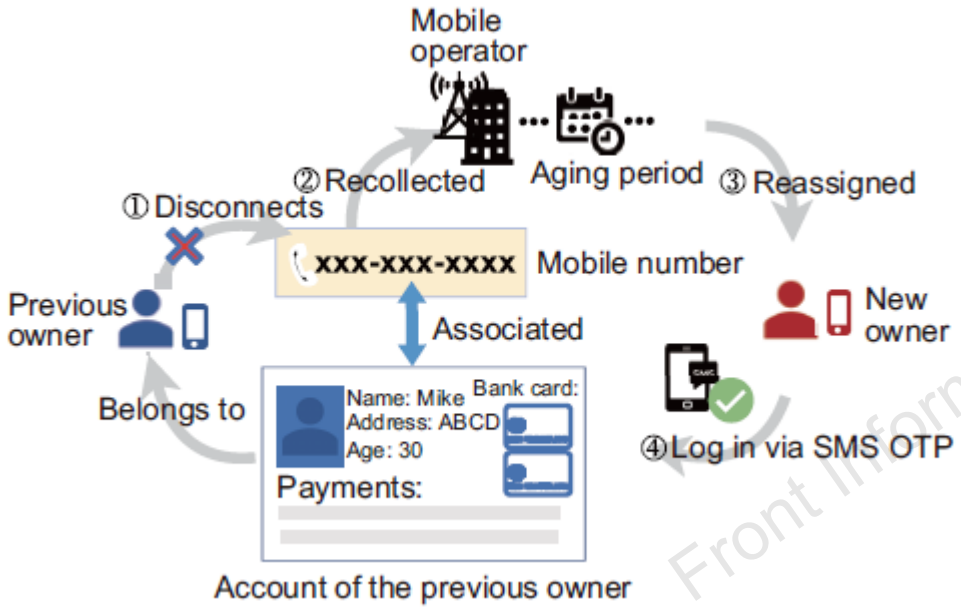
Min GAO, Shutong CHEN, Yangbo GAO, Zhenhua ZHANG, Yu CHEN, Yupeng Li, Qiongzan YE, Xin WANG, Yang CHEN. 2024. Detecting compromised accounts caused by phone number recycling on e-commerce platforms: taking Meituan as an example. *Frontiers of Information Technology & Electronic Engineering*, 25(8):1077-1095. https://doi.org/10.1631/FITEE.2300291

Detecting compromised accounts caused by phone number recycling on e-commerce platforms: taking Meituan as an example

Key words: Phone number recycling; Neural networks; E-commerce; Compromised account detection

Corresponding author: Yang CHEN
Email: chenyang@fudan.edu.cn
ORCID: https://orcid.org/0000-0003-4749-3060

Motivation



The process of phone number recycling

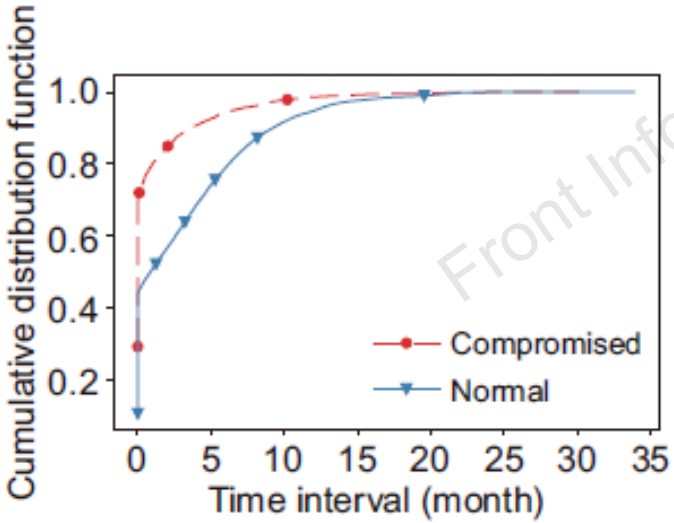
- **Security Risks of Phone Number Recycling (PNR):** PNR can lead to unauthorized access to accounts, increasing fraud risk and causing monetary loss as well as reputation damage.
- **Current Solutions:** Existing methods are both costly and ineffective. ☹
- **Need for an Effective Method:** There is an urgent need for cost-effective solutions that use only application data to identify compromised accounts.

Data-Driven Study

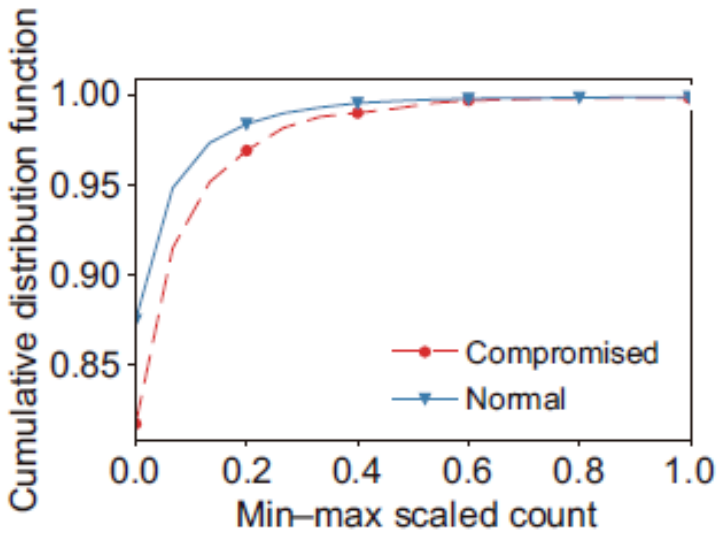
- We have conducted a data-driven study on the compromised accounts and the normal ones from several behavioral patterns.

- We have discovered several distinctive features that would distinguish compromised accounts from normal ones.

➢ Analysis on card binding related features



Average time interval of bank card bindings



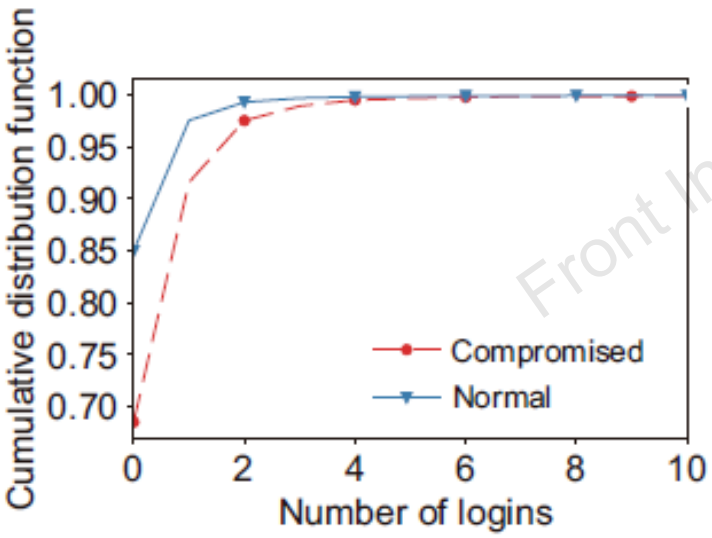
Number of failed bank card binding attempts

Data-Driven Study

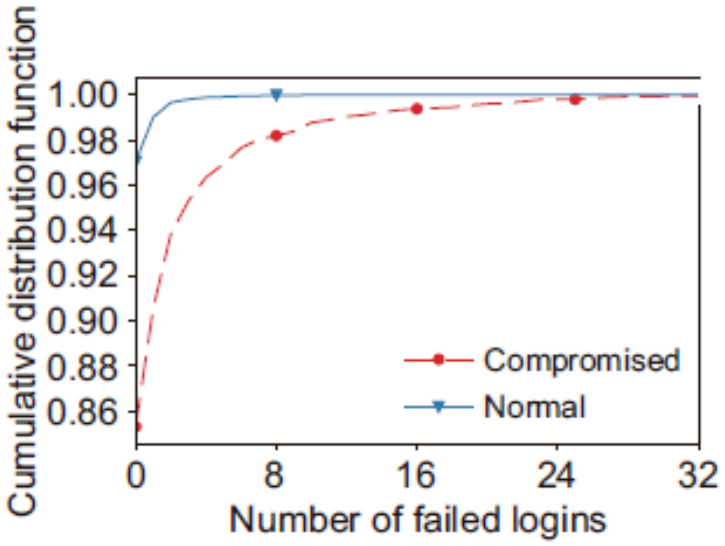
- We have conducted a data-driven study on the compromised accounts and the normal ones from several behavioral patterns.

- We have discovered several distinctive features that would distinguish compromised accounts from normal ones.

➢ Analysis on login related features



Number of logins on a new device by SMS OTP



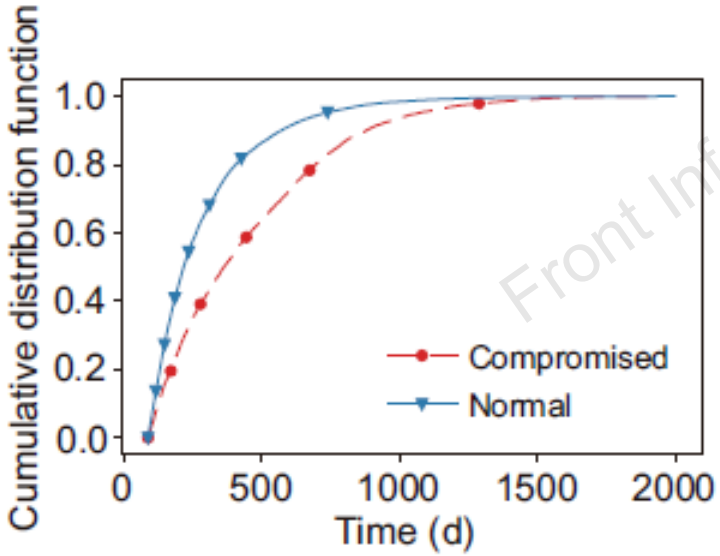
Number of failed logins

Data-Driven Study

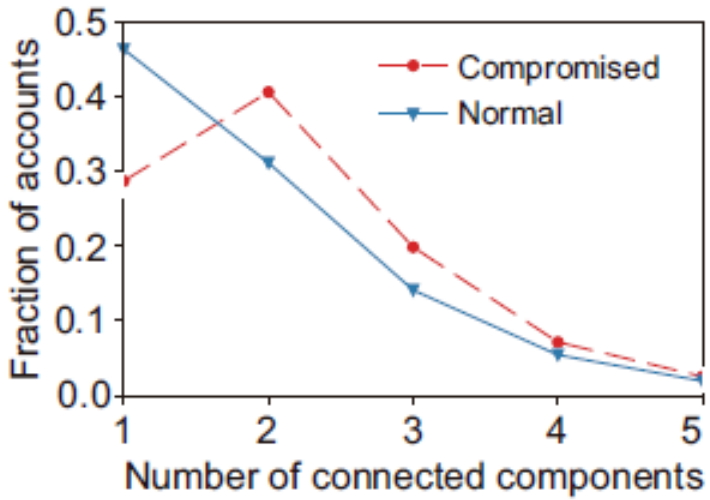
- We have conducted a data-driven study on the compromised accounts and the normal ones from several behavioral patterns.

- We have discovered several distinctive features that would distinguish compromised accounts from normal ones.

➢ Analysis on silent period and behavior sequences related features



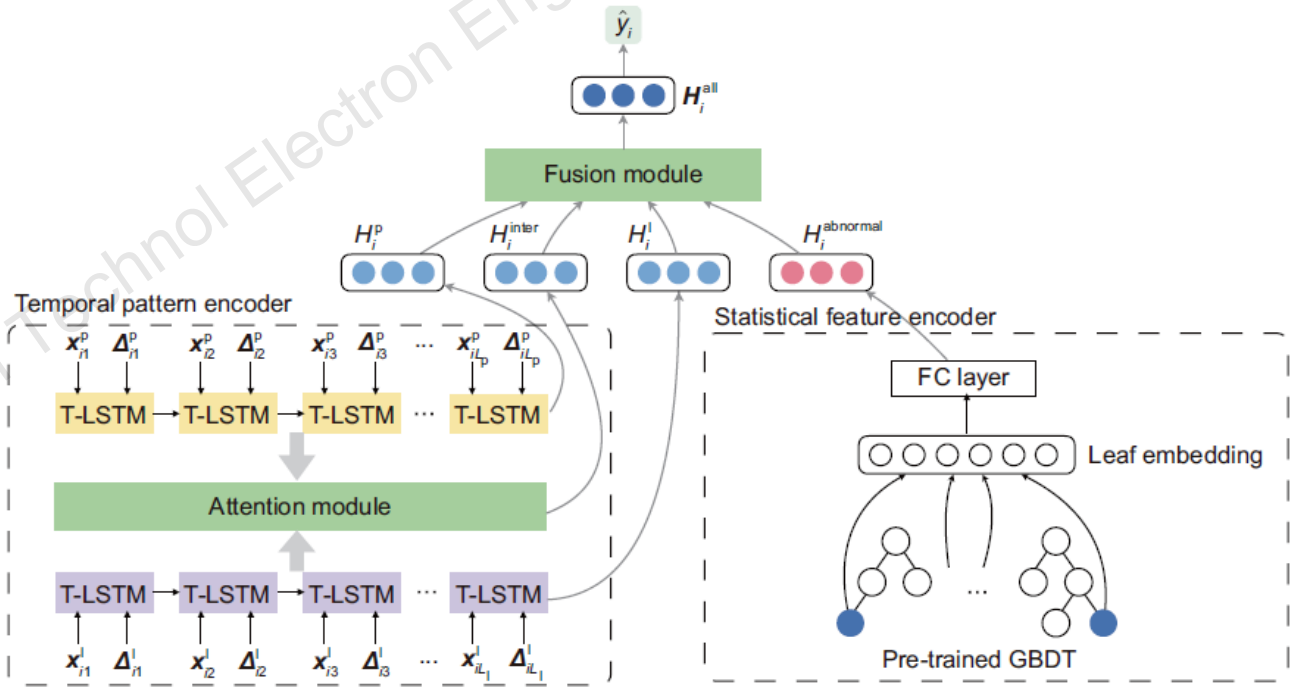
The maximum length of silent periods



Number of connected components in an account's behavior sequences

Method

- We propose an approach named TSF to detect compromised accounts due to PNR.
- The temporal pattern encoder is designed to capture behavioral evolutionary interaction.
- The statistical feature encoder is employed to capture significant operation features.

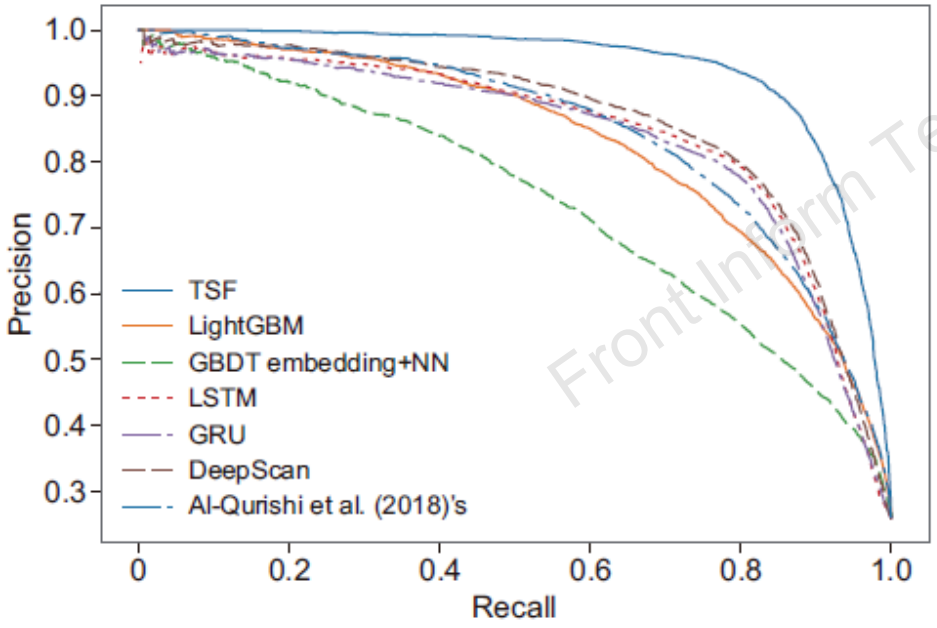


Overview of the TSF model

Major Results

Table 1 Statistics of the Meituan dataset

Type	Number of accounts	Average payment per account	Average login per account	Number of payment records	Number of login records
Compromised	32 388	166	55	5 368 566	1 815 494
Normal	91 575	315	48	28 863 693	4 687 487



Precision–recall curves of TSF and the baselines for the Meituan dataset

Detection results

- The proposed TSF outperforms all the baseline models, with a 3.89%–4.95% AUC increase and a 16.55%–50.49% recall increase for the Meituan dataset.
- The ensemble models outperform the sequential models with higher AUC performance.

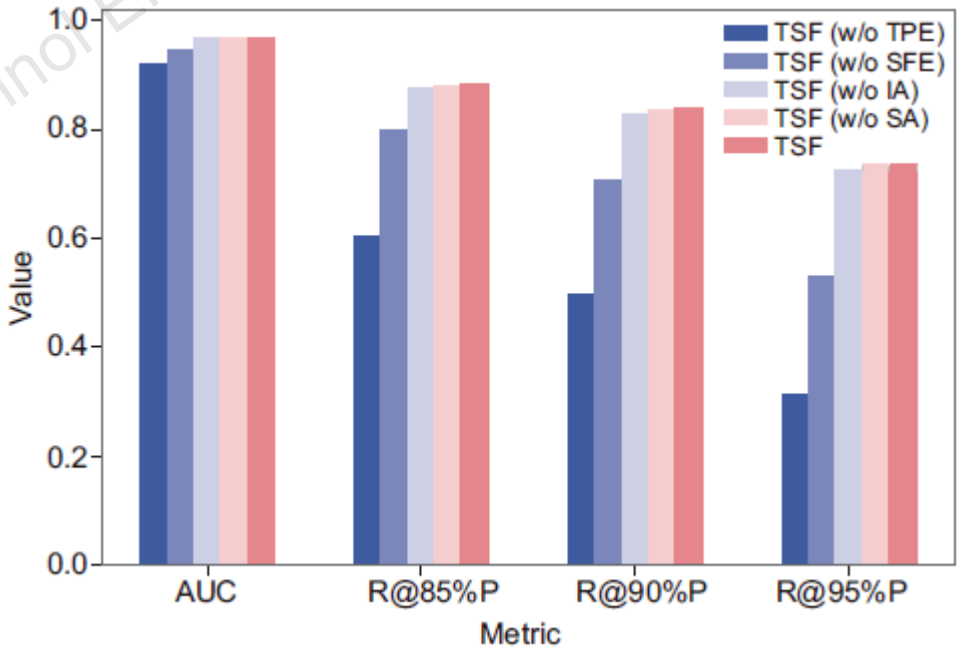
Major Results

Table 1 Statistics of the Meituan dataset

Type	Number of accounts	Average payment per account	Average login per account	Number of payment records	Number of login records
Compromised	32 388	166	55	5 368 566	1 815 494
Normal	91 575	315	48	28 863 693	4 687 487

Ablation study

- All modules of our model contribute to the task of detecting compromised accounts via PNR for the Meituan dataset.
- Removing the temporal pattern encoder has the most severe impact on the performance of our model.



Ablation study for the Meituan dataset

Conclusions

- We focused on the problem of detecting compromised accounts via PNR, a widely prevalent problem in e-commerce platforms. We analyzed and observed that the compromised accounts and the normal ones have distinctive behaviors and evolution patterns.

- We proposed the TSF model to detect the compromised accounts from the normal ones. TSF considers behavioral evolutionary interaction and significant operation features of all accounts.

- Extensive experiments on two real-world datasets demonstrated that our model significantly outperforms some carefully selected baselines.

Future Work

- We will build the association graphs between accounts and entities (such as devices and IP addresses) and utilize graph neural networks to identify the compromised accounts.
- We expand our approach to distinguishing more fine-grained types like lost, stolen, and other abnormally lost numbers for compromised accounts.
- We will continue to develop more comprehensive strategies like data augmentation and transfer learning techniques, for further addressing class imbalance challenges.

Author Information

Yang CHEN is a Professor within the School of Computer Science at Fudan University, China and a Vice Director of the Shanghai Key Lab of Intelligent Information Processing. He leads the Big Data and Networking (DataNET) group since 2014. Before joining Fudan, he was a postdoctoral associate at the Department of Computer Science, Duke University, USA. He received his BSc and Ph.D. degrees from Department of Electronic Engineering, Tsinghua University in 2004 and 2009, respectively. His research interests include social computing, Internet architecture and mobile computing. He served as an OC / TPC Member for many international conferences, including SOSP, SIGCOMM, WWW, MobiSys, ICDCS, IJCAI, AAAI, IWQoS, ICCCN, GLOBECOM, and ICC. He is a Senior Member of ACM, IEEE and CCF.

Author Information

Min GAO received the B.S. degree in the School of Mathematics and Computer Science, Anshan Normal University, China, in 2018, and the M.S. degree in the School of Mathematics and Information, Fujian Normal University, China, in 2021. She is currently working towards the Ph.D. degree in the School of Computer Science, Fudan University, China. Her main research interests include machine learning, social network analysis, and complex networks.