LBSNShield: Malicious Account Detection in Location-Based Social Networks

Yuan Xuan[1,2,3], **Yang Chen**[1,2,3 a], **Huiying Li**[1,2], **Pan Hui**[4], **Lei Shi**[3]

[1] School of Computer Science, Fudan University, China

[2] Engineering Research Center of Cyber Security Auditing and Monitoring, Ministry of Education, China

[3] The State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, China

[4] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong

{yxuan12, chenyang, huiyingli13}@fudan.edu.cn
panhui@cse.ust.hk, shil@ios.ac.cn

Abstract

Given the popularity of GPS-enabled smart devices, location-based social networks (LBSNs) have attracted numerous users around the world. The openness of LBSN platforms has also made themselves the targets of malicious attackers. In LBSNs, attackers can register a number of fake identities and let them post spam reviews or fake check-ins. Therefore, discovering and blocking the malicious accounts are vital for the experience of legitimate users. In this paper, we investigated how to accurately detect malicious accounts in LBSNs. We collected rich user data from a popular LBSN in China, so-called Dianping. We then built a crowdsourcing based annotation platform to mark legitimate and malicious accounts. By examining the annotated data set, we selected a number of key features to distinguish between these two types of accounts. Based on these features, we built LBSNShield, a machine learning-based malicious account detection system. According to our extensive evaluation, our system can achieve an F1-score of 0.89.

Author Keywords

Crowdsourcing; Machine Learning; Location-based Social Networks; Malicious Account Detection

ACM Classification Keywords

J.4 [Social And Behavioral Sciences]: Sociology

[a] Corresponding Author

CSCW '16 Companion, February 27 - March 02, 2016, San Francisco, CA, USA

ACM 978-1-4503-3950-6/16/02.
http://dx.doi.org/10.1145/2818052.2869094

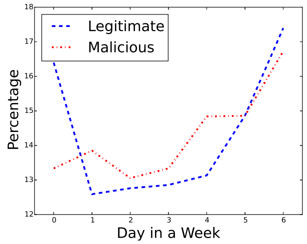

Figure 1: Check-in Pattern (Weekly)

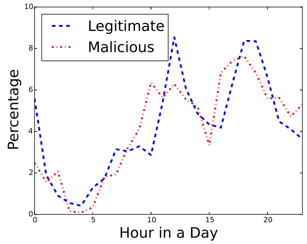

Figure 2: Check-in Pattern (Daily)

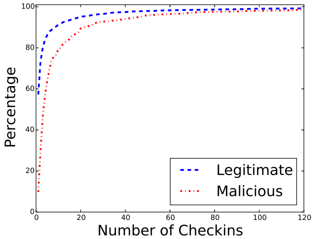

Figure 3: Total Number of Check-ins (Cumulative Distribution Function)

1 Background and Related Work

Nowadays smart devices, including phones, tablets, and watches, have been widely used all over the world. Most of these devices are GPS-enabled, allowing people to track their latest locations. Meanwhile, online social networks have become a key component in people's daily-life. Location-Based Social Networks (LBSNs), such as Foursquare and Dianping [1], have attracted hundreds of millions of users around the world. However, due to the open nature of such platforms, attackers would also join these platforms, and badly hurt the experience of legitimate users. For example, they can register a number of malicious accounts to post misleading reviews, or conduct fake check-ins [5]. Therefore, it is very important for a LBSN platform to quickly identify these accounts and block them.

There are a number of literatures about malicious account detection in online social networks (OSNs), e.g., social graph-based approaches [1] and machine learning-based approaches [4]. However, most of the those existing solutions do not take the location information into account. One known solution for detecting malicious accounts in LBSNs is done by Zhang et al. [5]. This work reveals several key differences between legitimate users and malicious accounts. However, the data set they used was quite small, including only 2000 users who have visited a specific hotpot restaurant. Therefore, their results cannot ensure whether their proposed solution works for the entire user population.

2 Our Approach

To enhance the user experience in LBSNs, we aim to eliminate the negative impact of malicious accounts. We conducted a data-driven study by using the real data of more than one million LBSN users. Based on the collected data, we conduct a series of studies, and have made the following three key contributions: (1) We build a crowdsourcing based annotation system to utilize the "wisdom of the crowd" to check whether a user is a legitimate one or a malicious one. (2) Given the nature of LBSNs, we select a number of features to distinguish between legitimate users and malicious accounts. We can see malicious accounts act significantly differently from legitimate users. (3) We build a machine learning-based framework for malicious account detection, and we have tested a series of classic machine learning algorithms, including Support Vector Machine (SVM), Decision Tree, Random Forests, and several Bayesian methods. According to our evaluation, we can see that the framework can identify malicious users accurately.

2.1 Data collection

The data collection was done from Feb. 15, 2015 to May 27, 2015. We have developed our own crawler for Dianping[2], which can extract the key information of any given Dianping user. Similar to other OSN platforms, Dianping applies a very strict per-IP rate limit policy. To speed up the crawling, we use a distributed data crawling framework, so-called crowd crawling [2]. We use 40 virtual instances from the East Asia data center of the Microsoft Azure cloud computing platform, and each of these virtual instances has a unique IP address. All the crawlers work together to fetch the data set, and our crawling efficiency is much higher than running crawlers via the same IP address. For each user, we have collected her ID, #checkins, #follows, #fans, #reviews, last login date, registration date, birthday, city, gender, and check-in timeline.

2.2 Crowdsourcing based annotation

The data set contains 1,019,195 users' profiles and user-generated contents (UGCs). In particular, we only examine

[1] http://www.dianping.com/aboutus

[2] https://github.com/chenyang03/Dianping_Crawler

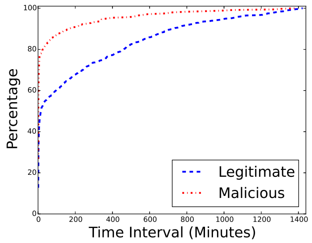

Figure 4: Time Interval
(Cumulative Distribution Function)

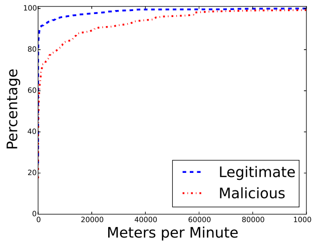

Figure 5: Travelling Speed
(Cumulative Distribution Function)

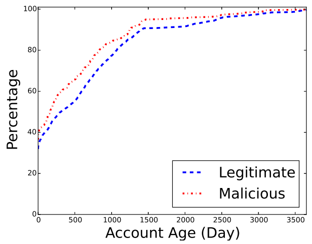

Figure 6: Account Age
(Cumulative Distribution Function)

the users who have at least posted a review or conducted a check-in in Dianping. Unfortunately, the data set itself cannot tell us which accounts are malicious. As we need this information for our further study, we build a crowdsourcing based annotation system to check against each user. We recruit ten volunteers to do the annotation, and all of them are graduate students from the School of Computer Science at Fudan University. To limit the annotation overhead, we randomly select the data of 4,395 users for the annotation system. For each volunteer, once he or she accesses our annotation system, a randomly chosen user's profile and timeline will be displayed, and the volunteer can select from "malicious" and "legitimate" to tell the system his or her opinion. We ensure each user would receive at least 3 independent votes, and we make the conclusion by simple majority. If there is no difference between the numbers of "malicious" and "legitimate" votes, we will examine the user by ourselves. Finally, we find that 398 of the studied accounts are malicious, and 3,997 are legitimate.

2.3 Key Features

As shown in Fig. 1 - 6, we compare malicious users with legitimate users by looking at several key features.

- We show the percentage of check-ins on each day of a week in Fig. 1, and we can see that malicious accounts are more active on weekdays, while legitimate accounts are more active on weekends.
- We display the percentage of check-ins on each hour of a day in Fig. 2, and we find that malicious accounts are more active in work hours, while legitimate accounts are more active during lunch/dinner time.
- Malicious accounts conduct more check-ins than legitimate users (Fig. 3). They have a shorter checkin interval (Fig. 4), a higher movement speed (Fig. 5), and a smaller account age (Fig. 6).

Algorithm	Settings
Decision Tree	Confidence Factor C=0.25, M=2
Random Forests	100 trees, 3 random features each
SVM	Radial Basis Function, C=32768.0, g=2.0

Table 1: Settings for detection algorithms.

In short, all these features can be used to distinguish between malicious users and legitimate users.

2.4 Machine Learning-Based Malicious Account Detection

Based on the aforementioned key features, we introduce a machine learning-based framework, known as LBSNShield. Similar to [3], a number of classic machine learning algorithms are used, including Bayes Net, Bayes Logistic Regression, Naive Bayes, Decision Tree, Random Forests, and Support Vector Machine (SVM).

To get a balanced experimental dataset, we pick all 398 malicious accounts, and randomly select 398 legitimate users from our annotated data set. Briefly, we use the grid search method to find the best parameters and use them to train the model. We use cross-validation to check out whether the selected features and algorithms are effective. Note that features are normalized when using SVM. Key settings in our study are shown in Table 1.

2.5 Evaluation

To examine the correctness of this system, we apply the following three key metrics, i.e., precision, recall, and F1-score, to evaluate the accuracy of classification. Precision is the fraction of detected malicious users that are real. Recall is the fraction of real malicious users that are detected. F1 score is harmonic mean of precision and recall. By do-

Algorithm	Precision	Recall	F1-score
Decision Tree	0.83	0.80	0.82
Random Forests	0.81	0.81	0.81
SVM	0.86	0.93	0.89

Table 2: Precision, Recall, and F1-score of three tested methods.

Exp. Round	Precision	Recall	F1-score
1	0.9247	0.9089	0.9167
2	0.9224	0.9224	0.9224
3	0.9210	0.9154	0.9182
4	0.9229	0.8919	0.9071
5	0.9196	0.9327	0.9261

Table 3: Result of five parallel experiments using SVM.

ing 10-fold cross-validation, we observe that three Bayesian methods perform poorly on our experimental dataset, so we only show the performance of SVM, Decision Tree, and Random Forests in Table 2. In our experiment, SVM outperforms all other algorithms and is able to achieve an F1-score of 0.89. To further check the stability of our method, we simulate the real-world scenario by training the SVM model using a smaller training set and test it on the dataset containing all 4,395 labeled users. We randomly select 300 legitimate users and 300 malicious users to train the model, and use the trained model to predict malicious accounts from the dataset of 4,395 users. We conduct the same experiment independently for 5 times, and the result is shown in Table 3. The result clearly indicates that our method is stable and accurate.

3 Conclusion and Future Work

In this paper, we investigate the attacks in LBSNs by using real data set collected from a representative LBSN in China. Our study shows the difference between legitimate users and malicious users from different angles. Moreover, we propose a framework of a defense system known as LBSNShield. LBSNShield can discover malicious users in Dianping with a high accuracy.

For future work, on one hand, we will further investigate the key features for classification. We hope to find more useful features in LBSNs, and improve the accuracy of malicious account detection. On the other hand, we plan to expand the generality of LBSNShield. We expect it can work well not only for Dianping, but also for other LBSN platforms.

References

[1] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and et al. Aiding the Detection of Fake Accounts in Large Scale Social Online Services. In *Proc. of NSDI, 2012*.

[2] Cong Ding, Yang Chen, and Xiaoming Fu. Crowd Crawling: Towards Collaborative Data Collection for Large-scale Online Social Networks. In *Proc. of ACM COSN, 2013*.

[3] Gang Wang, Tianyi Wang, Haitao Zheng, and Ben Y. Zhao. Man vs. Machine: Practical Adversarial Detection of Malicious Crowdsourcing Workers. In *Proc. of USENIX Security, 2014*.

[4] Zhi Yang, Christo Wilson, and et al. 2014. Uncovering Social Network Sybils in the Wild. *ACM Trans. Knowl. Discov. Data* 8, 1 (2014), 2:1–2:29.

[5] Xiaokuan Zhang, Haizhong Zheng, Xiaolong Li, Suguo Du, and Haojin Zhu. You are Where You Have Been: Sybil Detection via Geo-location Analysis in OSNs. In *Proc. of IEEE GLOBECOM, 2014*.