Composite Behavioral Modeling for Identity Theft Detection in Online Social Networks

Cheng Wang, *Senior Member, IEEE*, Hangyu Zhu, and Bo Yang

***Abstract*— In this work, we aim at building a bridge from coarse behavioral data to an effective, quick-response, and robust behavioral model for online identity theft detection. We concentrate on this issue in online social networks (OSNs) where users usually have composite behavioral records, consisting of multidimensional low-quality data, e.g., offline check-ins and online user-generated content (UGC). As an insightful result, we validate that there is a complementary effect among different dimensions of records for modeling users' behavioral patterns. To deeply exploit such a complementary effect, we propose a *joint* (instead of *fused*) model to capture both online and offline features of a user's composite behavior. We evaluate the proposed joint model by comparing it with typical models and their fused model on two real-world datasets: Foursquare and Yelp. The experimental results show that our model outperforms the existing ones, with the area under the receiver operating characteristic curve (AUC) values 0.956 in Foursquare and 0.947 in Yelp, respectively. Particularly, the *recall* (true positive rate) can reach up to 65.3% in Foursquare and 72.2% in Yelp with the corresponding *disturbance rate* (false-positive rate) below 1%. It is worth mentioning that these performances can be achieved by examining only one composite behavior, which guarantees the low response latency of our method. This study would give the cybersecurity community new insights into whether and how real-time online identity authentication can be improved via modeling users' composite behavioral patterns.**

***Index Terms*— Composite behavioral modeling, identity theft detection, joint models, online social networks (OSNs).**

I. Introduction

With the rapid development of the Internet, more and more affairs, e.g., mailing [1], health caring [2], shopping [3], booking hotels, and purchasing tickets, are handled online [4]. Meanwhile, the Internet also brings sundry potential risks of invasions, such as losing financial information [5], identity theft [6], and privacy leakage [3]. Online accounts serve as the agents of users in the cyber world. Online identity theft is a typical online crime which is the deliberate use of another person's account [7], usually as a method to gain a financial advantage or obtain credit and other benefits in another person's name. As a matter of fact, compromised accounts are usually the portals of most cybercrimes [1], such as blackmail [5], fraud [8], and spam [9], [10]. Thus, identity theft detection is essential to guarantee users' security in the cyberworld.

Traditional identity authentication methods are mostly based on access control schemes, e.g., passwords and tokens [11], [12]. But users have some overheads in managing dedicated passwords or tokens. Accordingly, the biometric identification [13]–[15] is delicately introduced to start the era of password-free. However, some disadvantages make these access control schemes still incapable of being effective in real-time online services [16], [17].

1) They are not *nonintrusive*. Users have to spend extra time in the authentication.
2) They are not *continuous*. The defending system will fail to take further protection once the access control is broken.

Behavior-based suspicious account detection [16], [18], [19] is a highly anticipated solution to pursue a nonintrusive and continuous identity authentication for online services. It depends on capturing users' suspicious behavior patterns to discriminate the suspicious accounts. The problem can be divided into two categories: fake/sybil account detection [20] and compromised account detection [21]. The fake/sybil account's behaviors usually do not conform to the behavioral pattern of the majority. Meantime, the compromised account usually behaves in a pattern that does not conform to its previous one, even behaves like fake/sybil accounts. It can be solved by capturing *mutations* of users' behavioral patterns.

Comparing with detecting compromised accounts, detecting fake/sybil accounts is relatively easy since the latter's behaviors are generally more detectable than the former's. It has been extensively studied and can be realized by various population-level approaches, e.g., clustering [22], [23], classification [5], [24]–[26] and statistical or empirical rules [8], [27], [28]. Thus, we *only* focus on the compromised account detection, commonly called *identity theft detection*, based on individual-level behavioral models.

Recently, researchers have proposed the individual-level identity theft detection methods by using suspicious behavior detection [9], [29]–[35]. The efficacy of these methods significantly depends on the sufficiency of behavior records. They are usually suffering from the low-quality of behavior records due to data collecting limitations or some privacy issues [3]. In particular, when a method only utilizes a specific dimension of behavioral data, the efficacy damaged by poor data is possibly enlarged and the scope of application is

Manuscript received March 11, 2020; revised May 10, 2021; accepted June 9, 2021. The work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61972287, in part by the Major Project of the Ministry of Industry and Information Technology of China under Grant TC200H01J, and in part by the Municipal Human Resources Development Program for Outstanding Young Talents in Shanghai. *(Corresponding author: Cheng Wang.)*

The authors are with the Key Laboratory of the Ministry of Education for Embedded System and Service Computing, Department of Computer Science, Tongji University, Shanghai 201804, China (e-mail: chengwang@tongji.edu.cn; 1830823@tongji.edu.cn; boyang@tongji.edu.cn).

Digital Object Identifier 10.1109/TCSS.2021.3092007

limited. Unfortunately, many existing works just concentrate on a specific dimension of users' behavior, such as keystroke [29], clickstream [32], [36], touch-interaction [37], and user generated content (UGC) [9], [33], [34], [38].

In this article, we propose an approach to detect identity theft by using multidimensional behavioral records which are possibly insufficient in each dimension. According to such characteristics, we choose the online social network (OSN) as a typical scenario where most users' behaviors are coarsely recorded [39]. In the Internet era, users' behaviors are composited by offline behaviors, online behaviors, social behaviors, and perceptual/cognitive behaviors. The behavioral data can be collected in many applications, such as offline check-ins in location-based services (LBSs), online tips-posting in instant messaging services, and social relationship-making in online social services. Accordingly, we design our method based on users' composite behaviors by these categories.

In OSNs, user behavioral data that can be used for online identity theft detection are often too low-quality or restricted to build qualified behavioral models due to the difficulty of data collection, the requirement of user privacy, and the fact that some users have a few several behavioral records. We devote ourselves to proving that a high-quality (effective, quick-response, and robust) behavioral model can be obtained by integrally using multidimensional behavioral data, even though the data is extremely insufficient in each dimension.

Generally, there are two paradigms to integrate behavioral data: the *fused* and *joint* manners. Fused models are a relatively simple and straightforward kind of composite behavior models (CBMs). They first capture features in each behavior space and then make a comprehensive metric based on these features in different dimensions. With the possible complementary effect among different behavior spaces, they can act as a feasible solution for integration [7], [17]. However, the identification efficacy can be further improved, since fused models neglect potential links among different spaces of behaviors. We take an example where a person posted a picture in an OSN when he/she visited a park. If this composite behavior is simply separated into two independent parts: he/she once posted a picture and he/she once visited a park, the difficulty in relocating him/her from a group of users is possibly increased, since there are more users satisfy these two simple conditions comparing to the original condition. In contrast, the joint model can sufficiently exploit the correlations between behaviors in different dimensions, then increases the certainty of users' behavior patterns, which contributes to a better identification efficacy. The underlying logic for the difference between the joint and fused models can be also explained by the well-known *Chain Rule for Entropy* [40], which indicates that the entropy of multiple simultaneous events is no more than the sum of the entropies of each individual event, and are equal if the events are independent. It shows that the joint behavior has lower uncertainty comparing to the sum of the uncertainty in each component [41].

Therefore, to fully utilize potential information in composite behaviors for user profiling, we propose a *joint* model, specifically, a joint probabilistic generative model based on Bayesian networks called CBM. It offers a composition of the typical features in two different behavior spaces: check-in location in offline behavior space and UGC in online behavior space. Considering the composite behavior of a user, we assume that the generative mechanism is as follows. When a user plans to visit a venue and simultaneously post tips online, he/she subconsciously selects a specific behavioral pattern according to his/her behavioral distribution. Then, he/she comes up with a topic and a targeted venue based on the present pattern's topic and venue distributions, respectively. Finally, his/her comments are generated following the corresponding topic-word distribution. To estimate the parameters of the mentioned distributions, we adopt the collapsed Gibbs sampling [42].

Based on the joint model CBM, for each composite behavior, denoted by a triple-tuple $(u, v, \mathcal{D})$, we can calculate the chance of user u visiting venue v and posting a tip online with a set of words $\mathcal{D}$. Taking into account different levels of activity of different users, we devise a *relative anomalous score* S_r to measure the occurrence rate of each composite behavior $(u, v, \mathcal{D})$. By these approaches, we finally realize real-time detection (i.e., judging by only one composite behavior) for identity theft suspects.

We evaluate our joint model by comparing it with three typical models and their fused model [17] on two real-world OSN datasets: Foursquare [43] and Yelp [44]. We adopt the area under the receiver operating characteristic curve (AUC) as the detection efficacy. Particularly, the *recall* [true positive rate (TPR)] reaches up to 65.3% in Foursquare and 72.2% in Yelp, respectively, with the corresponding *disturbance rate* [false-positive rate (FPR)] below 1%, while the fused model can only achieve 60.8% and 60.4% in the same condition, respectively. Note that this performance can be achieved by examining only one composite behavior per authentication, which guarantees the low response latency of our detection method. As an insightful result, we learn that the complementary effect does exist among different dimensions of low-quality records for modeling users' behaviors.

The main contributions are summarized into three folds.

1) We propose a joint model, CBM, to capture both online and offline features of a user's composite behavior to fully exploit coarse behavioral data.
2) We devise a relative anomalous score S_r to measure the occurrence rate of each composite behavior for realizing real-time identity theft detection.
3) We perform experiments on two real-world datasets to demonstrate the effectiveness of CBM. The results show that our model outperforms the existing models and has the low response latency.

The rest of this article is organized as follows. We give an overview of our solution in Section II. Then, we present our method in Section III, and make the validation in Section IV. We provide a literature review in Section V. Finally, we draw conclusions in Section VI.

II. Overview of Our Solution

Online identity theft occurs when a thief steals a user's personal data and impersonates the user's account. Generally,

TABLE I
NOTATIONS OF PARAMETERS

Variable	Description
w	the word in UGC
v	the venue or place
π_u	the community memberships of user u, expressed by a multinomial distribution over communities
θ_c	the interests of community c, expressed by a multinomial distribution over topics
ϑ_c	a multinomial distribution over spatial items specific to community c
ϕ_z	a multinomial distribution over words specific to topic z
$\alpha, \beta, \gamma, \eta$	Dirichlet priors to multinomial distributions θ_c, ϕ_z, π_u and ϑ_c, respectively

a thief usually first gathers information about a targeted user to steal his/her identity and then use the stolen identity to interact with other people to get further benefits [4]. Criminals in different online services usually have different motivations.

An OSN user's behavior is usually composed of online and offline behaviors occurring in different behavioral spaces [17]. Based on this fact, we aim at devising a joint model to embrace them into a unified model to deeply extract information.

Before presenting our joint model, named CBM, we provide some conceptions as the preparations. The relevant notations are listed in Table I.

Definition 1 (Composite Behavior): A composite behavior, denoted by a four-tuple $(u, v, \mathcal{D}, t)$, indicates that at time t, user u visits venue v and simultaneously posts a tip consisting of a set of words $\mathcal{D}$ online.

In this work, the representation of a composite behavior can be simplified into a triple-tuple $(u, v, \mathcal{D})$. We remark that for a composite behavior, the occurring time t is a significant factor. Two types of *time attributes* play important roles in digging potential information for improving the identification. The first is the *sequential correlation* of behaviors. However, in some OSNs, the time intervals between adjacent behavioral records are usually overlong, which leads that the sequential correlations cannot be captured effectively. The second is the *temporal property* of behaviors, e.g., periodicity and preference variance over time. However, in some OSNs, the occurring time is recorded with a low resolution, e.g., by day, which shields the possible dependency of a user's behavior on the occurring time. Thus, it is difficult to obtain reliable time-related features of users' behaviors. Since we aim to propose a practical method based on uncustomized datasets of user behaviors, we only concentrate on the dependency between a user's check-in location and tip-posting content of each behavior, taking no account of the impact of specific occurring time in this work. Thus, the representation of a composite behavior can be simplified into a triple-tuple $(u, v, \mathcal{D})$ without confusion in this article. The graphical representation of our joint model CBM is demonstrated in Fig. 1.

Our model is mainly based on the following two assumptions.

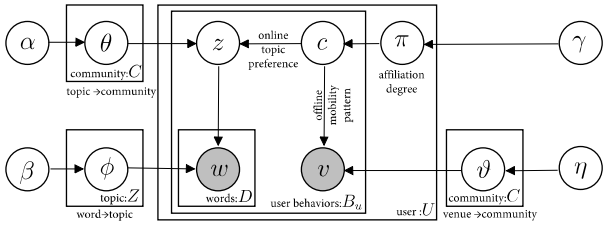


Fig. 1. Graphical representation of the joint model. The parameters and notions are explained in Table I.

1) Each user behaves in multiple patterns with different possibilities.
2) Users with similar behavioral patterns have similar interests in topics and places.

To describe the features of users' behaviors, we first introduce the *topic* of tips.

Definition 2 (Topic, [45]): Given a set of words $\mathcal{W}$, a topic z is represented by a multinomial distribution over words, denoted by ϕ_z, whose each component $\phi_{z,w}$ denotes the probability of word w occurring in topic z.

Next, we formulate a specific *behavioral pattern* of users by a conception called *community*.

Definition 3 (Community): A community is a set of users with the same behavioral pattern. Let $\mathcal{C}$ denote the set of all communities. A community $c \in \mathcal{C}$ has two critical parameters.

1) A topic distribution θ_c, whose component, say $\theta_{c,z}$, indicates the probability that the users in community c send a message with topic z.
2) A spatial distribution ϑ_c, whose component, say $\vartheta_{c,v}$, represents the chance that users in community c visit venue v.

More specifically, we assume that a community is formed by the following procedure. Each user u is included in communities according to a multinomial distribution, denoted by π_u. That is, each component of π_u, say $\pi_{u,c}$, denotes u's affiliation degree to community c. Similarly, we allocate each community c with a topic distribution θ_c to represent its online topic preference and a spatial distribution ϑ_c to represent its offline mobility pattern.

III. METHOD

A. CBM

Generally, users take actions according to their regular behavioral patterns which are represented by the corresponding communities (Definition 3). We present the behavioral generative process in Algorithm 1: When a user u is going to visit a venue and post online tips there, he/she subconsciously selects a specific behavioral pattern, denoted by community c, according to his/her community distribution π_u (line 11). Then, he/she comes up with a topic z and a targeted venue v based on the present community's topic and venue distributions (θ_c and ϑ_c, respectively) (lines 12 and 13). Finally, the words of his/her tips in $\mathcal{D}$ are generated following the topic-word distribution ϕ_z (line 15).

Exact inference of our joint model CBM is difficult due to the intractable normalizing constant of the posterior

Algorithm 1 Joint Probabilistic Generative Process

for *each community* $c \in \mathcal{C}$ **do**
  Sample the distribution over topics $\theta_c \sim Dirichlet(\cdot|\alpha)$
  Sample the distribution over venues $\vartheta_c \sim Dirichlet(\cdot|\eta)$
end for
for *each topic* $z \in \mathcal{Z}$ **do**
  Sample the distribution over words $\phi_z \sim Dirichlet(\cdot|\beta)$
end for
for *each user* $u \in \mathcal{U}$ **do**
  Sample the distribution over communities $\pi_u \sim Dirichlet(\cdot|\gamma)$
  for *each composite behavior* $(u, v, \mathcal{D}) \in \mathcal{B}_u$ **do**
    Sample a community indicator $c \sim Multi(\pi_u)$
    Sample a topic indicator $z \sim Multi(\theta_c)$
    Sample a venue $v \sim Multi(\vartheta_c)$
    for *each word* $w \in \mathcal{D}$ **do**
      Sample a word $w \sim Multi(\phi_z)$
    end for
  end for
end for

distribution [42]. We adopt collapsed Gibbs sampling for approximately estimating distributions (i.e., θ, ϑ, ϕ, and π). As for the hyperparameters, we take the fixed values, i.e., $\alpha = 50/Z$, $\gamma = 50/C$, and $\beta = \eta = 0.01$, where Z and C are the numbers of topics and communities, respectively.

In each iteration, for each composite behavior $(u, v, \mathcal{D})$, we first sample community c according to the following equation:

$$P(c|\mathbf{c}^{\neg}, \mathbf{z}, \mathbf{v}, u) \propto \left(n_{u,c}^{\neg} + \gamma\right) \frac{n_{c,z}^{\neg} + \alpha}{\sum_{z'} \left(n_{c,z'}^{\neg} + \alpha\right)} \frac{n_{c,v}^{\neg} + \eta}{\sum_{v'} \left(n_{c,v'}^{\neg} + \eta\right)} \tag{1}$$

where $\mathbf{c}^{\neg}$ denotes the community allocation for all composite behaviors except the current one; $\mathbf{z}$ denotes the topic allocation for all composite behaviors; $n_{u,c}$ denotes the number of times that community c is generated by user u; $n_{c,z}$ denotes the number of times that topic z is generated by community c; $n_{c,v}$ denotes the number of times that venue v is visited by users in community c; a superscript $\neg$ denotes something except the current one.

Then, given a community c, we sample topic z according to the following equation:

$$P(z|\mathbf{z}^{\neg}, \mathbf{c}, \mathcal{D}) \propto \left(n_{c,z}^{\neg} + \alpha\right) \prod_{w \in \mathcal{D}} \frac{n_{z,w}^{\neg} + \beta}{\sum_{w'} \left(n_{z,w'}^{\neg} + \beta\right)} \tag{2}$$

where $n_{z,w}$ denotes the number of times that word w is generated by topic z.

The inference algorithm is presented in Algorithm 2. We first randomly initialize the topic and community assignments for each composite behavior (lines 2–4). Then, we update the community and topic assignments for each composite behavior based on (1) and (2) in each iteration (lines 6–9). Finally, we estimate the parameters, test the coming cases and update

Algorithm 2 Inference Algorithm of Joint Model CBM

Require: user composite behavior collection $\mathcal{B}$, number of iterations I, start saving step I_b, saving lag I_s, start training sequence number N_b, end training sequence number N_e, hyperparameters α, β, γ and η
Ensure: estimated parameters $\hat{\theta}$, $\hat{\vartheta}$, $\hat{\phi}$, $\hat{\pi}$
1: Create temporary variables $\theta^{\mathbf{sum}}$, $\vartheta^{\mathbf{sum}}$, $\phi^{\mathbf{sum}}$ and $\pi^{\mathbf{sum}}$, initialize them with zero, set testing sequence number $N_t = 0$ and let $\mathcal{B}(N_t)$ denote the corresponding training collection for testing behaviors which sequence number values N_t
2: **for** *each composite behavior* $(u, v, \mathcal{D}) \in \mathcal{B}(N_t)$ **do**
3: Sample community and topic randomly
4: **end for**
5: **for** $iteration = 1$ to I **do**
6: **for** *each behavior* $(u, v, \mathcal{D}) \in \mathcal{B}(N_t)$ **do**
7: Sample community c according to Eq. (1)
8: Sample topic z according to Eq. (2)
9: **end for**
10: **if** $(iteration > I_b)$ and $(iteration \bmod I_s == 0)$ **then**
11: **return** model parameters as follows:

$$\theta_{c,z} = \frac{n_{c,z} + \alpha}{\sum_{z'} \left(n_{c,z'} + \alpha\right)}; \quad \vartheta_{c,v} = \frac{n_{c,v} + \eta}{\sum_{v'} \left(n_{c,v'} + \eta\right)}$$

$$\pi_{u,c} = \frac{n_{u,c} + \gamma}{\sum_{c'} \left(n_{u,c'} + \gamma\right)}; \quad \phi_{z,w} = \frac{n_{z,w} + \beta}{\sum_{w'} \left(n_{z,w'} + \beta\right)}$$

12: Evaluate corresponding test cases and update N_t++; N_b++; N_e++
13: **end if**
14: **end for**

the training set for every I_s iterations since I_bth iteration (lines 10–13) to address concept drift.

To overcome the problem of data insufficiency, we adopt the tensor decomposition [46] to discover their potential behaviors. In our experiment, we use the Twitter-latent Dirichlet allocation (LDA) [47] to obtain each UGC's topic and construct a tensor $\mathbf{A} \in \mathbb{R}^{N \times M \times L}$, with three dimensions standing for users, venues, and topics. Then, $\mathbf{A}(u, v, z)$ denotes the frequency that a user u posting a message on topic z in venue v. We can decompose $\mathbf{A}$ into the multiplication of a core tensor $\mathbf{S} \in \mathbb{R}^{d_U \times d_V \times d_Z}$ and three matrices, $\mathbf{U} \in \mathbb{R}^{N \times d_U}$, $\mathbf{V} \in \mathbb{R}^{M \times d_V}$, and $\mathbf{Z} \in \mathbb{R}^{L \times d_Z}$, if using a tucker decomposition model, where d_U, d_V, and d_Z denote the number of latent factors; N, M, and L denote the number of users, venues and topics. An objective function to control the errors is defined as

$$\begin{aligned}\mathcal{L}(\mathbf{S}, \mathbf{U}, \mathbf{V}, \mathbf{Z}) &= \frac{1}{2}\|\mathbf{A} - \mathbf{S} \times_U \mathbf{U} \times_V \mathbf{V} \times_Z \mathbf{Z}\|^2 \\ &\quad + \frac{\lambda}{2}\left(\|\mathbf{S}\|^2 + \|\mathbf{U}\|^2 + \|\mathbf{V}\|^2 + \|\mathbf{Z}\|^2 + \sum_{(i,j) \in \mathcal{F}} u_i^T u_j\right)\end{aligned}$$

where $\mathcal{F}$ is a set of friend pairs (i, j). $\mathbf{A}^* = \mathbf{S} \times_U \mathbf{U} \times_V \mathbf{V} \times_Z \mathbf{Z}$ is the potential frequency tensor, and $\mathbf{A}^*(u, v, z)$ denotes the frequency that user u may post a message on topic z in venue

v. A higher $\mathbf{A}^*(u, v, z)$ indicates that the user u has a higher chance to do this kind of behavior in the future. We limit the competition space to the behavior space of u's friends, that is

$$\{(u, v, z)|\mathbf{A}(u', v, z) > 0, (u, u') \in \mathcal{F}\}$$

and select the top 20 behaviors as his/her latent behaviors to improve data quality.

B. Identity Theft Detection Scheme

By the parameters $\hat{\psi} = \{\hat{\theta}, \hat{\vartheta}, \hat{\phi}, \hat{\pi}\}$ learned from the inference algorithm (Algorithm 2), we estimate the *logarithmic anomalous score* (S_l) of a composite behavior $(u, v, \mathcal{D})$ by the following equation:

$$\begin{aligned} S_l(u, v, \mathcal{D}) &= -\lg P(v, \mathcal{D}|u) \\ &= -\lg\left(\sum_c \hat{\pi}_{u,c}\hat{\vartheta}_{c,v}\sum_z \hat{\theta}_{c,z}\left(\prod_{w\in\mathcal{D}}\hat{\phi}_{z,w}\right)^{\frac{1}{|\mathcal{D}|}}\right). \end{aligned} \tag{3}$$

However, we may mistake some normal behaviors occurring with low probability, e.g., the normal behaviors of users whose behavioral diversity and entropy are both high, for suspicious behaviors. Thus, we propose a *relative anomalous score* (S_r) to indicate the trust level of each behavior by the following equation:

$$S_r(u, v, \mathcal{D}) = 1 - P(u|v, \mathcal{D}) = 1 - \frac{P(v, \mathcal{D}|u)P(u)}{\sum_{u'} P(v, \mathcal{D}|u')P(u')}. \tag{4}$$

For reducing computational complexity, we randomly select $n = 40$ users to estimate the relative anomalous score S_r for each composite behavior. The selection process of hyperparameter n is omitted due to space limitations. Our experimental results in Section IV show that the approach based on S_r outperforms one based on S_l.

IV. Evaluation

In this section, we present the experimental results to evaluate the proposed joint model CBM, and validate the efficacy of the joint model for identity theft detection on real-world OSN datasets.

A. Datasets

Our experiments are conducted on two real-life OSN datasets: Foursquare [43] and Yelp [44], which are two well-known online social networking service providers. Foursquare is an LBS provider and encourages users to share their current locations and comments with others. The adopted Foursquare dataset contains the check-in history of 31 494 users in LA. Yelp is another popular location-based social networking service provider, which publishes crowd-sourced reviews about local businesses. The adopted Yelp dataset contains the tips of 80 593 users. In both datasets, there are no URLs or other sensitive terms. Both datasets contain users' social ties and behavioral records. Each social tie contains *user-ID* and *friend-ID*. Each behavior record contains the *user-ID*, *venue-ID*, *timestamp*, and *UGC*. The basic statistics are shown in Table II. Examples of our behavior records are illustrated in Table III.

TABLE II
STATISTICS OF FOURSQUARE AND YELP DATASETS

	Foursquare	Yelp
# of users	31,493	80,592
# of venue	143,923	42,051
# of check-ins	267,319	491,393

TABLE III
USER'S BEHAVIOR RECORDS IN FOURSQUARE DATASET

User-ID (Anonymized)	Venue-ID (Anonymized)	Timestamp	Message Content
1	1	1299135219	Pen is better! The class sizes are only 20:1!!! GO PANTHERS!
1	2	1299135270	GO PANTHERS!
1	3	1299135328	Save a whale, eat a Sea King!
1	4	1301004689	Best school in the world. PV High is an uber fail compared to here.
1	4	1303711907	The best teachers only come from Pen.
1	5	1303971421	DJ Mike too the rescue!

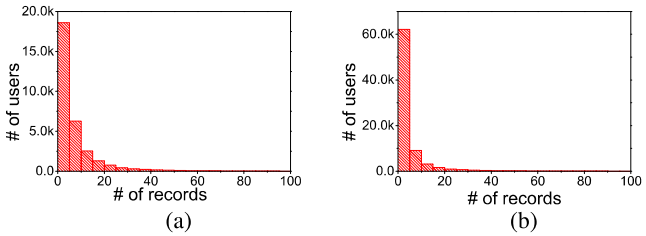


Fig. 2. Distribution of user record counts. (a) Foursquare. (b) Yelp.

We count each user's records and present the results in Fig. 2. It shows that most users have less than five records in both datasets. The quality of these datasets is too *poor* to model individual-level behavioral patterns for the majority of users, which confronts our method with a big challenge.

B. Experiment Settings

1) Simulation of Postintrusion Behavior Pattern: The so-called *postintrusion behavior* refers to the behavior of a thief via the compromised account. We simulate three typical kinds of postintrusion behaviors based on different kinds of scenarios. Specifically, in each experiment, we randomly simulate 5% of all behavior records in the testing set as *anomalous behaviors* and repeat ten times.

a) Behavioral displacement: Most thieves usually take actions under specific aims. Such actions can be easily detected since it is quite different from normal behaviors. We focus on a harder scenario, where thieves show no specific aims and just do as usual in the compromised account.

Accordingly, we swap two user's behavioral records to simulate the scenario.

b) Behavioral imitation: Some extremely cunning thieves try to imitate the normal user's behavioral pattern and maintain part of the victim's behavioral pattern to get further benefits from the victim's friends. It is harder to detect this kind of postintrusion behaviors. We note that in our work it will make no sense if thieves completely imitate behavioral patterns of victims. Accordingly, we simulate two kinds of variations, i.e., the guise in venue and guise in content, respectively. We swap two normal user's behavioral records that have similar venues (two venues have similar tags) to simulate the guise in the venue. Besides, we swap two normal user's behavioral records that have a similar topic to simulate a scenario where thieves may imitate victims' habits to cheat their friends.

c) Random synthesis: To simulate intangible behavioral patterns, we randomly generate behavioral records as postintrusion behaviors.

2) Representative Models: We compare our joint model CBM to some representative models in OSNs. For two different dimensional behaviors, we choose CF-KDE and LDA as baseline models, respectively. For offline check-in behaviors, mixture kernel density estimate (MKDE) is a typical spatial model describing a user's offline behavioral pattern [48]. However, it assumes that users tend to behave like their friends in the same chance and it has not quantified the potential influence of different friends. To improve its performance, we introduce a collaborative filtering method to cooperate with MKDE, which is named CF-KDE. For online UGC, LDA has been successfully applied for analyzing text from user messages on online social networks [45], [49]. The LDA detection algorithm uses users' documents as an input and detects the corresponding topics. For online social networks with small message lengths, topic detection is shown to be less efficient. For this reason, we aggregate the UGC of each user and his/her friends in the training set as a document and then run topic modeling on the documents. In Table IV, we list the features of these models. Next, we will give a detailed description of how to deploy them in this work.

a) CF-KDE: Before presenting the CF-KDE model, we introduce the MKDE to give a brief prior knowledge. MKDE mainly utilizes a bivariate density function in the following equations to capture the spatial distribution for each user:

$$f_{\mathrm{KDE}}(e|E,h) = \frac{1}{n}\sum_{j=1}^{n} K_h(e-e^j) \tag{5}$$

$$K_h(x) = \frac{1}{2\pi h}\exp\left(-\frac{1}{2}x^T H^{-1}x\right), \quad \mathbf{H} = \begin{pmatrix} h & 0 \\ 0 & h \end{pmatrix} \tag{6}$$

$$f_{\mathrm{MKDE}}(e|E,h) = \alpha f_{\mathrm{KD}}(e|E_1) + (1-\alpha) f_{\mathrm{KD}}(e|E_2). \tag{7}$$

In (5), $E = \{e^1, \ldots, e^n\}$ is a set of historical behavioral records for a user and $e^j = \langle x, y\rangle$ is a 2-D spatial location (i.e., an offline behavior). Equation (6) is a kernel function and $\mathbf{H}$ is the bandwidth matrix. MKDE adopts (7), where E_1 is a set of an individual's historical behavioral records (individual component), E_2 is a set of his/her friends' historical behavioral records (social component), and α is the weight variable for the individual component. In this article, to detect identity thieves, we compute a surprise index S_e in (8) for each behavior e, defined as the negative log-probability of individual u's conducting behavior e

$$S_e = -\log f_{\mathrm{MKDE}}(e|E_u, h_u). \tag{8}$$

Furthermore, we can select the top-N behaviors with the highest S_e as suspicious behaviors.

We introduce a collaborative filtering method to improve performance. Based on the historical behavioral records, it establishes a user–venue matrix $\mathbf{R}_{|\mathbf{U}|\times|\mathbf{V}|}$, where U and V are the number of users and venues, respectively; $\mathbf{R}_{ij} = 1$ if user i has visited venue j in the training set, otherwise $\mathbf{R}_{ij} = 0$. We adopt a matrix factorization method with an objective function in (9) to obtain feature vectors for each user and venue

$$L = \min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\sum_{i=1}^{U}\sum_{j=1}^{V}\left(\mathbf{R}_{ij} - u_i^T v_j\right)^2 + \frac{\lambda_1}{2}\sum_{i=1}^{U} u_i^T u_i + \frac{\lambda_2}{2}\sum_{j=1}^{V} v_j^T v_j. \tag{9}$$

Specifically, we let

$$u_i = \left(u_i^{(1)}, u_i^{(2)}, \ldots, u_i^{(k)}\right)^T \text{ and } v_j = \left(v_j^{(1)}, v_j^{(2)}, \ldots, v_j^{(k)}\right)^T.$$

We adopt a stochastic gradient descent algorithm in (10) and (11) in the optimization process

$$u_i^{(k)} \leftarrow u_i^{(k)} - \alpha\left(\sum_{j=1}^{V}\left(\mathbf{R}_{ij} - u_i^T v_j\right)v_j^{(k)} + \lambda_1\, u_i^{(k)}\right) \tag{10}$$

$$v_j^{(k)} \leftarrow v_j^{(k)} - \alpha\left(\sum_{i=1}^{U}\left(\mathbf{R}_{ij} - u_i^T v_j\right)u_i^{(k)} + \lambda_2 v_j^{(k)}\right). \tag{11}$$

Consequently, we can figure out $\hat{\mathbf{R}} = \mathbf{U}^T\mathbf{V}$, and use $\hat{r}_{ij} = u_i^T v_j$ as the weight variable for the KDE model. To detect anomalous behaviors, we use (12) to measure the surprising index for each behavior e

$$S_e = -\log\frac{\sum_{j=1}^{n}\hat{r}_{uj}K_h(e-e^j)}{\sum_{j=1}^{n}\hat{r}_{uj}}. \tag{12}$$

We assert that the top-N behaviors with the highest S_e are suspicious behaviors.

b) LDA: A user's online behavior pattern can be denoted as the mixing proportions for topics. We aggregate the UGC of each user and his/her friends in the training set as a document, then use LDA to obtain each user's historical topic distribution θ_{his}. To get their present behavioral topic distributions θ_{new}

TABLE IV
BEHAVIORS ADOPTED IN DIFFERENT MODELS

	Online UGC	Offline Check-in
CF-KDE	NO	YES
LDA	YES	NO
FUSED	YES	YES
JOINT	YES	YES

TABLE V
CONFUSION MATRIX FOR BINARY CLASSIFICATION

Predicted Condition	True Condition	
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

in the testing set. For each behavior, we count the number of words assigned to the kth topic, and denote it by $n(k)$. The kth component of the topic proportion vector can be computed by

$$\theta_{\text{new}}^{(k)} = \frac{n(k) + \alpha}{\sum_{i=1}^{K}(n(i) + \alpha)} \tag{13}$$

where K is the number of topics, and α is a hyperparameter.

To detect anomalous behaviors, we measure the distance between a user's historical and present topic distribution by using the Jensen-Shannon (JS) divergence in the following equations:

$$D_{\text{KL}}(\theta_{\text{his}}, \theta_{\text{new}}) = \sum_{i=1}^{K} \theta_{\text{his}}^{(i)} \cdot \ln\left(\frac{\theta_{\text{his}}^{(i)}}{\theta_{\text{new}}^{(i)}}\right) \tag{14}$$

$$D_{\text{JS}}(\theta_{\text{his}}, \theta_{\text{new}}) = \frac{1}{2}[D_{\text{KL}}(\theta_{\text{his}}, M) + D_{\text{KL}}(\theta_{\text{new}}, M)] \tag{15}$$

where $M = ((\theta_{\text{his}} + \theta_{\text{new}})/2)$. We consider that the top-N behaviors with the highest $D_{\text{JS}}(\theta_{\text{his}}, \theta_{\text{new}})$ are suspicious behaviors.

c) Fused model: Egele *et al.* [7] propose COMPA which directly combines use users' explicit behavior features, e.g., languages, links, message sources, and so on. In our case, we introduce a fused model [17], which combines users' implicit behavior features discovered by CF-KDE and LDA to detect identity theft. We try different thresholds for the CF-KDE model and the LDA model (i.e., different classifiers). For each pair (i.e., a CF-KDE model and an LDA model), we treat any behavior that fails to pass either the identification model as suspicious behavior and compute TPR and FPR to draw the receiver operating characteristic (ROC) curve and estimate the AUC value.

3) Metrics: For the convenience of description, we first give a confusion matrix in Table V.

In the experiments, we set *anomalous behaviors* as *positive instances*, and focus on the following four metrics, since the identity theft detection is essentially an *imbalanced binary classification problem* [50].

a) TPR/Recall: TPR is computed by $[\text{TP}/(\text{TP} + \text{FN})]$, and indicates the proportion of true positive instances in all positive instances (i.e., the proportion of anomalous behaviors that are detected in all anomalous behaviors). It is also known as *recall*. Specifically, we named it *detection rate*.

b) FPR: FPR is computed by $[\text{FP}/(\text{FP} + \text{TN})]$, and indicates the proportion of false-positive instances in all negative instances (i.e., the proportion of normal behaviors that are mistaken for anomalous behaviors in all normal behaviors). Specifically, we named it *disturbance rate*.

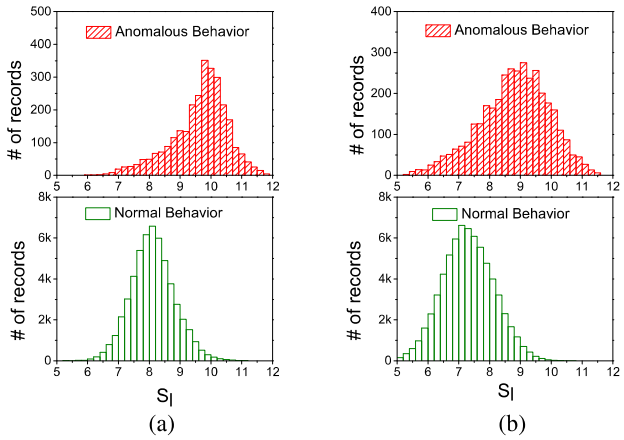


Fig. 3. Histogram of *logarithmic anomalous score* S_l [defined in (3)] for each behavior. (a) Foursquare. (b) Yelp.

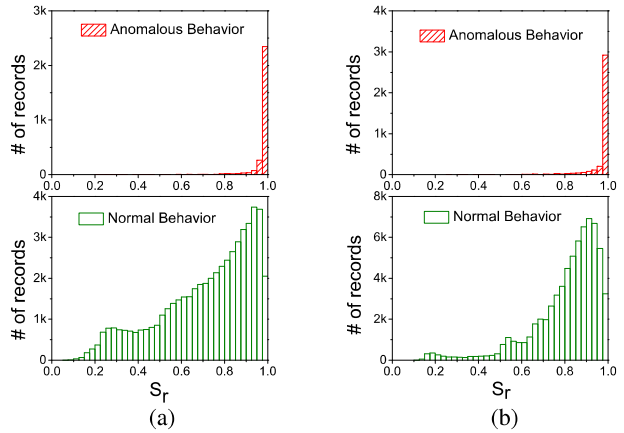


Fig. 4. Histogram of *relative anomalous score* S_r [defined in (4)] for each behavior. (a) Foursquare. (b) Yelp.

c) Precision: The precision is computed by $[\text{TP}/(\text{TP} + \text{FP})]$, and indicates the proportion of true positive instances in all predicted positive instances (i.e., the proportion of anomalous behaviors that are detected in all suspected cases).

d) AUC: Given a rank of all test behaviors, the AUC value can be interpreted as the probability that a classifier/predictor will rank a randomly chosen positive instance higher than a randomly chosen negative one.

4) Threshold Selection: It is an important issue in classification tasks. Specifically, we take a case where $C = 30$ and $Z = 20$ an example to present the threshold selection strategy. The parameter sensitivity analysis will be conducted in Section IV-B5. We compare the distribution of *logarithmic anomalous score* S_l (or *relative anomalous score* S_r) for normal behaviors with that for anomalous behaviors. Figs. 3 and 4 present the differences between normal and anomalous behaviors in terms of the distributions of S_l and S_r, respectively. They show that the differences are both significant, and the difference in terms of S_r is much more obvious.

To obtain a reasonable threshold, we take fivefold cross validation for the training set and focus on the performance where the threshold changes from 0.975 to 1, since this range contains 81.5% (81.4%) of all anomalous behaviors and 3.9% (4.8%) of all normal behaviors in Foursquare (Yelp).

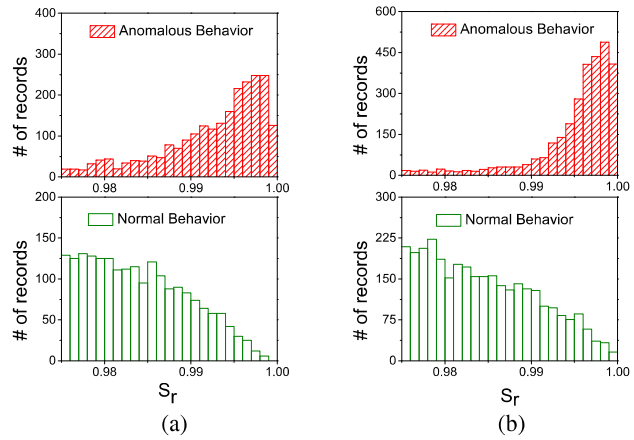


Fig. 5. Partial of the distribution of *relative anomalous score* S_r [defined in (4)] for each behavior. (a) Foursquare. (b) Yelp.

TABLE VI

SUMMARY OF DIFFERENT METRICS WITH THE THRESHOLD 0.989 FOR FOURSQUARE AND 0.992 FOR YELP, RESPECTIVELY

	Foursquare	Yelp
Precision	79.91%	83.55%
Recall (TPR)	62.32%	68.75%
FPR	0.85%	0.71%
AUC	0.956	0.947
TNR	99.15%	99.29%
FNR	37.68%	31.25%
Accuracy	97.26%	97.76%
F1	0.700	0.754

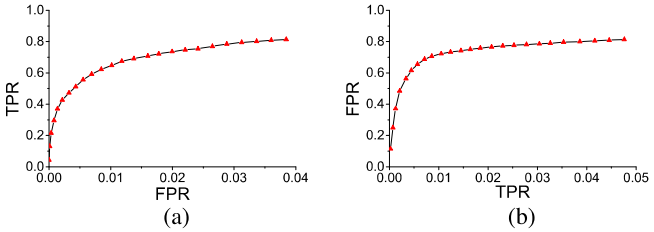


Fig. 6. Partial of ROC curve of identity theft detection. (a) Foursquare. (b) Yelp.

The detailed trade-offs are demonstrated in Figs. 5 and 6 from different aspects. To optimize the trade-offs of detection performance, we define the detection *Cost* in the following equation:

$$\text{Cost} = \frac{\#\text{ of newly mistaken normal behaviors}}{\#\text{ of newly identified anomalous behaviors}}. \quad (16)$$

We present the threshold-cost curve in Fig. 7. It shows that a smaller threshold usually corresponds to a larger cost.

We select the minimum threshold satisfying that the corresponding cost is less than 1. Thus, we choose 0.989 and 0.992 as the thresholds for Foursquare and Yelp, respectively. Under them, our joint model CBM reaches 62.32% (68.75%) in TPR and 0.85% (0.71%) in FPR on Foursquare (Yelp). Refer to Table VI for details.

5) Parameter Sensitivity Analysis: Parameter tuning is another important part of our work. The performance of our model is indeed sensitive to the number of communities (C) and topics (Z). Therefore, we study the impact of varying

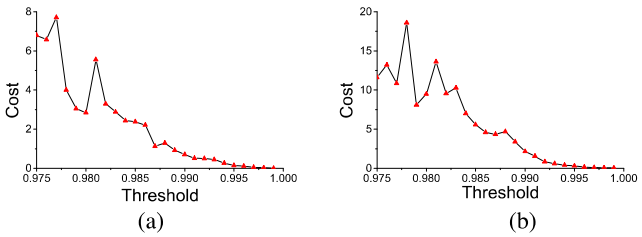


Fig. 7. Detection costs with different thresholds. (a) Foursquare. (b) Yelp.

TABLE VII

AUC ON FOURSQUARE (YELP) DATASET

	$C = 10$	$C = 20$	$C = 30$
$Z = 10$	0.876 (0.910)	0.945 (0.936)	0.953 (0.945)
$Z = 20$	0.917 (0.915)	0.946 (0.938)	**0.956** (**0.947**)
$Z = 30$	0.922 (0.917)	0.947 (0.938)	0.957 (0.947)

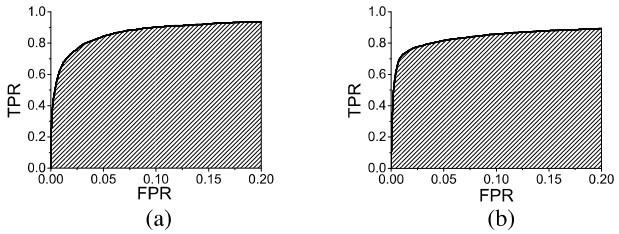


Fig. 8. ROC curves of identity theft detection via the joint model CBM. (a) Foursquare. (b) Yelp.

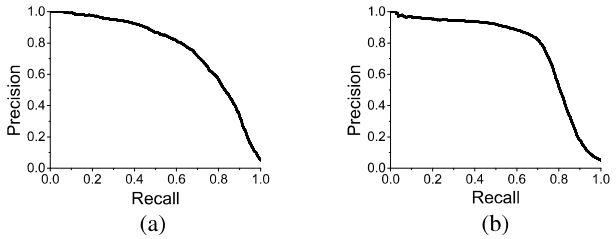


Fig. 9. Precision–recall curves of identity theft detection via the joint model CBM. (a) Foursquare. (b) Yelp.

TABLE VIII

DETECTION RATES WITH DISTURBANCE RATES

	Foursquare	Yelp
Disturbance Rate= 0.1%	30.8%	31.7%
Disturbance Rate= 1.0%	65.3%	72.2%

parameters in our model. We select the *relative anomalous score* S_r as the test variable, and evaluate the performance of our model by changing the values of C and Z. The experimental results are summarized in Table VII.

From the results on both datasets, the detection efficacy goes stable when Z reaches 20 and C has a larger impact on the efficacy. Thus, we set $C = 30$ and $Z = 20$ in our joint model, and present the ROC and precision–recall curves in Figs. 8 and 9, respectively. Specifically, we present detection rate (TPR) in Table VIII, where the disturbance rate (FPR) reaches 1% and 0.1%, respectively.

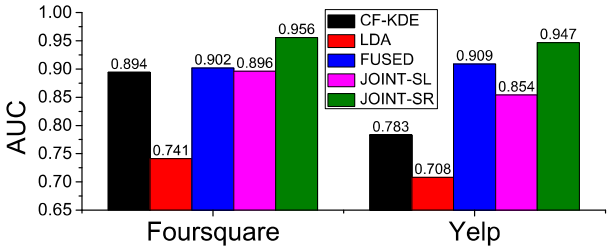


Fig. 10. Identity theft detection efficacy.

C. Performance Comparison

We compare the performance of our method with the typical ones in terms of *detection efficacy* (AUC) and *response latency*. The latter denotes the number of behaviors in the test set needed to accumulate for detecting a specific identity theft case.

1) Detection Efficacy Analysis: In Fig. 10, we present the results of all comparison methods.

Our joint model outperforms all other methods on the two datasets. The AUC value reaches 0.956 and 0.947 under normal behavior spam attacks in Foursquare and Yelp datasets, respectively.

There are three reasons for the outstanding performance. First, it embraces different types of behaviors and exploits them in a unified model. Second, it takes advantage of the community members' and friends' behavior information to overcome the data insufficiency and concept drift [51] in individual-level behavioral patterns. Finally, it utilizes correlations among different behavioral spaces.

For partially behavioral imitation attacks, our joint model also shows a nice performance. The detail results can be found in Figs. 11–13.

From the results, we have several other interesting observations.

1) LDA model performs poorly in both datasets which may indicate its performance is strongly sensitive to the data quality.
2) CF-KDE and LDA model performs not well in Yelp dataset comparing to Foursquare dataset, but the fused model [17] observes a surprising reversion.
3) The joint model based on *relative anomalous score* S_r outperforms the model based on *logarithmic anomalous score* S_l.
4) The joint model (i.e., JOINT-SR, the joint model in the following content of Section IV-C all refer to the joint model based on S_r) is indeed superior to the fused model.

For random behavior attacks, our joint model shows better performance. Specifically, we apply the *logarithmic anomalous score* (S_l) in (3) for detecting the kind of attack. Besides, we present the details in Table IX, where the disturbance rate (FPR) reaches 1% and 0.1%, respectively.

2) Response Latency Analysis: For each model, we also evaluate the relationship between the efficacy and response latency (i.e., a response latency k means that the identity theft is detected based on k recent continuous behaviors). Figs. 14 and 15 demonstrate the AUC values and TPRs via different response latency in each model on both datasets.

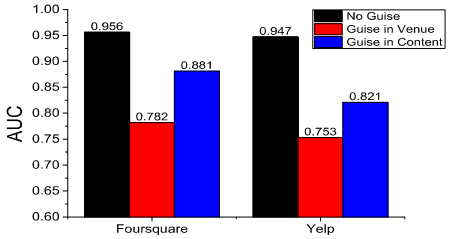


Fig. 11. Detection efficacy (AUC) via joint model in different scenarios.

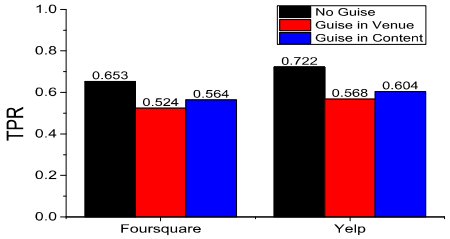


Fig. 12. Detection rate (TPR) via joint model in different scenarios with disturbance rate = 0.01.

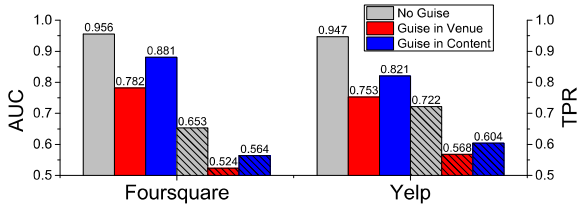


Fig. 13. Efficacy (AUC) and detection rate (TPR) of identity theft detection via the joint model CBM in different scenarios with disturbance rate = 0.01 (FPR = 0.01). Painted ones denote AUC and shaded ones denote TPR.

TABLE IX
PERFORMANCE FOR RANDOM BEHAVIOR ATTACK

	Precision	Recall	FPR	AUC
Foursquare	83.08%	93.99%	1.00%	0.995
	97.82%	86.58%	0.10%	
Yelp	82.43%	88.95%	1.00%	0.995
	97.57%	76.09%	0.10%	

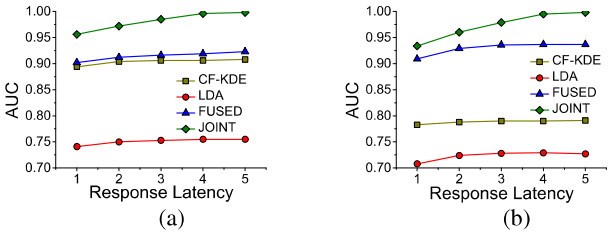


Fig. 14. Identity theft detection efficacy via different response latency (i.e., the number of behaviors in the test set we cumulated). (a) Foursquare. (b) Yelp.

The experimental results indicate that our joint model CBM is superior to all other methods. The AUC values of our joint model can reach 0.998 in both Foursquare and Yelp with five test behavioral records. The detection rates (TPRs) of our joint model can reach 93.8% in Foursquare and 97.0% in Yelp with five test behavioral records and disturbance rates (FPRs) of 1.0%.

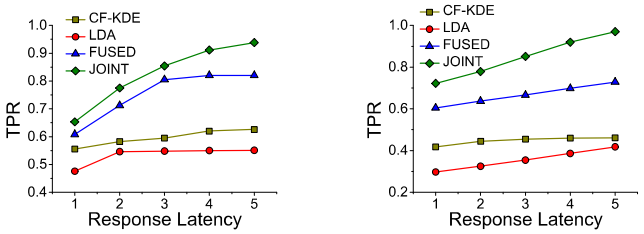


Fig. 15. Detection rates (TPR) via different response latency with disturbance rate = 0.01 (FPR = 0.01). (a) Foursquare. (b) Yelp.

V. Literature Review

To prevent and detect identity theft in online services, developers have designed various authentication methods to identify a user's identity.

Traditional password-based (username-password) authentication methods are still widely used. But the password is easy to leak, forget, and copy. Then, an authentication method adopted a physical token instead of a password [52], but it was easy to lose the token. Since the biological characteristics are hard to copy or change, more and more applications turn to utilize biometric identification technologies, such as fingerprint, face, iris, and speech recognition which are stable and not varying with time, for authentication [13], [14]. Sitova *et al.* [53] introduced hand movement, orientation, and grasp (HMOG), a set of behavioral features to continuously authenticate smartphone users. Rajoub and Zwiggelaar [15] used thermal imaging to monitor the periorbital region's thermal variations and test whether it can offer a discriminative signature for detecting deception. However, these biometric technologies usually require expensive hardware devices which makes it inconvenient and difficult to popularize.

Another drawback for these methods above is that they are redundant steps, which require users to spend extra time passing identification. Besides, they are all disposable identification measures. Once a criminal breaks through the wall, the defending system will fail to take further protection. On the contrary, when attackers facing with a continuous authentication system, they have to spend a prolonged time fooling the system. Increasingly, researchers and security experts realized that they cannot ensure users' security just by building higher and stronger digital walls around everything [16]. Thus, it is urgent to establish a non-intrusive and continuous authentication system.

Recently, researchers found that users' behavior can identify their identity and judge their personality. A study on three months of credit card records for 1.1 million people showed that four spatiotemporal points are enough to uniquely reidentify 90% of individuals [3]. Research studies found that computers outpacing humans in personality judgment presented significant opportunities and challenges in the areas of psychological assessment, marketing, and privacy [54]. Abouelenien *et al.* [30] explored a multimodal deception detection approach that relied on a novel dataset of 149 multimodal recordings, and integrated multiple physiological, linguistic, and thermal features. These works indicated that users' behavior patterns can represent their identities. Many studies turn to utilize users' behavior patterns for identifications. Behavior-based methods were born at the right moment, which plays important roles in a wide range of tasks including preventing and detecting identity theft. Typically, behavior-based user identification includes two phases: user profiling and user identifying.

User profiling is a process to characterize a user with his/her history behavioral data. Some works focus on statistical characteristics, such as the mean, variance, median, or frequency of a variable, to establish the user profile. Naini *et al.* [55] 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. But it mainly relied on experts' experience since different cases usually have different characteristics. Egele *et al.* [7] proposed a behavior-based method to identify compromises of individual high-profile accounts. However, it required high-profile accounts which were difficult to obtain.

Other researchers discovered other features, such as tracing patterns, topic and spatial distributions, to describe user identity. Ruan *et al.* [32] conducted a study on online user behavior by collecting and analyzing user clickstreams of a well-known OSN. Lesaege *et al.* [31] developed a topic model extending the LDA to identify the active users. Viswanath *et al.* [56] presented a technique based on principal component analysis (PCA) that accurately modeled the "like" behavior of normal users in Facebook and identified significant deviations from it as anomalous behaviors. Zaeem *et al.* [33] proposed an approach that involved the novel collection of online news stories and reports on the topic of identity theft. Lichman and Smyth [48] proposed MKDE model to accurately characterize and predict the spatial pattern of an individual's events. Tsikerdekis and Zeadally [57] presented a detection method based on nonverbal behavior for identity deception, which can be applied to many types of social media. These methods above mainly concentrated on a specific dimension of the composite behavior and seldom thought about utilizing multidimensional behavior data. Sekara *et al.* [58] explored the complex interaction between social and geospatial behavior and demonstrated that social behavior can be predicted with high precision. It indicated that composite behavior features can identify one's identity. Yin *et al.* [42] proposed a probabilistic generative model combining the use of spatiotemporal data and semantic information to predict user's behavior. Nilizadeh *et al.* [49] presented POISED, a system that leverages the differences in propagation between benign and malicious messages on social networks to identify spam and other unwanted content. These studies implied that composite behavior features are possibly helpful for user identification.

User identification is a process to match the same user in two datasets or distinguish anomalous users/behaviors. User identifying can be applied to a variety of tasks, such as detecting anomalous users or match users across different data sources. Mazzawi *et al.* [59] presented a novel approach for detecting malicious user activity in databases by checking user's self-consistency and global consistency. Shabtai *et al.* [60] presented a behavior-based anomaly detection system for detecting meaningful deviations in a mobile application's network to protect mobile device users and

cellular infrastructure companies from malicious applications. Lee and Kim [34] proposed a suspicious URL detection system for Twitter to detect anomalous behavior. Thomas *et al.* [9] leveraged Monarch's feature collection infrastructure to study distinctions among 11 million URLs drawn from email and Twitter. These works mainly detected population-level anomalous behaviors which indicated a strong difference to other behaviors. They did not take into account the coherence of a user's behavioral records since their works implied that anomalous accounts were created by criminals (i.e., fake accounts). Cao *et al.* [23] designed and implemented a malicious account detection system for detecting both fake and compromised real user accounts. Zhou *et al.* [61] designed an FRUI algorithm to match users among multiple OSNs. Hao *et al.* [62] proposed a novel framework for user identification in cyberphysical space.

Most of the existing related works mainly considered specific dimensions of users' behavior. Sufficient high-quality data are necessary for these works. In this study, we aim to build high-performance behavioral models based on low-quality behavioral data by integrating different dimensions of behavioral records that are usually too sparse to support qualified models. The most relevant work to our study is [17], where different dimensions of behavioral records are *fused* to build a CBM for detecting identity theft under a special pattern.

VI. Conclusion and Future Work

We investigate the feasibility of building a ladder from low-quality behavioral data to a high-performance behavioral model for user identification in OSNs. By deeply exploiting the complementary effect among OSN users' multidimensional behaviors, we propose a joint probabilistic generative model by integrating online and offline behaviors. When the designed joint model is applied to identity theft detection in OSNs, its comprehensive performance, in terms of the detection efficacy, response latency, and robustness, is validated by extensive evaluations on real-life OSN datasets. Particularly, the joint model significantly outperforms the existing fused model.

Our behavior-based method mainly aims at detecting identity thieves after the access control of the account is broken. Then, it is easy and promising to incorporate our method into traditional methods to solve the identity theft problem better.

Acknowledgment

The authors sincerely thank the anonymous referees for their helpful comments and suggestions.

References

[1] J. Onaolapo, E. Mariconti, and G. Stringhini, "What happens after you are pwnd: Understanding the use of leaked Webmail credentials in the wild," in *Proc. Internet Meas. Conf.*, Nov. 2016, pp. 65–79.

[2] A. Mohan, "A medical domain collaborative anomaly detection framework for identifying medical identity theft," in *Proc. Int. Conf. Collaboration Technol. Syst. (CTS)*, May 2014, pp. 428–435.

[3] Y.-A. de Montjoye, L. Radaelli, V. K. Singh, and A. S. Pentland, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–539, Jan. 2015.

[4] P. Hyman, "Cybercrime: It's serious, but exactly how serious?" *Commun. ACM*, vol. 56, no. 3, pp. 18–20, Mar. 2013.

[5] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: Automated identity theft attacks on social networks," in *Proc. 18th Int. Conf. World Wide Web (WWW)*, 2009, pp. 551–560.

[6] J. Lynch, "Identity theft in cyberspace: Crime control methods and their effectiveness in combating phishing attacks," *Berkeley Technol. Law J.*, vol. 20, no. 1, pp. 259–300, 2005.

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

[8] T. C. Pratt, K. Holtfreter, and M. D. Reisig, "Routine online activity and Internet fraud targeting: Extending the generality of routine activity theory," *J. Res. Crime Delinquency*, vol. 47, no. 3, pp. 267–296, Aug. 2010.

[9] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in *Proc. IEEE Symp. Secur. Privacy*, May 2011, pp. 447–462.

[10] H. Li *et al.*, "Bimodal distribution and co-bursting in review spam detection," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 1063–1072.

[11] A. M. Marshall and B. C. Tompsett, "Identity theft in an online world," *Comput. Law Secur. Rev.*, vol. 21, no. 2, pp. 128–137, Jan. 2005.

[12] B. Schneier, "Two-factor authentication: Too little, too late," *Commun. ACM*, vol. 48, no. 4, p. 136, Apr. 2005.

[13] M. V. Ruiz-Blondet, Z. Jin, and S. Laszlo, "CEREBRE: A novel method for very high accuracy event-related potential biometric identification," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 7, pp. 1618–1629, Jul. 2016.

[14] R. D. Labati, A. Genovese, E. Muñoz, V. Piuri, F. Scotti, and G. Sforza, "Biometric recognition in automated border control: A survey," *ACM Comput. Surv.*, vol. 49, no. 2, p. 24, 2016.

[15] B. A. Rajoub and R. Zwiggelaar, "Thermal facial analysis for deception detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 6, pp. 1015–1023, Jun. 2014.

[16] M. M. Waldrop, "How to hack the hackers: The human side of cybercrime," *Nature*, vol. 533, no. 7602, pp. 164–167, May 2016.

[17] C. Wang, B. Yang, J. Cui, and C. Wang, "Fusing behavioral projection models for identity theft detection in online social networks," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 4, pp. 637–648, Aug. 2019.

[18] C. Shen, Y. Li, Y. Chen, X. Guan, and R. A. Maxion, "Performance analysis of multi-motion sensor behavior for active smartphone authentication," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 1, pp. 48–62, Jan. 2018.

[19] C. Wang and H. Zhu, "Representing fine-grained co-occurrences for behavior-based fraud detection in online payment services," *IEEE Trans. Dependable Secure Comput.*, early access, May 4, 2020, doi: 10.1109/TDSC.2020.2991872.

[20] H. Zheng *et al.*, "Smoke screener or straight shooter: Detecting elite sybil attacks in user-review social networks," in *Proc. 25th Annu. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, San Diego, CA, USA, Feb. 2018, pp. 259–300.

[21] R. T. Mercuri, "Scoping identity theft," *Commun. ACM*, vol. 49, no. 5, pp. 17–21, May 2006.

[22] G. Stringhini, P. Mourlanne, G. Jacob, M. Egele, C. Kruegel, and G. Vigna, "EVILCOHORT: Detecting communities of malicious accounts on online services," in *Proc. USENIX Secur.*, 2015, pp. 563–578.

[23] Q. Cao, X. Yang, J. Yu, and C. Palow, "Uncovering large groups of active malicious accounts in online social networks," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2014, pp. 477–488.

[24] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *Proc. 26th Annu. Comput. Secur. Appl. Conf. (ACSAC)*, 2010, pp. 1–9.

[25] Y. Yao, B. Viswanath, J. Cryan, H. Zheng, and B. Y. Zhao, "Automated crowdturfing attacks and defenses in online review systems," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Dallas, TX, USA, Oct. 2017, pp. 1143–1158.

[26] C. Wang, C. Wang, H. Zhu, and J. Cui, "LAW: Learning automatic windows for online payment fraud detection," *IEEE Trans. Dependable Secure Comput.*, early access, Nov. 16, 2020, doi: 10.1109/TDSC.2020.3037784.

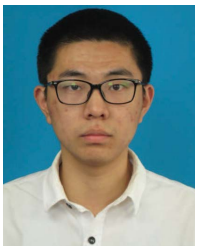
[27] F. Ahmed and M. Abulaish, "A generic statistical approach for spam detection in online social networks," *Comput. Commun.*, vol. 36, nos. 10–11, pp. 1120–1129, Jun. 2013.

[28] G. R. Milne, L. I. Labrecque, and C. Cromer, "Toward an understanding of the online consumer's risky behavior and protection practices," *J. Consum. Affairs*, vol. 43, no. 3, pp. 449–473, Sep. 2009.

[29] A. Abo-Alian, N. L. Badr, and M. F. Tolba, "Keystroke dynamics-based user authentication service for cloud computing," *Concurrency Comput., Pract. Exper.*, vol. 28, no. 9, pp. 2567–2585, 2016.
[30] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, "Detecting deceptive behavior via integration of discriminative features from multiple modalities," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 5, pp. 1042–1055, May 2017.
[31] C. Lesaege, F. Schnitzler, A. Lambert, and J.-R. Vigouroux, "Time-aware user identification with topic models," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 997–1002.
[32] X. Ruan, Z. Wu, H. Wang, and S. Jajodia, "Profiling online social behaviors for compromised account detection," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 176–187, Jan. 2016.
[33] R. N. Zaeem, M. Manoharan, Y. Yang, and K. S. Barber, "Modeling and analysis of identity threat behaviors through text mining of identity theft stories," *Comput. Secur.*, vol. 65, pp. 50–63, Mar. 2017.
[34] S. Lee and J. Kim, "WarningBird: Detecting suspicious URLs in Twitter stream," in *Proc. 19th Annu. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, San Diego, CA, USA, Feb. 2012, pp. 1–13.
[35] H. Li, Y. Ge, R. Hong, and H. Zhu, "Point-of-interest recommendations: Learning potential check-ins from friends," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 975–984.
[36] C. Shen, Y. Chen, X. Guan, and R. A. Maxion, "Pattern-growth based mining mouse-interaction behavior for an active user authentication system," *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 2, pp. 335–349, Mar. 2020, doi: 10.1109/TDSC.2017.2771295.
[37] C. Shen, Y. Zhang, X. Guan, and R. A. Maxion, "Performance analysis of touch-interaction behavior for active smartphone authentication," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 3, pp. 498–513, Mar. 2016.
[38] N. Hernandez, M. Rahman, R. Recabarren, and B. Carbunar, "Fraud de-anonymization for fun and profit," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Toronto, ON, Canada, Oct. 2018, pp. 115–130.
[39] C. Wang, J. Zhou, and B. Yang, "From footprint to friendship: Modeling user followership in mobile social networks from check-in data," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 825–828.
[40] T. M. Cover and J. A. Thomas, "Entropy, relative entropy and mutual information," *Elements Inf. Theory*, vol. 2, no. 1, pp. 12–13, 1991.
[41] C. Song, Z. Qu, N. Blumm, and A. L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, p. 1018, 2010.
[42] H. Yin *et al.*, "Discovering interpretable geo-social communities for user behavior prediction," in *Proc. IEEE 32nd Int. Conf. Data Eng. (ICDE)*, May 2016, pp. 942–953.
[43] J. Bao, Y. Zheng, and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in *Proc. 20th Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL)*, 2012, pp. 199–208.
[44] S. Kc and A. Mukherjee, "On the temporal dynamics of opinion spamming: Case studies on yelp," in *Proc. 25th Int. Conf. World Wide Web*, Apr. 2016, pp. 369–379.
[45] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[46] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proc. ACM SIGKDD*, 2014, pp. 25–34.
[47] W. X. Zhao *et al.*, "Comparing Twitter and traditional media using topic models," in *Proc. ECIR*, 2011, pp. 338–349.
[48] M. Lichman and P. Smyth, "Modeling human location data with mixtures of kernel densities," in *Proc. ACM SIGKDD*, 2014, pp. 35–44.
[49] S. Nilizadeh *et al.*, "POISED: Spotting Twitter spam off the beaten paths," in *Proc. ACM CCS*, 2017, pp. 1159–1174.
[50] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[51] E. Bursztein *et al.*, "Handcrafted fraud and extortion: Manual account hijacking in the wild," in *Proc. ACM IMC*, 2014, pp. 347–358.
[52] E. Dauterman, H. Corrigan-Gibbs, D. Mazières, D. Boneh, and D. Rizzo, "True2F: Backdoor-resistant authentication tokens," in *Proc. IEEE Symp. Secur. Privacy (SP)*, San Francisco, CA, USA, May 2019, pp. 398–416.
[53] Z. Sitová *et al.*, "HMOG: New behavioral biometric features for continuous authentication of smartphone users," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 5, pp. 877–892, May 2016.
[54] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 4, pp. 1036–1040, 2015.
[55] F. M. Naini, J. Unnikrishnan, P. Thiran, and M. Vetterli, "Where you are is who you are: User identification by matching statistics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 2, pp. 358–372, Feb. 2016.
[56] B. Viswanath *et al.*, "Towards detecting anomalous user behavior in online social networks," in *Proc. USENIX Secur.*, 2014, pp. 223–238.
[57] M. Tsikerdekis and S. Zeadally, "Multiple account identity deception detection in social media using nonverbal behavior," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 8, pp. 1311–1321, Aug. 2014.
[58] V. Sekara, A. Stopczynski, and S. Lehmann, "Fundamental structures of dynamic social networks," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 36, pp. 9977–9982, 2016.
[59] H. Mazzawi, G. Dalal, D. Rozenblatz, L. Ein-Dorx, M. Niniox, and O. Lavi, "Anomaly detection in large databases using behavioral patterning," in *Proc. IEEE 33rd Int. Conf. Data Eng. (ICDE)*, Apr. 2017, pp. 1140–1149.
[60] A. Shabtai, L. Tenenboim-Chekina, D. Mimran, L. Rokach, B. Shapira, and Y. Elovici, "Mobile malware detection through analysis of deviations in application network behavior," *Comput. Secur.*, vol. 43, pp. 1–18, Jun. 2014.
[61] X. Zhou, X. Liang, H. Zhang, and Y. Ma, "Cross-platform identification of anonymous identical users in multiple social media networks," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 411–424, Feb. 2016.
[62] T. Hao, J. Zhou, Y. Cheng, L. Huang, and H. Wu, "User identification in cyber-physical space: A case study on mobile query logs and trajectories," in *Proc. ACM SIGSPATIAL*, 2016, pp. 71:1–71:4.

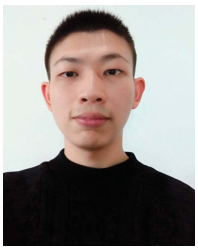
Cheng Wang (Senior Member, IEEE) received the Ph.D. degree from the Department of Computer Science, Tongji University, Shanghai, China, in 2011.

His research area is network information service. His research interests include cyberspace security, mobile social networks, topic modeling, and knowledge graph.



Hangyu Zhu received the master's degree from the Department of Computer Science and Technology, Tongji University, Shanghai, China, in March, 2021, where he is currently pursuing the Ph.D. degree with the Department of Computer Science.

His research interests include anomaly detection, behavior modeling, and network representation learning.



Bo Yang received the B.S. degree from the School of Mathematical Sciences, Tongji University, Shanghai, China, in 2014, where he is currently pursuing the master's degree with the Department of Computer Science.

His research interests include behavior modeling and cybersecurity.