system exploited by Qihoo 360 displays the possible trajectory of pseudo base stations with location data of spam messages in all days. The "Kirin" system exploited by Tencent can detect spam messages sent from pseudo base stations and get their positions using big data techniques, it alarms instantly when suspect pseudo base stations are detected. Nevertheless, the lack of spam messages and pseudo base stations general analysis and solutions formulation tools remains unsolved. Several visualization systems[7-9] have been exploited for spam messages analysis, but they are almost static with less interactions and without solutions generator, so that cannot meet the exploration and decision needs of users.

3. Data Abstraction and Task Analysis

3.1. Data Abstraction
This work mainly uses two types of data collected in Beijing city. The detailed information is described as follows:

Spam messages data includes contents of spam messages, receiving timestamp, connection with pseudo base station timestamp, approximate longitude latitude locations (to protect users’ privacy) collected by the nearest real base stations, contents of spam messages, md5, phone number used to send message. A total of 3 million sample spam messages was collected in two months.

Road network data comprises more than 11000 geojson format road segments in the city.

3.2. Task Analysis
For the purpose of making a final decision to formulate the police routes plan, decision makers need to follow a series of analysis tasks. The very first task is finding out the entire spatial-temporal behaviour patterns of pseudo base stations. Afterwards, extract themes of spam messages among different business areas and figure out their distribution characteristics. Finally, formulate police routes striking pseudo base stations with comprehensive consideration of patterns and characteristics obtained above.

- Spatial-temporal behaviour patterns extraction: Where do pseudo base stations occur frequently? What time is the high incidence period of pseudo base stations?
Themes extraction and distribution observation: Which kind of spam messages do criminals favour most? What are the relationships among time, regions, and themes of spam messages that pseudo base stations most like?

Pseudo base stations striking routes formulation: Considering the conclusions obtained above, what routes would be the best plan to strike pseudo base stations?

4. System Design and Implementation

4.1. Design Principles
In order to design an insightful visualization, the system shall meet several requirements:
- The basis for working with locations is a map visualization. This visualization needs to show spam messages positions, pseudo base stations predicted moving directions, police routes directions.
- The exploration of spam messages themes distribution needs to be interactive and intuitive. Not least important is a visualization of the distribution result. But just it displays the relationship between themes, time and business areas.
- The final decision of police routes needs to be specific and realistic. Comprehensive consider with police resources gathering positions and pseudo base station activity characteristics. Make directly on the system detailed plans which can be finally carried out.

4.2. System Overview
As shown in Fig. 1, the entire visual analysis system mainly consists of three modules: a pseudo base stations behavior mining module, a spam messages classification and themes extraction module, an integration analysis and decision making module. The behavior mining module employs an interactive visual method to explore spatial-temporal patterns of pseudo base stations activities. The themes extraction module utilizes the natural language processing approach to classify spam messages into several clusters. The decision making module supports to formulate final visual police routes after a comprehensive consideration of all clues acquired among above steps.

4.3. Pseudo Base Stations Routes Prediction
Before analysis, the dataset is cleaned complying with some rules: First, spam messages with wrong receive timestamp and connection timestamp are abandoned, such as the receive timestamp is earlier than connection timestamp, a too long interval between connect time and receive time. Second, meaningless text messages that were split up due to the limit of the length of text are abandoned.

Pseudo base stations are often found at certain high-frequency points. Connect all these high-frequency points in a certain area, then it can reflect the movements of the pseudo base stations in that area. Based on this idea, the approach clusters the high-frequency points in the same business area with DBSCAN algorithm, which is based on density of points. Afterwards, curve-fit roads which represent predicted tracks of the pseudo base stations in each cluster are gotten.
4.4. Themes Extraction
It is hardly impossible to cluster spam messages with traditional machine learning clustering method because of the short length. The approach employs FastText[10] here, a library for efficient learning of word representations and sentence classification. The approach first marks more than 10 thousand messages with correct labels and segment them with JieBa as a training text, then classifies the rest messages with the model after training. The approach can easily and fast extract keywords of short texts. It allows users to get classifications of new dataset in an immediate speed.

4.5. Visualization Design and Interface
Fig. 2 indicates the pseudo base station analysis system which employs a set of visualization techniques that can assist users in formulating patrol routes. This system uses a modular framework to settle multi-structures complexity. Map view is the central module, in addition to that, it adds two other modules, namely, theme relationships module, solution generation module. The system is encoded in coherent visual mode, which provides users an immediate feedback.

**Fig.2** The visual interface. (a) The map view displays behavior patterns of pseudo base stations. (b) The relationship view shows relationships among spatial-temporal-thematic of spam messages. (c) The solution view enables decision makers formulate police routes plan here.

**Map view:** The map view indicates pseudo base stations behavior patterns, final decision of police routes intuitively (Fig. 2(a)). The moving short lines indicate possible routes of pseudo base stations. The color of one short line indicates the most frequent activity range of a pseudo base station. The density of short lines indicates the activity density of pseudo base stations. Moreover, to help users tell the routes and detail process of movements, the map view employs dynamic trajectory which includes a starting point and an end point and a movement route between them for each possible pseudo base station. According to the dynamic routes, users can easily predict next actions of pseudo base stations.

**Theme relationships view:** The theme relationships view shows relationships between themes of spam messages, business areas and peak time periods in all day (Fig. 2(b)). The first show of theme relationships view represents the most frequently occurring themes and time periods of spam messages in selected business area. The second show can be displayed in three different forms in case of changing center perspective to theme, time period or business area, so that it supports frequency comparison in details among other dimensions. With the help of pseudo base stations movements in
map view, users interactively tab between first and second shows to obtain the spatial-temporal-thematic patterns.

**Decision view:** The decision view allows users to make final decisions of deploying police resources in an intuitive way. Decision makers plan routes through the analysis obtained from map view and theme relationships view in previous exploration, set waypoints in the form of business areas chronologically. Start points are set to public security branch office of each administrative districts. After that, the map view is switched to display final decisions in a dynamic visual way.

### 4.6. Interactive Exploration

Our system supports abundant interactions which enable users to explore a specific level of time period and observe a smaller area of region. Besides that, decision makers can get an instant changing view of police routes plan.

**Exploration in a specific level of time period:** A time brush in top bar of map view enables users to select a time range. The map view will get an instant change response when dragging or moving the time brush.

**Observation in several focus regions, time period or theme:** Multi-scenes in theme relationships view display spatial-temporal-thematic complexity in different insights. Meanwhile, the map view would span and zoom to fit a view when changing among these scenes.

**Verification during decision making:** The system provide decision makers a list of business areas ordered by predicted time of pseudo base stations. They can connect these areas as their desire referring to the map view and theme relationships view. After that, the map view switches to a new scene to display final police routes.

### 5. Case Study

To demonstrate this system in a realistic usage scenario, a case study analyzing pseudo base stations of Beijing city is reported. A spam messages dataset extracted in 2017 are used for evaluation of the approach of this work. In the following part, three specific questions are solved. What is the general spatial-temporal pattern of pseudo base stations? What are the relationships among spam message themes, time period and regions? Is there a specific plan for striking pseudo base stations during a time?

#### 5.1. The General Spatial-Temporal Pattern of Pseudo Base Stations

Heatmap is firstly utilized to display the change in quantity of spam messages during 24 hours. As shown in Fig. 3(a), x-axis represents 24 hours, red color indicates high occurrence of spam messages. Blue color indicates low occurrence. Black position indicates that no spam messages occurs. In this way, the occurrence patterns of pseudo base stations during a day are found out.

Taking a first glance at generally spam messages spatial distribution (Fig. 3(b)), A rough impression of pseudo base stations density is presented. After clustering and curve-fitting in a period of time, here a weekday period during 2017/4/10 to 2017/4/14 is chosen for deep exploration. Pseudo base stations in four regions of obvious characteristics are identified (Table 1).

| Top regions | Time periods-Themes |
|-------------|---------------------|
| 1 GuoMao area | 7:00-10:00; 16:00-19:00 invoice agent; 18:00-01:00 porn service |
| 2 DaXing district | 10:00-19:00 estate, invoice agent, bank fraud; |
| 3 Sanyuanqiao area | 10:00-17:00 loan, convert score; |
| 4 WangJing area | 20:00-24:00 porn service; 10:00-17:00 loan; 8:00-20:00 invoice agent; |
5.2. The Relationships among Themes, Time Periods and Regions

With interacting in relationship view (Fig. 3(d)), the top pollution areas and time periods can be concluded. Spam messages in different themes do distinct degree harm to society. The spatial-temporal-thematic distribution of regions in top four harm degree levels are summarized in Table 2.

Table 2. Top victim areas

| Frequent time     | Top regions       | Time periods-Themes                                      |
|-------------------|-------------------|----------------------------------------------------------|
| 10h, 17h-19h      | JianWai street    | invoice agent; convert score; fake certificates; loan; stock |
| 9h-12h, 16h-18h   | WangJing area     | estate; credit card transaction; invoice agent; bank fraud; |
| 9h-11h, 17h-18h   | DaHongmen area    | invoice agent; estate; bank fraud; convert score         |
| 10h-11h, 17h-19h  | BeiYuan area      | invoice agent; porn service; bank fraud                  |

Fig. 3 Spam Messages distribution and PBS occurrence prediction

Fig. 4 Police routes plan in final decision
5.3. A Specific Plan for Striking Pseudo Base Stations in Weekdays

After analyzing such patterns and distribution obtained above, it begins to formulate police routes of final decision. Certain places generated by systems in a time order of pseudo base stations movements are connected. Then the map view, displays the final routes that are just planned (Fig. 4).

6. Conclusion

In this paper, an interactive visual analysis approach of multi-factors driven formulating pseudo base stations striking solutions is proposed. The main feature of the system is that it allows users explore detailed behavior patterns of pseudo base stations interactively in overview and focus insights, and makes available police routes plan intuitively. The applicability of the approach is shown in a case study which demonstrates the practicability and effectiveness in analyzing spam messages with spatial-temporal-thematic features. In future work, the approach will be improved with more considerable features.

Acknowledgments

This work is derived from Chinavis2017 data challenge, the first prize of challenge one. After going through the approach, a security department executive commended that, it is a novel idea that worth to be attempt. This was made public in Chinavis2017 conference by Prof. Zhao of HuNan university.

References

[1] Idreos S, Papaemmanouil O, Chaudhuri S. Overview of data exploration techniques[C]. Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. ACM, 2015: 277-281.

[2] Xumin S, Caixia L. Improved algorithm identifying spam message of pseudo base station based on digital signature[C]. Electronics Information and Emergency Communication (ICEIEC), 2017 7th IEEE International Conference on. IEEE, 2017: 176-179.

[3] Wang Z, Sheen B, Wong M. Response mechanism for detection of a faked base station: U.S. Patent 9,872,237[P]. 2018-1-16.

[4] HUANG Q, Hengbin C U I, Laijun Y A N. Method and device for identifying short messages from pseudo base stations: U.S. Patent Application 10/045,166[P]. 2018-8-7.

[5] Shao J, Zhu D, Jin H, et al. A Joint Detection Method for Identifying Pseudo Base Station Based on Abnormal Access Parameters[J]. DEStech Transactions on Engineering and Technology Research, 2016 (iceta).

[6] Chen W, Guo F, Wang F Y. A survey of traffic data visualization[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(6): 2970-2984.

[7] Jia Y, Lyu B, Li Y. Temporal and Spatial Analysis of Fake Base Stations[C]. Journal of Physics: Conference Series. IOP Publishing, 2018, 1060(1): 012019.

[8] Li, Y., Yang, H., Hao, A., et al.: Detecting and tracking pseudo base stations in GSM signal hijacking and frauds: a visualized approach. Inf. Secur. Comput. Fraud 5(1), 1–8 (2017)

[9] Zhang H, Tang X, Li C, et al. PBSVis: A Visual System for Studying Behavior Patterns of Pseudo Base Stations[C]. International Conference of Pioneering Computer Scientists, Engineers and Educators. Springer, Singapore, 2018: 599-610.

[10] Joulin A, Grave E, Bojanowski P, et al. Bag of tricks for efficient text classification[J]. arXiv preprint arXiv:1607.01759, 2016.