IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

REPRESENTING FINE GRAINED CO OCCURRENCES FOR BEHAVIOR BASED FRAUD DETECTION

T. KALAISELVI[1], M. KARTHIK[2]
1 ASSOCIATE PROFESSOR 2 PG SCHOLAR
COMPUTER SCIENCE AND ENGINEERING
ERODE SENGUNTHAR ENGINEERING COLLEGE, TAMILNADU, INDIA

ABSTRACT

The vigorous development of e-commerce breeds cybercrime. Online payment fraud detection, a challenge faced by online service, plays an important role in rapidly evolving e-commerce. Behavior-based methods are recognized as a promising method for online payment fraud detection. However, it is a big challenge to build high-resolution behavioral models by using low-quality behavioral data. In this work, we mainly address this problem from data enhancement for behavioral modeling. We extract fine-grained co-occurrence relationships of transactional attributes by using a knowledge graph. Furthermore, we adopt the heterogeneous network embedding to learn and improve representing comprehensive relationships. Particularly, we explore customized network embedding schemes for different types of behavioral models, such as the population-level models, individual-level models, and generalized-agent based models. The performance gain of our method is validated by the experiments over the real dataset from a commercial bank. It can help representative behavioral models improve significantly the performance of online banking payment fraud detection. To the best of our knowledge, this is the first work to realize data enhancement for diversified behavior models by implementing network embedding algorithms on attribute-level co-occurrence relationships.

Keywords: OSN website, Embedding schemes and Fraud detection.

1.INTRODUCTION

Online payment services have penetrated into people's lives. The increased convenience, though, comes with inherent security risks [1]. The cybercrime involving online payment services often has the characteristics of diversification, specialization, industrialization, concealment, scenario, and cross-region, which makes the security prevention and control of online payment extremely challenging [2]. There is an urgent need for realizing effective and comprehensive online payment fraud detection. The behavior-based method is recognized as an effective paradigm for online payment fraud detection [3]. Gen rally, its advantages can be summarized as follows: Firstly, behavior based methods adopt the non intrusion detection scheme to guarantee the user experience without user operation in the implementation process. Secondly, it changes the fraud detection pattern from one-time to continuous and can verify each transaction. Thirdly, even if the fraudster imitates the daily operation habits of the victim, the fraudster must deviate from the user behavior to gain the benefit of the victim. The deviation can be detected by behavior based methods. Finally, this behavior-based method can be used cooperatively as a second security line, rather than replacing with other types of detection methods.

2. LITERATURE SURVEY

REPRESENTING FINE-GRAINED CO-OCCURRENCES FOR BEHAVIOR-BASED FRAUD DETECTION IN ONLINE PAYMENT SERVICES

The vigorous development of e-commerce breeds cybercrime. Online payment fraud detection, a challenge faced by online service, plays an important role in rapidly evolving e-commerce. Behavior-based methods are recognized as a promising method for online payment fraud detection. However, it is a big challenge to build high-resolution behavioral models by using low-quality behavioral data. In this work, we mainly address this problem from data enhancement for behavioral modeling. We extract fine-grained co-occurrence relationships of transactional attributes by using a knowledge graph. Furthermore, we adopt the heterogeneous network embedding to learn and improve representing comprehensive relationships. Particularly, we explore customized network embedding schemes for different types of behavioral models, such as the population-level models, individual-level models, and generalized-agent-based models. The performance gain of our method is validated by the experiments over the real dataset from a commercial bank. It can help representative behavioral models improve significantly the performance of online banking payment fraud detection. To the best of our knowledge, this is the first work to realize data enhancement for diversified behavior models by implementing network embedding algorithms on attribute-level co-occurrence relationships.

PROFILING ONLINE SOCIAL BEHAVIORS FOR COMPROMISED ACCOUNT DETECTION

Account compromization is a serious threat to users of online social networks (OSNs). While relentless spammers exploit the established trust relationships between account owners and their friends to efficiently spread malicious spam, timely detection of compromised accounts is quite challenging due to the well established trust relationship between the service providers, account owners, and their friends. In this paper, we study the social behaviors of OSN users, i.e., their usage of OSN services, and the application of which in detecting the compromised accounts. In particular, we propose a set of social behavioral features that can effectively characterize the user social activities on OSNs. We validate the efficacy of these behavioral fea- tures by collecting and analyzing real user clickstreams to an OSN website. Based on our measurement study, we devise individual user's social behavioral profile by combining its respective behavioral feature metrics. A social behavioral profile accurately reflects a user's OSN activity patterns. While an authentic owner conforms to its account's social behavioral profile involuntarily, it is hard and costly for impostors to feign. We evaluate the capability of the social behavioral profiles in distinguishing different OSN users, and our experimental results show the social behavioral profiles can accurately differentiate individual OSN users and detect compromised accounts.

PRIVACY AGAINST STATISTICAL MATCHING: INTER-USER CORRELATION

Modern applications significantly enhance user experience by adapting to each user' s individual condition and/or preferences. While this adaptation can greatly improve utility or be essential for the application to work (e.g., for ride-sharing applications), the exposure of user data to the application presents a significant privacy threat to the users, even when the traces are anonymized, since the statistical matching of an anonymized trace to prior user behavior can identify a user and their habits. Because of the current and growing algorithmic and computational capabilities of adversaries, provable privacy guarantees as a function of the degree of anonymization and obfuscation of the traces are necessary. Our previous work has established the requirements on anonymization and obfuscation in the case that data traces are independent between users. However, the data traces of different users will be dependent in many applications, and an adversary can potentially exploit such. In this paper, we consider the impact of correlation between user traces on their privacy. First, we demonstrate that the adversary can readily identify the association graph, revealing which user data traces are correlated. Next, we demonstrate that the adversary can use this association graph to break user privacy with significantly shorter traces than in the case when traces are independent between users, and that independent obfuscation of the data traces is often insufficient to remedy such. Finally, we discuss how the users can employ dependency in their obfuscation to improve their privacy.

TOWARDS DETECTING COMPROMISED ACCOUNTS ON SOCIAL NETWORKS

Compromising social network accounts has become a profitable course of action for cybercriminals. By hijacking control of a popular media or business account, attackers can distribute their malicious messages or disseminate fake information to a large user base. The impacts of these incidents range from a tarnished reputation to multi-billion dollar monetary losses on financial markets. In our previous work, we demonstrated how we can detect large-scale compromises (i.e., so-called campaigns) of regular online social network users. In this work, we show how we can use similar techniques to identify compromises of individual high-profile accounts. High-profile accounts frequently have one characteristic that makes this detection reliable-they show consistent behavior over time. We show that our system, were it deployed, would have been able to detect and prevent three real-world attacks against popular companies and news agencies. Furthermore, our system, in contrast to popular media, would not have fallen for a staged compromise instigated by a US restaurant chain for publicity reasons.

AUTHENTICATION OF SMARTPHONE USERS USING BEHAVIORAL BIOMETRICS

Smartphones and tablets have become ubiquitous in our daily lives. Smartphones, in particular, have become more than personal assistants. These devices have provided new avenues for consumers to play, work, and socialize whenever and wherever they want. Smartphones are small in size, so they are easy to handle and to stow and carry in users' pockets or purses. However, mobile devices are also susceptible to various problems. One of the greatest concerns is the possibility of breach in security and privacy if the device is seized by an outside party. It is possible that threats can come from friends as well as strangers. Due to the size of smart devices, they can be easily lost and may expose details of users' private lives. In addition, this might enable pervasive observation or imitation of one's movements and activities, such as sending messages to contacts, accessing private communication, shopping with a credit card, and relaying information about where one has been. This paper highlights the potential risks that occur when smartphones are stolen or seized, discusses the concept of continuous authentication, and analyzes current approaches and mechanisms of behavioral biometrics with respect to methodology, associated datasets and evaluation approaches.

WRONGDOING MONITOR: A GRAPH-BASED BEHAVIORAL ANOMALY DETECTION IN CYBER SECURITY

The so-called behavioral anomaly detection (BAD) is expected to solve effectively a variety of security issues by detecting the deviances from normal behavioral patterns of protected agents. We propose a new graph-based behavioral modeling paradigm for BAD problem, named behavioral identification graph (BIG), which has distinct advantages over existing methods by mining deeply the property-level (as an enhancement to the event-level) associations in behavioral data. Under BIG, the behavioral properties and their co-occurrence associations in behavioral data are modeled as the entities and relationships of graph, respectively; furthermore, behavioral properties and events are both vectorized by a devised event-property composite model, and the behavioral patterns of agents are finally represented as a multidimensional spatial distribution of behavioral properties. Consequently, for a behavior, the intensity of its behavioral anomaly can be transformed into the spatial decentrality of its behavioral agent and properties which contain both fine-grained information between behavioral properties and coarse-grained information between behavioral events. To the best of our knowledge, this is the first work to improve behavioral modeling for anomaly detection by integrating inter (event-level) and intra (property-level) associations of behaviors into a unified graph and space. Our method is validated by four representative security issues, i.e., fraud detection in online payment services (by transaction behaviors), intrusion detection in network communication services (by traffic behaviors), insider threat detection in organizational information systems (by system behaviors), and compromise detection in social networking services (by trajectory behaviors).

3. EXISTING SYSTEM

Vedran et al. [19] explored the complex interaction between social and geospatial behavior and demonstrated that social behavior could be predicted with high precision. Yin et al. [4] proposed a probabilistic generative model combining use spatiotemporal data and semantic information to predict user behavior. Naini et al. [7] studied the task of identifying the users by matching the histograms of their data in the anonymous dataset with the histograms from the original dataset. Egele et al. [8] proposed a behavior-based method to identify compromises of high-profile accounts. Ruan et al. [3] conducted a study on online user behavior by collecting and analyzing user clickstreams of a well known OSN.

Rzecki et al. [20] designed a data acquisition system to analyze the execution of single-finger gestures on a mobile device screen and indicated the best classification
method for person recognition based on proposed surveys. Alzubaidi et al. [9] investigated the representative methods for user authentication on smartphone devices in smartphone authentication including seven types of behavioral biometrics, which are hand waving, gait, touchscreen, keystroke, voice, signature and general profiling.

Lee and Kim [21] proposed a suspicious URL detection system to recognize user anomalous behaviors on Twitter. Cao et al. [11] designed and implemented a malicious account detection system for detecting both fake and compromised real user accounts. Zhou et al. [12] proposed an FRUI algorithm to match users among multiple OSNs. Stringhini et al. [22] designed a system named EVILCOHORT, which can detect malicious accounts on any online service with the mapping between an online account and an IP address. Meng et al. [23] presented a static sentence-level attention model for text-based speaker change detection by formulating it as a matching problem of utterances before and after a certain decision point. Rawat et al. [24] proposed three methodologies to cope up with suspicious and anomalous activities, such as continuous creation of fake user accounts, hacking of accounts and other illegitimate acts in social networks.

VanDam et al. [25] focused on studying compromised accounts in Twitter to understand who were hackers, what type of content did hackers tweet, and what features could help distinguish between compromised tweets and normal tweets. They also showed that extra meta-information could help improve the detection of compromised accounts.

Zhao et al. [26] proposed a semi-supervised network embedding model by adopting graph convolutional network that is capable of capturing both local and global structure of protein-protein interactions network even there is no any information associated with each vertex. Li et al. [27] incorporated word semantic relations in the latent topic learning by the word embedding method to solve that the Dirichlet Multinomial Mixture model does not have access to background knowledge when modeling short texts.

Baqueri et al. [28] presented a framework to model residents travel and activities outside the study area as part of the complete activity-travel schedule by introducing the external travel to address the distorted travel patterns. Chen et al. [29] proposed a collaborative and adversarial network (CAN), which explicitly models the common features between two sentences for enhancing sentence similarity modeling. Catolino et al. [30] devised and evaluated the performance of a new change prediction model that further exploit developerrelated factors (e.g., number of developers working on a class) as predictors of change-proneness of classes. Liu et al. [31] proposed a novel method for disaggregating the coarse-scale values of the group-level features in the nested data to overcome the limitation in terms of their predictive performance, especially the difficulty in identifying potential cross-scale interactions between the local and group-level features when applied to datasets with limited training examples.

4. PROPOSED SYSTEM

The system proposes a novel effective data enhancement scheme for behavioral modeling by representing and mining more fine-grained attribute-level co-occurrences. We adopt the heterogeneous relation networks to represent the attribute-level co-occurrences, and extract those relationships by heterogeneous network embedding algorithms in depth.

The system devises a unified interface between network embedding algorithms and behavioral models by customizing the preserved relationship networks according to the classification of behavioral models.

The system implements the proposed methods on a real-world online banking payment service scenario. It is validated that our methods significantly outperform the state of-the-art classifiers in terms of a set of representative metrics in online fraud detection.

5. SYSTEM REQUIREMENTS

H/W System Configuration

Processor	-	Intel i3 processor
RAM	-	4 GB (min)
Hard Disk	-	500 GB
Key Board	-	Standard Windows Keyboard
Mouse	-	Two or Three Button Mouse
Monitor	-	SVGA

SOFTWARE REQUIREMENTS

Operating System	-	Windows 10
Coding Language	-	Java/J2EE (JSP, Servlet)
Front End	-	J2EE
Back End	-	MySQL

6. MODULES

USER MODULE

In this module the user will search the Server that are available and buy the service that the user is interested the user will redirect to the online transaction page after selecting the service then if the user has the transactional id and password they can login and pay the amount to the service provider account by online transfer method.

SERVICE MODULE

In this module the Service Provider will add the new services or products to view for the user. Then the user will search the services that are available and select the required service and then move to the transaction page to pay the money for activation of the service

BANK TRANSACTION MODULE

In transaction module the user will get redirect to this page after the selection of the service that they want for online payment. In this transaction module first the user will apply for the loan for buying the service and the bank admin should approve the loan after successfully verifying the details of the user once the bank approved the loan the amount will automatically credit to the user account in his account number. Then the user can directly transfer the amount to the service provider account.

ADMIN MODULE

In admin module the admin will authorize the users and add the services that are launched newly to view in the user side. Admin can check directly his bank account details and purchase transaction behaviors after login. In admin module we will have all the user details and their transactions behavior. customized co-occurrence relation networks, and introduce the technique of heterogeneous network embedding to represent online transaction data for different types of behavioral models, e.g., the individual-level and population-level models. The methods are validated by the implementation on a real-world dataset. They outperform the state of- the-art classifiers with lightweight feature engineering methods. Therefore, our methods can also serve as a feasible paradigm of automatic feature engineering. There are some interesting issues left to study: (1) An interesting future work is to extend the data enhancement scheme into other types of behavioral models, e.g.,

the group-level models and generalized-agent-based models, except the population-level and individual-level models studied in this work. (2) It would be interesting to investigate the dedicated enhancement schemes for more advanced individual-level models, since the adopted naive individual-level model does not fully capture the advantages of the proposed data representation scheme based on the techniques of heterogeneous network embedding. (3) It is anticipated to demonstrate the generality of the proposed method by applying it to different real-life application scenarios.

7. REFERENCES

[1] B. Cao, M. Mao, S. Viidu, and P. S. Yu, "Hitfraud: A broad learning approach for collective fraud detection in heterogeneous information networks," in Proc. IEEE ICDM 2017, New Orleans, LA, USA, November 18-21, 2017, pp. 769–774.

[2] M. A. Ali, B. Arief, M. Emms, and A. P. A. van Moorsel, "Does the online card payment landscape unwittingly facilitate fraud?" IEEE Security & Privacy, vol. 15, no. 2, pp. 78–86, 2017.

[3] X. Ruan, Z. Wu, H. Wang, and S. Jajodia, "Profiling online social behaviors for compromised account detection," IEEE Trans. Information Forensics and Security, vol. 11, no. 1, pp. 176–187, 2016.

[4] H. Yin, Z. Hu, X. Zhou, H. Wang, K. Zheng, N. Q. V. Hung, and S. W. Sadiq, "Discovering interpretable geo-social communities for user behavior prediction," in Proc. IEEE ICDE 2016, Helsinki, Finland, May 16-20, 2016, pp. 942–953.

[5] Y.-A. De Montjoye, L. Radaelli, V. K. Singh et al., "Unique in the shopping mall: On the reidentifiability of credit card metadata," Science, vol. 347, no. 6221, pp. 536–539, 2015.

[6] A. Khodadadi, S. A. Hosseini, E. Tavakoli, and H. R. Rabiee, "Continuous-time user modeling in presence of badges: A probabilistic approach," ACM Trans. Knowledge Discovery from Data, vol. 12, no. 3, pp. 37:1–37:30, 2018.

[7] F. M. Naini, J. Unnikrishnan, P. Thiran, and M. Vetterli, "Where you are being who you are: User identification by matching statistics," IEEE Trans. Information Forensics and Security, vol. 11, no. 2, pp. 358–372, 2016.

[8] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "Towards detecting compromised accounts on social networks," IEEE Trans. Dependable and Secure Computing, vol. 14, no. 4, pp. 447–460, 2017.

[9] A. Alzubaidi and J. Kalita, "Authentication of smartphone users using behavioral biometrics," IEEE Communications Surveys and Tutorials, vol. 18, no. 3, pp. 1998–2026, 2016.

[10] H. Mazzawi, G. Dalaly, D. Rozenblatz, L. Ein-Dor, M. Ninio, and O. Lavi, "Anomaly detection in large databases using behavioral patterning," in Proc. IEEE ICDE 2017, pp. 1140–1149.