


Received January 24, 2018, accepted March 3, 2018, date of publication March 22, 2018, date of current version April 23, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2818116

# Modeling Privacy Leakage Risks in Large-Scale Social Networks

SUGUO DU<sup>1</sup>, XIAOLONG LI<sup>1</sup>, JINLI ZHONG<sup>1</sup>, LU ZHOU<sup>1</sup>, MINHUI XUE<sup>2</sup>,  
HAOJIN ZHU<sup>1</sup> , (Senior Member, IEEE), AND LIMIN SUN<sup>3,4</sup>, (Member, IEEE)

<sup>1</sup>Shanghai Jiao Tong University, Shanghai 200000, China

<sup>2</sup>New York University Shanghai, Shanghai 200122, China

<sup>3</sup>School of Cyber Security, University of Chinese Academy of Sciences, Beijing 101408, China

<sup>4</sup>Beijing Key Laboratory of IOT Information Security Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100864, China

Corresponding author: Haojin Zhu (zhuhaojin@gmail.com)

This work was supported by the National Natural Science Foundation of China under Grant 71671114, Grant U1405251, Grant U1401253, and Grant 61472418.

**ABSTRACT** The current culture that encourages online dating, and interaction makes large-scale social network users vulnerable to miscellaneous personal identifiable information leakage. To this end, we take a first step toward modeling privacy leakages in large-scale social networks from both technical and economic perspectives. From a technical perspective, we use Markov chain to propose a dynamic attack-defense tree-based model, which is temporal-aware, to characterize an attack effort made by an attacker and a corresponding countermeasure responded by a social network security defender. From an economic perspective, we use static game theory to analyze the ultimate strategies taken by the attacker and the defender, where both rational participants tend to maximize their utilities, with respect to their attack/defense costs. To validate the proposed approach, we perform extensive experimental evaluations on three real-world data sets, triggered by the survey of over 300 volunteers involved, which illuminates the privacy risk management of contemporary social network service providers.

**INDEX TERMS** Social network services, Data privacy, Information security.

## I. INTRODUCTION

Social networks including online social networks and mobile social networks are extremely popular nowadays. The latest statistics show that the number of active social media users has exceeded 3.28 billion [1]. Along with overwhelming popularity of social networks, the privacy leakage issues are posing a serious threat to the security of large-scale social networks. In 2016, more than 200 mobile applications and website were found to leak the sensitive consumer information by analyzing about four billion requests [2]. According to a recent research, more than 6.05 billion personal information has been disclosed in China [3].

Privacy issues in large-scale social networks are gaining an increasing attention recently due to the following reasons. Firstly, there is a strong economic motivation for the social network to collect the users' interests, habits, demographic information (e.g., race, economic status, sex, age, the level of education, income level, etc.) and even the online behavior history. This personal information is intentionally collected by the social network for *targeted advertising*,

which uses sophisticated machine learning/recommendation algorithms to target the most receptive audiences with certain traits, based on the product or person the advertiser is promoting. Secondly, the social network platform may be vulnerable to the various attacks. For example, in 2013, it was reported that the Facebook bug leaked the private contact information of 6 million users [4]. Another issue is that the social network platform may be reluctant to perform the privacy enhancement technology, such as “do not track” (DNT) option, which was proposed by FTC to enable people to avoid having their actions monitored online. In 2017, Twitter publicly announced that it is abandoning DNT privacy protection standard [5]. Thirdly, the compromise of the user privacy may be caused by the mis-operation of the users. The existing research pointed out that it is possible to infer 39.9% more personal information via de-anonymization and aggregation from multiple social networks [6], [7]. Based on the above challenges, it is highly desirable to model the privacy leakage in large-scale social networks.

The existing research on privacy issues in social networks mainly includes the preventative defense techniques [8] [9]. From a system point of view, it lacks a comprehensive yet well-defined security evaluation to allow the system administrator to identify the most critical privacy leaking threats and thus determine the appropriate defense strategy, which are more than important for the overall privacy preservation of the social networks. The existing risk analysis schemes in social network include HMM inference model [10], sensitivity analysis [11] based solutions. However, neither of them could be utilized to model the risk of the whole system.

In particular, modeling the privacy leakage risks in social networks should face the following research challenges. Firstly, the privacy defense strategy is directly correlated to how the privacy compromising attack is launched, which means that the security evaluation should consider both of the attack and the defense sides rather than any single one. Secondly, since the privacy related attack is a dynamic process, each attack step (or state) and the transitions among different states are highly important. Therefore, it is highly desirable to introduce a new approach to modeling the transitions among different attack/defense states, which can well capture the strategies of the attackers and defenders. Thirdly, most of the available solutions only consider the technological issues instead of the economical ones, which is of equal importance indeed. Lastly, but no less importantly, most of the existing security solutions only consider how to prevent an attack while fail to take the costs and gains into consideration. In practice, a rational attacker or defender may try to maximize its attack or defense benefits instead of blindly launching an attack or adopting a countermeasure.

In this study, we introduce a novel approach to model and analyze the privacy leakage issues in large-scale social networks. The proposed approach adopts attack-defense tree to describe a series of attack steps launched by the attackers to achieve their ultimate goals and the corresponding countermeasures that can be adopted by the social network security defenders. To further illustrate a dynamic attacking process, we propose a Markov chain based approach to model a temporal-aware attack-defense tree. Lastly, to analyze the potential strategies performed by the attacker and the defender, we introduce an attack-defense game, in which each rational participant tends to get the maximum utility based on the different utility, attack/defense strategies and the associated attack/defense cost. We performed extensive evaluations based on 3 real-world datasets involving 62, 699 users and the collected questionnaire survey of 304 volunteers.

To the best of our knowledge, this paper is the first to model the potential risk as well as attack/defense strategies in social networks. The main contributions of this paper are listed as follows:

- 1) We adopt the attack-defense tree based risk analysis model to describe attack/defense strategy of the attacker and the defender. The built attack-defense tree gives a comprehensive review on the reported security solutions.

- 2) We introduce a discrete Markov chain based model to abstract the logical and temporal order of the state transition of attack and defense.
- 3) We introduce a novel attack-defense game to model the interaction between the attacker and the defender, both of which may try to maximize their benefits. We model the attack-defense game as a static game and give a detailed analysis on its Nash Equilibrium.
- 4) We perform extensive evaluations based on 4 real-world datasets.

The reminder of this paper is as follows. Section II presents the related work. Section III proposes privacy risk evaluation model combining attack-defense tree, Markov chain and game theory. Section IV applies the model to real-world datasets. Finally, Section V gives our conclusion.

## II. RELATED WORK

There is an increasing interest to study how to de-anonymize or re-identify users across social networks, which mainly falls to the following two categories: profile based de-anonymization and structured based de-anonymization. Structure based de-anonymization works are based on the assumption that the different social networks of the same group users should show the similar network topology, which can be exploited for user identification [12], [13]. Profile based de-anonymization leverages the public information and semantic information on social media or social network sites to match users of different social networks. Iofciu *et al.* used tags to identify users across social tagging systems such as Delicious, StumbleUpon and Flickr [14]. Lai *et al.* [15] proposed to detect communities in social networks via users interests and de-anonymize users in communities. Li *et al.* [6] propose NHDS, which aims at de-anonymizing heterogeneous social networks by leveraging the network graph structure to significantly reduce the size of candidate set, and exploiting user profile information to identify the correct mapping users with a high confidence.

In addition, location privacy protection in location-based services is a long-standing topic [16]–[21]. The most popular approach to achieve location privacy in social networks is utilizing obfuscation techniques to coarse the spatial or temporal granularity of real locations [22]–[24]. But the service utility and the privacy protection are always a trade-off. In [25], it investigates the location privacy leakage issues in popular mobile social networks such as WeChat, Momo and Skout. The similar problem is also pointed out to exist in Facebook in [26].

Different from any previous works, this work mainly investigates how to model and measure the privacy risks in social networks.

## III. PRIVACY RISK MODELING IN SOCIAL NETWORKS

In this section, we propose a privacy risk assessment model of large scale social networks, which is comprised of the attack-defense tree model, Markov Chain based dynamic model analysis, as well as the attack-defense game based strategy analysis.

## A. PRIVACY RISK ASSESSMENT BASED ON ATTACK-DEFENSE TREE MODEL

### 1) INTRODUCTION TO ATTACK-DEFENSE TREE MODEL: GOAL-ORIENTED PERSPECTIVE

In this subsection, we will first give a brief introduction on attack-defense tree model. In the real-world attack event, privacy leakage cannot be achieved in a single step. Thus, it is rational to summarize all possible routes which can lead to privacy leakage and analyze their impact on the social network.

The attack tree model is a suitable method to achieve the requirements above, which describes the attack process step by step, from atomic attack event to the ultimate goal. In general, an attack tree model offers a goal-oriented perspective that facilitates the expression of multi-stage attacks [27]. The root node is the attack goal which represents the final objective of the attacker. The leaf node (or atomic attack) is a single step adopted by the attacker. During the process, the attacker uses different atomic attacks to achieve the sub-goals, which eventually achieve the ultimate goal of revealing users' privacy. Furthermore, logic gates (normally OR and AND gates) are utilized to demonstrate the relationship among attack events. For "AND gate", goal happens only if all events under the gate are completed. Whilst for "OR gate", completing any event under the gate would achieve the goal. Once a tree is established, each atomic attack event (namely the leaf node) would be assigned a corresponding value representing its success rate, and the root (ultimate goal) value would represent the security level of the whole system.

The attack tree model has been widely used since it is easy to understand its hierarchical structure and convenient for quantitative calculation. However, it is highly desirable to jointly consider both of the defense and attack strategies as a whole system. Clearly, social network operators have considered the security and privacy issues and implemented some defenses already. Thus, in this paper, we extend the attack tree into attack-defense tree so that both roles would be considered together with their interactions, which will be modeled as the game theory model in Section III.C as well.

### 2) BUILDING ATTACK-DEFENSE TREE FOR SOCIAL NETWORK PRIVACY

In this section, we focus on how to establish the attack-defense tree for social networks based on Table 1. The first thing is to identify the ultimate goal of the attack, namely, "Social network privacy disclosure" ( $G$ ). Unlike some traditional websites, social network has its specific characteristic, which suffers from two privacy leakage ways, the privacy leakage from operator side ( $M_1$ ) and the privacy leakage from the user side ( $M_2$ ). These two sub-goals are connected with "OR gate". By this way, we create the tree in a top-down manner. Finally, we add countermeasures related to sub-goals and turn it into the "attack-defense tree" (see Fig. 1).

TABLE 1. Atomic events' indexes of attacking difficulty.

	Indexes of attacking difficulty
$B_1$ : Bypass security scanning	1. Is application downloaded from trusted app store? 2. Is anti-virus software installed in terminal devices? 3. Is security scanning regularly done?
$B_2$ : Obtain terminal access	1. Will the app inform user of access rights? 2. Will user prohibits some access rights? 3. Will user stop installing if requests excessive rights?
$B_3$ : Get private information	1. Is private information stored in mobile device? 2. Is private information confidential?
$B_4$ : Collect user information	1. Is user information required in registration? 2. Will user write actual personal information?
$B_5$ : Platform accidental disclosure	1. Is the security level of social platform high? 2. Are security vulnerabilities regularly found and repaired? 3. Are platform programmers diligent?
$B_6$ : Obtain access permissions	1. Are friends the only ones with access authority? 2. Will user encrypt personal information? 3. Will user add strangers as friends?
$B_7$ : Analyze online information	1. Will user publish extremely private information? 2. Does the released information has certain regularity? 3. Is the user active?
$B_8$ : Compose personal information	1. Is there many interactions between victim and friends? 2. Will user pay much attention on private information?

For the left subtree  $M_1$ , we further divide it into two subtrees, namely, "Malware Attack" ( $M_3$ ) and "Social website leakage" ( $M_4$ ). The former refers to malicious social applications that are intended to obtain private information. To achieve such a goal, the attacker needs to take three subsequent actions: "Bypassing security scanning" ( $B_1$ ), "Obtaining terminal access" ( $B_2$ ) as well as "Getting private information" ( $B_3$ ). With malicious programs running in such applications, it has to bypass security scanning and detection. Otherwise, the installation would fail. Bypassing anti-virus software is included in  $B_1$  as well. During the installation, the malware needs to gain access to the terminal device. Otherwise, private information will be directly blocked from the malware. Finally, when gaining access to the raw data, private information would be leaked only if no encryption measure is conducted. As for the latter  $M_4$ , it refers to the privacy leakage when the social network platform is attacked or suffers from vulnerabilities. Usually, such a thing would happen when the platform "Collects user information" ( $B_4$ ) and suffers from "Platform accidental disclosure" ( $B_5$ ). The former means that the users upload a part of their private information to the platform. Thus when the platform is under an attack, the users' information is at risk as well.

As for the right subtree  $M_2$ , either "Individual leakage" ( $M_5$ ) or "Friends leakage" ( $M_6$ ) would make it happen. The former refers to compromising privacy due to user's own negligence such as regularly publishing his own locations. If the user doesn't set any access control on account's permission and is willing to publish his information on social networks, the attacker can easily compromise the user's privacy by launching the attack of "Gaining accounts' permission" ( $B_6$ ) and "Analyzing the Online Information" ( $B_7$ ). Similarly, for the latter  $M_6$ , the attacker is required to "Gain accounts' permission" ( $B_6$ ) from victim's friend and "Aggregate personal information" ( $B_8$ ).

In practice, the defense strategies are deployed on both of the operator side and the user side. So we establish the defense

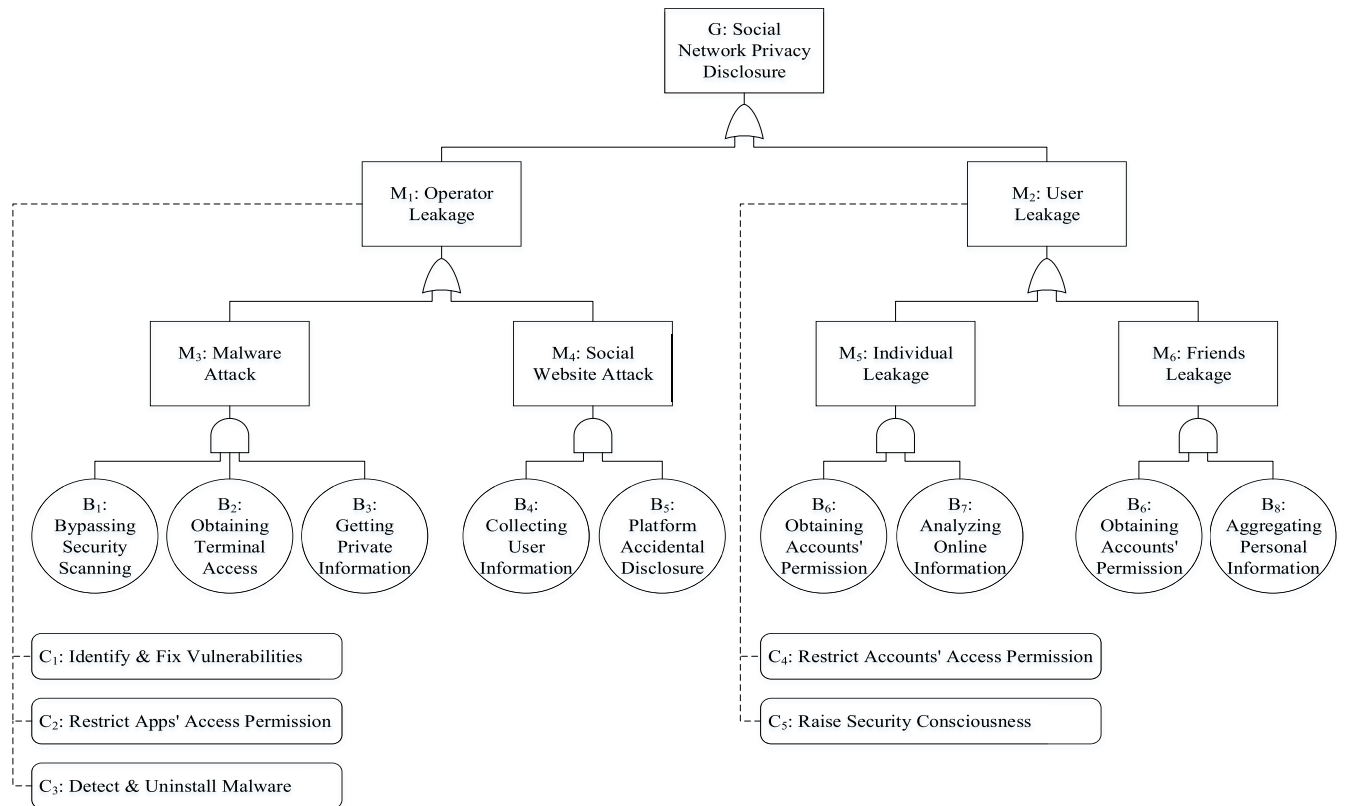


FIGURE 1. The attack-and-defense tree model of mobile social network privacy disclosure.

nodes corresponding to the sub-goals ( $M_1, M_2$ ) as follows. For the simplicity of the presentation, we omit some regular countermeasure such as “Regular maintenance”, “Technical updates”, which could prevent  $M_4$  to some extent as well.

- $C_1$ : “Identify & fix vulnerabilities”
- $C_2$ : “Restrict apps’ access permission”
- $C_3$ : “Detect & uninstall malware”
- $C_4$ : “Restrict accounts’ access permission”
- $C_5$ : “Raise security consciousness”

Based on the analysis above, it is observed that the atomic events should be in order in this attack-defense tree model. Here, we use  $\rightarrow$  to represent the order which is from the left to the right. In addition, we use lowercase letters to represent the success of the atomic attack and the uppercase to represent the opposite. Therefore, we could intuitively find the four attack routes, namely  $b_1 \rightarrow b_2 \rightarrow b_3, b_4 \rightarrow b_5, b_6 \rightarrow b_7,$  and  $b_6 \rightarrow b_8$ . However, the attack-defense tree could not reflect such an order. This motivates us to adopt the Markov model to capture the interactions between the attacker and the defender.

**B. MODELING A DYNAMIC SYSTEM VIA MARKOV MODEL**

1) INTRODUCTION TO MARKOV MODEL

Originally, Markov chain is used to describe the transition property of a random process. For random variables  $\{X_n, n = 0, 1, 2, \dots\}$  and its state space  $S = \{S_i, i = 0, 1, 2, \dots\},$

$X_n = S_i$  represents staying  $S_i$  at time  $n$ . The theory shows that the next state depends completely on the current state, and is unrelated to any of the previous ones. Thus the probability function would be

$$P\{X_{n+1} = j | X_n = i, X_{n-1}, \dots, X_1, X_0\} = P\{X_{n+1} = j | X_n = i\} = P_{ij} \tag{1}$$

Here  $P$  represents the transition matrix where  $P_{ij}(i \neq j)$  refers to the transition probability from  $S_i$  to  $S_j$  and  $P_{ii}$  refers to the probability of maintaining  $S_i$ . Given that  $|S| = m,$  the transition distribution can be represented as a  $m \times m$  transition matrix.

Given that the vector  $\pi$  describes the probability when the system reaches the balance, where  $\pi_i$  represents the probability of each state, then  $\pi$  must satisfy the following prerequisites.

$$\lim_{n \rightarrow \infty} \pi \cdot P^n = \pi \tag{2}$$

$$\sum_i \pi_i = 1 \tag{3}$$

Because of the random choices of attackers and defenders, the next state of users’ privacy leakage system is entirely decided by the current state rather than the future state, which is consistent with the assumption of Markov Chain. To further simplify the calculation, we have the following assumptions.

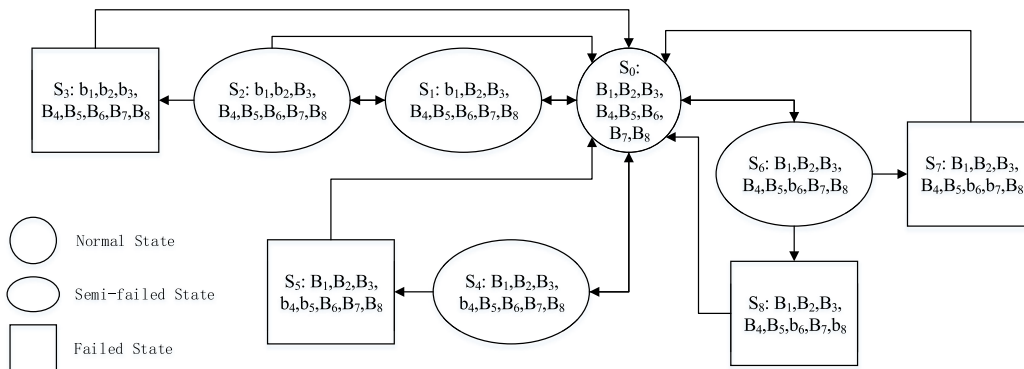


FIGURE 2. The state transition graph of mobile social network privacy disclosure.

- 1) Since the success rate of the atomic event is independent of each other, whatever the current state is, the transition probability would be the same. Otherwise, we could utilize the piecewise function to describe the difference since the state space is finite.
- 2) The attacker will not change the sub-goal halfway. Whenever he selects a sub-goal, he will not give up halfway. In other words, the atomic event is either achieved or failed, which is in line with discrete Markov model.
- 3) We consider state transition caused by only one atomic event since any other transition could be expressed with this method. For instance, transition caused by two events could be equally described as two adjacent transitions with the same probability.
- 4) Both of the attacker and the defender are assumed to understand not only the difficulty of attacks, but also the difficulty of defenses.
- 5) It is assumed that the attack and defense cannot happen simultaneously.
- 6) When the system reaches the failed state, the defender can always reboot the system.
- 7) We ignore self-loops of each state.

2) TRANSITION DIAGRAM PRODUCTION

Based on these assumptions, we describe systematic state as a combination constituted by the state of each atomic event. And it is divided into three categories, namely failed states, semi-failed states and normal states. The original situation of users' privacy is called the normal state. And once one of the above four attack routes succeeds, the system reaches failed state. Otherwise, it is semi-failed state. Finally, we have failed state  $\{S_0\}$ , semi-failed states  $\{S_1, S_2, S_4, S_6\}$ , and failed states  $\{S_3, S_5, S_7, S_8\}$ . The transition diagram is established as Figure 2.

3) TRANSITION PROBABILITY CALCULATION

Without loss of generality, we present a case study based on the collected questionnaire, which will be introduced in details in Section IV. Based on the Table 1 as well as the data from the questionnaire, we could transfer the textual

TABLE 2. Attacking and repairing technical difficulty of atomic events.

	$D_A$	$D_R$
$B_1$	2.88	2.12
$B_2$	3.70	1.30
$B_3$	2.69	2.31
$B_4$	1.92	3.08
$B_5$	4.00	1.00
$B_6$	2.97	2.03
$B_7$	2.85	2.15
$B_8$	3.25	1.75

information into numerical value with the assistance of Table 8 so that the difficulty of each atomic event ( $D_A$ ) could be calculated. For example, question 3,4 and 6 in the questionnaire are related to  $B_1$ . Thus we calculate numerical value of each question and take the average as the difficulty of  $B_1$  where numbers between 1 and 5 are used to represent difficulty levels. As for the defense difficulty, we consider the fact that the more effective the attack is, the weaker the defense implies. Otherwise, the defense mechanism would protect the system from being attacked to a great extent and under no circumstance will the attack be called effective. Thus, to simplify the calculation, we assume that the sum of success rate of any attack  $A$  and its corresponding repairing method  $R$  would be 5. Following this manner, we could obtain the difficulty of each atomic event as shown in Table 2.

Then we assume that the higher the success rate of an attack is, the lower transition probability related to the attack will be. Suppose that the current state is  $S_i$ ,  $Q_i$  represents all of the next states that the system may transit to.  $S_j \in Q_i$  refers to the most difficult attack, which implies that the transition probability  $P_{ij}$  would be the smallest if the attacker is rational. Similarly, a stronger defending strategy leads to a lower transition probability. Therefore, the transition frequency  $F_{ij}$  and the transition probability  $P_{ij}$  is as follows:

$$F_{ij} = \begin{cases} \frac{1}{D_A^u} & S_i \xrightarrow{\text{attack } u} S_j \\ \frac{1}{D_R^u} & S_i \xrightarrow{\text{defense } u} S_j \end{cases} \quad (4)$$

$$P_{ij} = \frac{F_{ij}}{\sum_{k \in Q_i} F_{ik}} \quad (5)$$

The generated matrix satisfies the requirement of Markov matrix that every element is no less than 0 and the sum of each row or each column is equal to 1, since the success rate of atomic event could never be less than 0 and  $P_{ij}$  is generated via frequency normalization. With the data from Table 2, the transition matrix is shown in the bottom of this page.

**C. STRATEGIES ANALYSIS BASED ON GAME THEORY**

Besides technical issues, the economic factors have an equally important impact on the behavior of the attackers and defenders. Different attack strategies will bring different benefits to the attacker with different costs. It is the same for the defender. Therefore, we introduce a game theory model to consider these economic factors and construct proper utility functions to reflect the effectiveness of the behaviors. Furthermore, we will discuss the dominant strategies under different conditions, which is expected to provide good suggestions for the defenders.

**1) GAME THEORY MODEL CONSTRUCTION**

Generally speaking, rational attackers and defenders endeavor to maximize their returns. However, increasing the benefit of either side would reduce that of the other, so that the two sides would never reach the maximum simultaneously. This feature is quite accordant with game theory where players maximize their utility functions determined by the cost and benefit. Therefore, we model it as a single-stage static game since both sides cannot know the other’s strategy before the action. Then we analyze the offensive and defensive behaviors of the two sides under different conditions by obtaining the Nash Equilibrium [28]. Here define the three key elements ( $O; S; U$ ) of the game theory model  $G$  as follows:

- 1) Participant: The participant set  $O$  is defined as  $O = \{O_1 : Attacker, O_2 : Defender\}$  where either of them does not know the other’s choice.
- 2) Strategy: The defense countermeasures are defined as  $\{C_i | i = 1, \dots, 5\}$  and the attack methods are defined as  $\{M_j | j = 3, \dots, 6\}$ , which are exactly the same as previous ones.
- 3) Utility function: The utility function is  $u_1(C_i, M_j) = ROI(C_i, M_j)$  for the defender, and it is  $u_2(C_i, M_j) = ROA(C_i, M_j)$  for the attacker.

**TABLE 3. Normalized form of single-phased security game theory model.**

Counter-measures	Attack			
	$M_3$		$\dots$	$M_6$
$C_1$	$ROI(C_1, M_3), ROA(C_1, M_3)$	$\dots$	$ROI(C_1, M_6), ROA(C_1, M_6)$	
$\dots$	$\dots$	$\dots$	$\dots$	
$C_5$	$ROI(C_5, M_3), ROA(C_5, M_3)$	$\dots$	$ROI(C_5, M_6), ROA(C_5, M_6)$	

According to the above definition, we could model the attack and defense on social network privacy as the single-stage static game, as described in Table 3.

Both the attackers and defenders are supposed in a complete information game since they basically know the strategies that the other side possesses. Then we assume that the players tend to select strategies with a certain probability in the long term, namely, the mixed strategy. Otherwise, if it is a pure strategy where the players choose one for all, then it could be represented with a probability 100%. For the defender, we define the probability distribution as  $p_C = (p_1, \dots, p_5)$ , where  $p_i \geq 0 (i = 1, \dots, 5)$  and  $\sum_{i=1}^5 p_i = 1$ . Similarly, for the attacker, it is  $q_M = (q_3, \dots, q_6)$  where  $q_j \geq 0 (j = 3, \dots, 6)$  and  $\sum_{j=3}^6 q_j = 1$ . Under the mixed strategy, the utility functions of both sides are expressed as follows:

$$u_1(S_{p_C}, S_{q_M}) = \sum_{i=1}^5 \sum_{j=3}^6 p_i \cdot q_j \cdot ROI(C_i, M_j) \tag{6}$$

$$u_2(S_{p_C}, S_{q_M}) = \sum_{i=1}^5 \sum_{j=3}^6 p_i \cdot q_j \cdot ROA(C_i, M_j) \tag{7}$$

There are two types of costs in attack-defense tree, including the attack cost and the defense cost [29]. We introduce  $ROA$  (Return of Attack) and  $ROI$  (Return of Investment) to measure the effectiveness of the costs. For  $ROI$ , it is the expected rate of return when conducting a certain countermeasure with a certain cost. Then, we define  $ALE$  as the annual expected loss if suffering an attack,  $RM$  as risk mitigation rate, and  $CI$  as the cost of investment. By using  $R_D = ALE/CI$  to represent the ratio of profit to cost for the defender, we have

$$ROI = R_D \times RM - 1 \tag{8}$$

Further, we define  $ROA$  as the expected rate of return when conducting a specific attack with a certain cost. Then, we define  $GE$  as the expected gain if achieving the attack,

0	0.288	0	0	0.432	0	0.280	0	0
0.636	0	0.364	0	0	0	0	0	0
0.292	0.477	0	0.231	0	0	0	0	0
1	0	0	0	0	0	0	0	0
0.565	0	0	0	0	0.435	0	0	0
1	0	0	0	0	0	0	0	0
0.428	0	0	0	0	0	0	0.305	0.267
1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0

**TABLE 4.** Gain matrix of the security game.

	$M_3$	$M_4$	$M_5$	$M_6$
$C_1$	-1, 0.33	-0.0625, -0.8438	-1, 0.5	-1, -0.5
$C_2$	-0.625, -0.4	-1, 0.25	-1, 0.5	-1, -0.5
$C_3$	-0.4375, -0.8333	-1, 0.25	-1, 0.5	-1, -0.5
$C_4$	-1, 0.33	-1, 0.25	-0.5, -0.25	-0.8333, -0.875
$C_5$	-0.25, -0.8	-0.1667, -0.5833	-0.25, -0.8125	-0.8333, -0.875

**TABLE 5.** Reduction gain matrix of the security game.

	$M_3$	$M_4$
$C_1$	-1, 0.33	-0.0625, -0.8438
$C_5$	-0.25, -0.8	-0.1667, -0.5833

$Cost_A$  as the cost to conduct such attack, and  $Cost_{AC}$  as the investment the defender would increase in the presence of such an attack. By using  $R_A = GE / (Cost_A + Cost_{AC})$  to represent the ratio of profit to cost for the attacker, we have

$$ROA = R_A \times (1 - RM) - 1 \quad (9)$$

According to the above definition, we empirically preset a series of level values for reference as listed in Table 9 and Table 10, where 0, 2, 4, 6, 8 are set for  $ALE$ ,  $CI$ ,  $GE$ ,  $Cost_A$  and  $Cost_{AC}$ , and 0, 0.25, 0.5, 0.75, 1 are set for  $RM$  as the risk coping factor. Then we could pre-evaluate the  $ROI$  of each protection strategy and the  $ROA$  of each attack strategy. For ‘‘Attack’’ items, we use sub-target  $M_3$  instead of leaf nodes  $B_1$ ,  $B_2$  and  $B_3$  for the convenience. It is because ‘‘AND gate’’ requires achieving all the leaf nodes otherwise the attacker gain cannot be achieved.

## 2) NASH EQUILIBRIUM CALCULATION

Based on the previous section, we can establish the gain matrix of the security game for both attacker and defender, as shown in Table 4. Here we introduce the ‘‘weakly dominant strategy’’ to further simplify the game theory model. If  $S'_1$  is a weakly dominant strategy for player 1, then for any strategy  $S_2$  of player 2,  $u_1(S'_1, S_2) \geq u_1(S''_1, S_2)$ , where  $S''_1$  refers to some choice for player 1, which means  $S''_1$  is dominated by  $S'_1$ . In this case,  $C_2$ ,  $C_3$ ,  $C_4$  are dominated by  $C_5$ , and  $M_5$  is dominated by  $M_6$ . Since the probability of  $M_5$  is negative, it is omitted as well. Then the simplified gain matrix is shown in Table 5.

Based on the game theory, we solve mixed strategies with the following equations.

$$\begin{cases} 0.33p_1 - 0.8p_5 = -0.8438p_1 - 0.5833p_5 \\ p_1 + p_5 = 1 \end{cases} \quad (10)$$

$$\begin{cases} -q_3 - 0.0625q_4 = -0.25q_3 - 0.1667q_4 \\ q_3 + q_4 = 1 \end{cases} \quad (11)$$

Then we have the results

$$\begin{cases} p_1 = 0.1558, p_5 = 0.8442 \\ q_3 = 0.122, q_4 = 0.878 \end{cases} \quad (12)$$

From the economic point of view, the attacker would choose  $M_4$  with the highest probability of 0.878 while the

defender would choose  $C_5$  with the probability of 0.8442. This reveals the characteristic of the countermeasure that is compared with other protective strategies, which implies that raising privacy awareness would be an effective way to reduce the privacy leakage risk. For instance, privacy consciousness would reduce the possibility to download malicious software, as well as publishing less private information on the social network. On the contrary, taking  $M_5$  or  $M_6$  as the attack strategy has a pretty low profit when the defender chooses  $C_5$ , which is not preferred by the attackers. Regardless of the high cost, attacking social websites is preferred since the attacker would obtain the high returns once succeeds.

## IV. EVALUATION AND DISCUSSIONS

based on the above evaluation model consisting of attack-defense tree, Markov chain and game theory, we conduct the evaluation on real-world datasets. We analyze how the users’ factors, difficulty of attacks/defense, and economic factors influence the privacy risk, which provide a practical guidance on users’ behavior in social networks. Two types of dataset are utilized in this paper. One is gathered through an online questionnaire survey while another is three real-world mobile social networks including: Facebook, Twitter and Myspace [30]. The former is utilized to understand people’s attitude towards privacy issues in the social network, and the latter reflects their actual behavior in real life.

### A. DATASET

- *Collected questionnaires*: The dataset contains 304 valid questionnaires completed by volunteers younger than 40 years old, where those younger than 20 years old and 20 to 30 years old respectively account for 10.20% and 87.17%. From another perspective, the proportion of male/female users is 49.34%/50.66%, or 42.43% are working while 57.57% are studying.
- *Facebook Dataset*: Facebook is an online social media and networking service, where users could share links, photos, and videos, post status updates, and exchange messages. The dataset consists of profiles of 24, 507 users such as gender, location, age group, etc.
- *Twitter Dataset*: Twitter is an online news and social networking service where users interact with messages. The dataset includes 28, 199 users’ profiles.
- *Myspace Dataset*: Myspace is a social networking website offering an interactive, user-submitted network of friends, personal profiles, blogs, groups, photos, music, and videos. The dataset consists of 9, 993 individuals’ profiles.

**TABLE 6.** comparison between categories.

	<20 years old	20-30 years old	Z
Not install any anti-virus software	19.35%	35.47%	-2.10
Do security scanning less than once a week	29.03%	47.55%	-2.13
Always concern for access rights	6.45%	23.77%	-3.38
	Male	Female	Z
Not install any anti-virus software	40.00%	26.62%	2.50
Do security scanning less than once a week	54.00%	38.31%	2.78
Never encrypt things published online	26.00%	12.99%	2.90
Only add friends in real life	38.00%	59.74%	-3.88
	Studying	On-the-job	Z
Login social network at least once a day	92.57%	78.29%	3.45
Treat personal regular activities as privacy	60.57%	73.64%	-2.44
No concern for access rights	9.14%	16.80%	-2.33

**B. THE USERS’ FACTORS RELATED TO SOCIAL NETWORK PRIVACY**

According to the collected questionnaires, on the one hand, most young people have security consciousness that they are basically familiar with the privacy settings of the mobile devices and social networks. For instance, 89.03% of the volunteers would choose or try to download applications from trusted app stores, and 87.17% would choose not to install the app when the permission is so excessive. On the other hand, users tend to overlook potential ways that may disclose private information. For example, half of the users, namely 50.66%, would store accounts and passwords in the mobile devices, and 33.55% would not keep the information under security protection.

In addition, we find that for different people, their behavior varies as well. Here we utilize two-proportion z-test to test whether one population proportion equals another population proportion, since it is a classic method to determine whether the difference between two is significant. Under the significance level  $\alpha$  of 0.05, the difference is called significant if  $Z > 1.96$  or  $Z < -1.96$ . Here, 1.96 refers to the threshold of two-tailed test when  $\alpha = 0.05$ . Then we have three comparisons (Table 6) including comparison between the users of different ages, between the male and the female, and between those studying and on-the-job. The z-test works only when the scale of population is larger than 30. Therefore we only compare those younger than 20 and those between 20 and 30 years old.

**C. IDENTIFYING DOMINANT ATTACKS FROM MARKOV ANALYSIS**

In this section, we discuss the dominant ones at presence of existing attacks which the defender should pay more attention to. Based on the transition matrix, when social network reaches balanced, the probability of each state is as follows.

- Normal state:  $\pi_0 = 0.390$
- Semi-failed state:  $\pi_1 = 0.136, \pi_2 = 0.050, \pi_4 = 0.169, \pi_6 = 0.109$
- Failed state:  $\pi_3 = 0.011, \pi_5 = 0.073, \pi_7 = 0.033, \pi_8 = 0.029$

Therefore, the probability of privacy disclosure  $FA$  would be the sum of failed state which is 0.147, and the availability of system  $AV$  would be  $AV = 1 - FA = 0.853$ . Based on this value, we could compare the risk issues of different social network systems and identify the dominant attack strategies for each system. Here, from the technical point of view, it is observed that state of the highest probability is  $S_5$ , followed by  $S_7$  and  $S_8$ . That is, for volunteers who are mostly between 20 and 30 years old, social network platform reveals most of the private information, mainly thanks to their high privacy consciousness.

Disclosure related to social network platform happens mainly for technical reasons. On the one hand, vulnerabilities are inevitable, especially those by human error. For example, Oauth 2.0 protocol, which is widely adopted by online social network, suffers from vulnerabilities [31]. Attackers could hijack callback domain of third party websites to malicious sites and use XSS attack to access to victims’ accounts, that is, obtaining users’ private information freely. On the other hand, social networking sites may not pay much attention to technical maintenance and do not regularly update technical barriers, lowering the technical barrier as well.

Privacy leakage on user side happens for lack of privacy awareness. Drawing on the survey, users would mostly post photos and locations since they could photograph and publish in social network wherever they go, thus inevitably exposing the location. Moreover, personal settings become another issue since users tend to omit the settings for complex operations or don’t know how to set permissions, thus losing the protection of their privacy. Furthermore, the majority of users’ online information is true in order to maintain the existing social circle, making their personal information accessible to others. Therefore, the hidden danger revealed by this evaluation is consistent with the reality.

**D. PRIVACY DEFENSE STRATEGIES UNDER DIFFERENT ATTACK/DEFENSE DIFFICULTY**

In this section, we provide suggestions on defense strategies in a dynamic system. In practice, the difficulty of a certain atomic event changes all the time. Thus we analyze its impact on social network system. In the simulation, we set the repairing difficulty as a fixed value and continuously alter the attack difficulty, and vice versa. Results show that increasing the difficulty of “Platform accidental disclosure” ( $B_5$ ) will increase the security level of social network system by 0.94%, which is 0.70% higher than increasing “Obtain access authority” ( $B_6$ ). Since the original value is between 0 to 5, the security level doesn’t seem to improve significantly. But under the condition of limited resources, we could start by protecting from  $B_5$ . Specifically, the change of system secu-



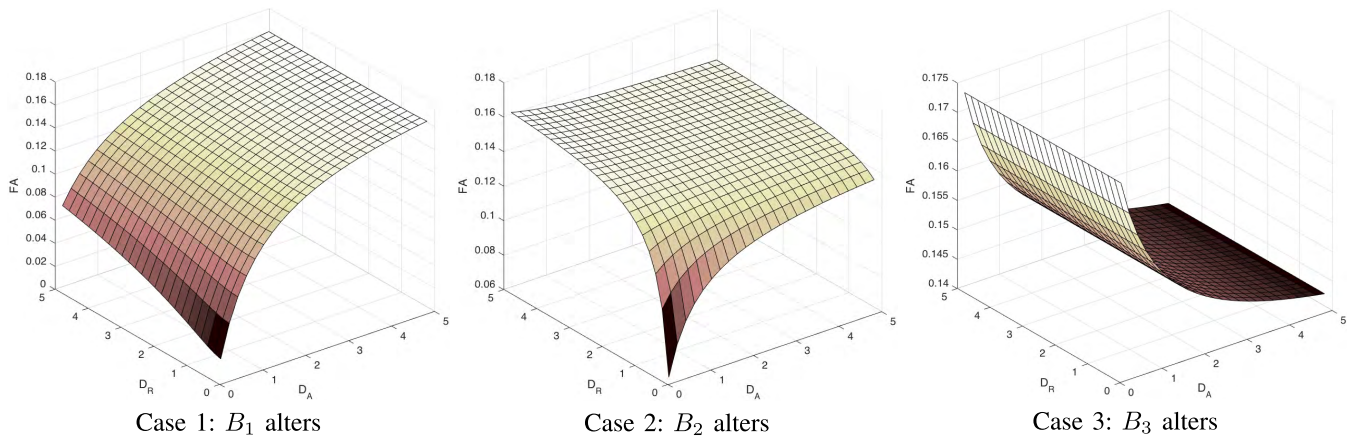


FIGURE 3. Changes of system FA.

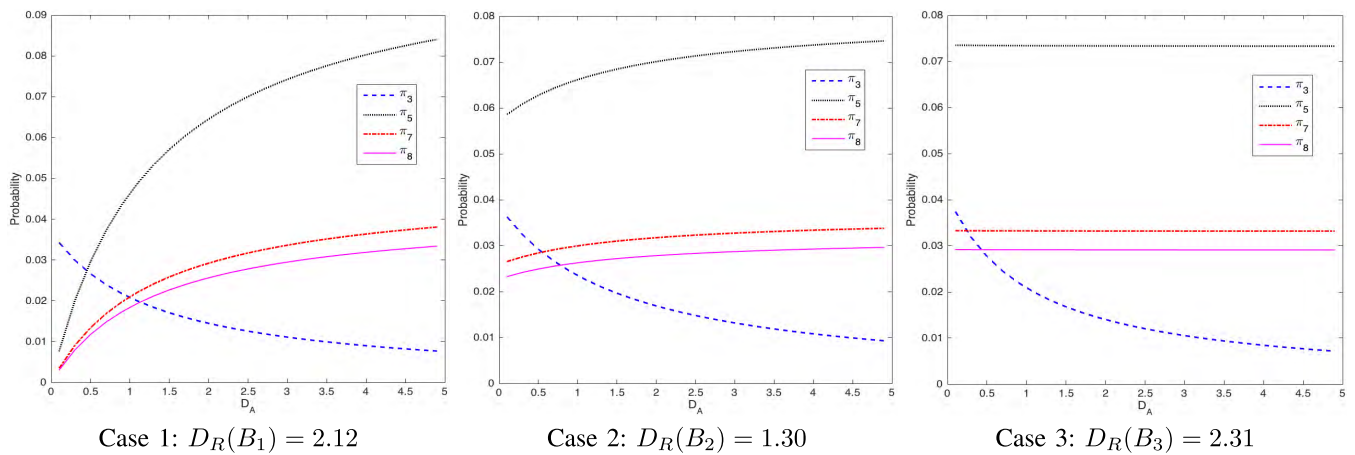


FIGURE 4. System FA decomposition.

ity level could be divided into three categories (see Fig.3, Fig.4):

- 1)  $B_1$ : System risk increases even if it is more difficult to conduct the attack  $B_1$ . That is, when bypassing security scanning becomes harder, privacy disclosure would be more possible to happen. It is because we suppose that attacker could learn the difficulty of attacks and after a number of attempts, rational attacker would master the difficulty level of each attack and tend to choose the relatively easier one, making social network suffer greater risk instead. In this case, he would turn to attack the social network platform. Therefore, simply improving the defense of a certain attack would lower the probability of privacy disclosure of this kind, but may increase the risk of the whole system.
- 2)  $B_2, B_4, B_6$  (take  $B_2$  for example since these three types are similar): When repairing difficulty is low, the system risk increases with the increase of the attack difficulty. That is, when users focus more on device settings and restrict access permissions to online posts, the system is under a greater risk instead. In contrast, when repairing difficulty becomes higher, system risk decreases with increase of attack difficulty. Especially,

when the difficulties of both repair event and attack event are low, the risk of social network system is low as well. It is obvious that the attacker prefers to those with low attack difficulty and high repair difficulty. Thus for the defender, increasing the difficulty of these attacks would be relatively efficient.

- 3)  $B_3, B_5, B_7, B_8$  (take  $B_3$  for instance): For these attacks, the risk of privacy disclosure decreases with the increase of attack difficulty. These attack events share the same characteristic that they contribute to the last step of attack route, which is related to the assumption that the defender would always bring the system back to work even in failed state. According to Case 3 of Fig.4,  $B_5$  has the most significant change, thus preventing from “Platform Accidental Disclosure” is of the high priority. In addition, it is found that the complex attack process could effectively assure privacy security. For example, an attacker has to either successively achieve  $B_1, B_2, B_3$  or  $B_4, B_5$ , while system risk decreases less when increasing difficulty of  $B_3$  than  $B_5$ . Therefore, it is necessary to increase the difficulty of last step for a simple attack process, which is consistent with the system structure, the simulation results and our common sense.

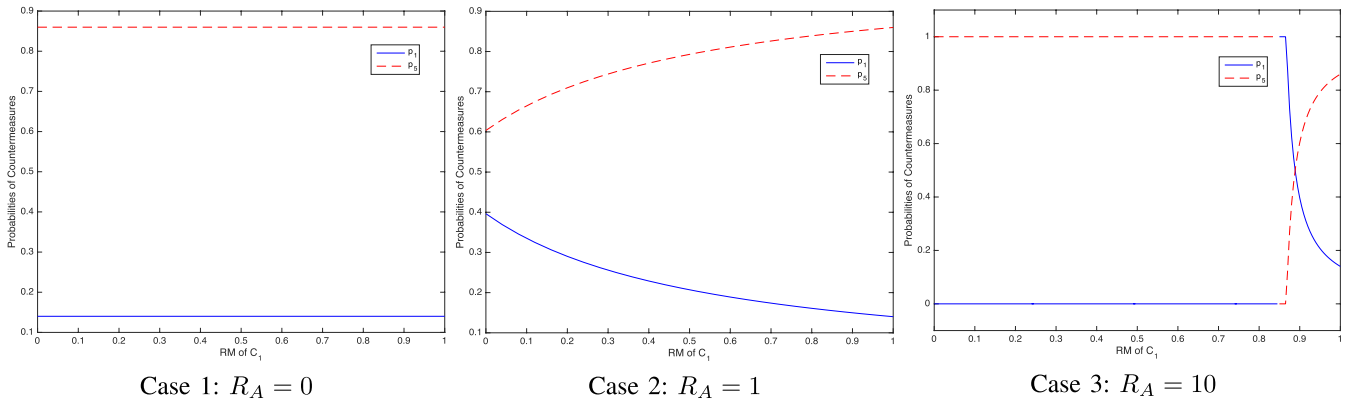


FIGURE 5. Equilibrium strategy of defenders.

TABLE 7. User profiles from social network.

Facebook	Gender	Location	Hometown	Phone
	74.96%	52.86%	47.77%	2.48%
	Language	Company	Status	Birthday
	27.58%	38.58%	20.53%	10.38%
Twitter	Relatives	Interests	About Me	Age Group
	13.38%	37.73%	27.54%	68.74%
Myspace	Location	Links	About Me	Connections
	91.51%	71.73%	82.67%	99.67%
	Age	Gender	Status	Interests
	41.19%	76.61%	46.07%	18.35%
	Hometown	Body Type	Ethnicity	Religion
15.67%	12.55%	13.66%	18.05%	
Children	Education	Occupation	Income	
13.73%	15.00%	15.57%	3.86%	

E. INFLUENCE OF ECONOMIC FACTORS

In face of various attackers, we offer guidance on defense strategies when attackers have different attitude towards profitability. For instance, we illustrate the impact of profit to cost ratio  $R_A$  in terms of  $M_4$ . In details, we set  $R_A$  at three different values and simulate new equilibrium when continuously changing risk mitigation rate  $RM$  of  $C_1$ . While in real life, according to Fig.5 and Table 7, which is the proportion of users’ private information clawed in social network, we find that improving security consciousness is basically the dominant strategy. Considering over 90% users release locations in twitter and around 80% of them are true based on the survey, over 70% of real location information is just exposed to attackers. And with hundreds of millions of Twitter users, it is high time that we should equip ourselves of privacy consciousness.

1) Fix  $R_A$  to be 0, which means that attacker has no economical profit by implementing the  $M_4$  attack. Take “WikiLeaks” for example, this non-profit organization releases secret information only to inform the public of important news. In this case, the defender would choose  $C_1$  with probability 0.14 and  $C_5$  with 0.86 in Nash Equilibrium. Thus when facing such an attack, raising privacy consciousness would be much more effective than identifying and fixing vulnerabilities.

- Fix  $R_A$  to be 1. Then the probability of selecting  $C_1$  increases compared to Case 1, since  $C_1$  is intended to prevent  $M_4$  while  $C_5$  is a fundamental solution. Thus we can tell that when the profit of selected attack strategy increases, the probability of selecting corresponding defense strategy increases as well. According to Fig.5,  $C_1$  decreases with the increase of  $RM$  for the decrease of return of attack, thus rational defender would still prefer strategy  $C_5$ . For instance, fishing url camouflages itself as an interesting site to collect personal information. But users would be more cautious about this site than the former “WikiLeaks” since they realize that attacker may benefit from their behavior.
- Fix  $R_A$  to be 10. Then we find that when  $RM$  is less than 0.85, the defender would only choose  $C_5$  since  $C_1$  is a dominated strategy for weak risk mitigation. In contrast, when  $RM$  is larger than the threshold, the defender would select  $C_1$  with certainty. But  $C_1$  decreases when  $RM$  keeps increasing, which implies that attacker turns to other strategies since  $C_1$  has blocked the system from  $M_4$  attack. And this is a great reflection of the game theory. Take “WannaCry” attack for instance. At the beginning of its propagation, users have to utilize strategy  $C_5$  through self-awareness protection. But when it is widely spread, the operator would take strategy  $C_1$ , making the attacker stop this attack and turn to others.

V. CONCLUSION

In this paper, We firstly proposed a privacy disclosure attack-defense tree to describe a series of attack steps launched by the attackers to achieve their ultimate goals and the corresponding countermeasures that can be adopted by the social network security defenders. To further illustrate a dynamic attacking process, we secondly extended a Markov chain based approach to model a temporal-aware attack-defense tree. Lastly, we introduced an attack-defense game to analyze the potential strategies performed by the attacker and the defender. To validate the proposed model, we used two types of datasets: One is gathered through an online questionnaire survey; the other is from three real-world mobile social networks. We performed extensive evaluations in which we

find: concerning technical difficulty, raising attack difficulty of single event doesn't necessarily reduce system risk, since attackers will turn to lower attack difficulty event which with relatively higher defending difficulty; concerning economic factors, operators responding to mobile social network platform attack only have limited effect, hence, improving personal awareness of privacy protection is the effective strategies to reduce system risk.

#### APPENDIX A

- 1) Do you think it is more likely to reveal personal information when logging into social network via mobile devices (phones, tablets, etc.) than computer?
  - (A) Yes
  - (B) No
- 2) How often do you usually log into social network via mobile devices?
  - (A) At least 3-4 times a day
  - (B) 1-2 times a day
  - (C) 5-6 times a week
  - (D) 3-4 times a week or less
- 3) Where do you usually download apps?
  - (A) Only from trusted app stores
  - (B) Basically from trusted app store
  - (C) Never consider the source and download directly
- 4) Have your mobile devices (phone, tablet, etc.) installed anti-virus software?
  - (A) Yes
  - (B) No
- 5) What is your installed anti-virus software?
  - (A) Inherent software of device
  - (B) Not installed
  - (C) Others \_\_\_\_\_ (Write the name here)
- 6) How often do you perform security scan?
  - (A) Once a day
  - (B) Every 3-4 days
  - (C) Once a week
  - (D) Every 7 days or less
- 7) Are you concerned about the access rights when installing software on mobile devices?
  - (A) Always concerned
  - (B) Mostly concerned
  - (C) Seldom concerned
  - (D) Never
- 8) Will you prohibit some access rights in some cases?
  - (A) Yes
  - (B) No
  - (C) Occasionally
- 9) Will you stop installing when the app requests excessive rights?
  - (A) Stop installing
  - (B) It depends
  - (C) Insist on installation
- 10) What stored in mobile devices do you think is related to private information? (multiple choices)
  - (A) Contacts and messages
  - (B) Photos
  - (C) Diary or daily arrangement
  - (D) Accounts and passwords
- 11) Is private informations on the mobile device protected by security measures?
  - (A) None of them
  - (B) Some of them
  - (C) All of them
- 12) What percent of social network do you think requires personal information in registration?
  - (A) 80-100%
  - (B) 60-80%
  - (C) 40-60%
  - (D) Less than 40%
- 13) What percent of information you fill in when registering social network is real?
  - (A) All of them
  - (B) Most part
  - (C) Small part
  - (D) None of them
- 14) Would you set access authority when posting information on social network?
  - (A) Only I have the right
  - (B) Only friends have the right
  - (C) Some contents set access authority or some of the friends have the right
  - (D) Don't know the set thing and posting by default or anyone has the right
- 15) Will you encrypt your posts on social network?
  - (A) Encrypt all information
  - (B) Encrypt most of the information
  - (C) Encrypt small part of information
  - (D) Encrypt none of information
- 16) What would you do when adding your friends?
  - (A) Add anyone as long as he applies, even strangers
  - (B) It depends for strangers
  - (C) Only friends I know in real life
- 17) How is your interaction with friends?
  - (A) Interact every day
  - (B) Only reply to interesting content
  - (C) Do not reply
- 18) What information do you think would inadvertently disclose when interacting with friends? (Can have multiple choices)
  - (A) Location information
  - (B) Hobbies
  - (C) Daily arrangements (travel, shopping, etc.)
  - (D) Regular activities
- 19) Which of the following information will you post on social network? (Can have multiple choices)

TABLE 8. Score conversion rule.

Question No.	atomic event	(A)	(B)	(C)	(D)
1	-				-
2	B <sub>7</sub>	2	3	4	5
3	B <sub>1</sub>	5	3	1	-
4	B <sub>1</sub>	4	1	-	-
5	-				-
6	B <sub>1</sub>	5	4	2	1
7	B <sub>2</sub>	5	4	2	1
8	B <sub>2</sub>	5	1	3	-
9	B <sub>2</sub>	5	3	1	-
10	B <sub>3</sub>	1 item 1, 2 items 2, 3 items 4, 4 items 5			
11	B <sub>3</sub>	1	3	5	-
12	B <sub>4</sub>	1	2	4	5
13	B <sub>4</sub>	1	2	4	5
14	B <sub>6</sub>	5	4	2	1
15	B <sub>6</sub>	5	4	2	1
16	B <sub>6</sub>	1	3	5	-
17	B <sub>8</sub>	1	3	5	-
18	B <sub>8</sub>	1 item 1, 2 items 2, 3 items 4, 4 items 5			
19	B <sub>7</sub>	1 item 1, 2 items 2, 3 items 4, 4 items 5			
20	-				-

TABLE 9. Evaluation of ROI.

Attack	ALE	Countermeasures	RM	CI	ROI
M <sub>3</sub>	6	C <sub>2</sub>	0.25	4	-0.625
		C <sub>3</sub>	0.75	8	-0.4375
		C <sub>1</sub> , C <sub>4</sub>	0	-	-1
		C <sub>5</sub>	0.75	6	-0.25
M <sub>4</sub>	10	C <sub>1</sub>	0.75	8	-0.0625
		C <sub>2</sub> , C <sub>3</sub> , C <sub>4</sub>	0	-	-1
		C <sub>5</sub>	0.5	6	-0.1667
M <sub>5</sub>	6	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub>	0	-	-1
		C <sub>4</sub>	0.25	4	-0.5
		C <sub>5</sub>	0.75	6	-0.25
M <sub>6</sub>	2	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub>	0	-	-1
		C <sub>4</sub> , C <sub>5</sub>	0.5	6	-0.8333

TABLE 10. Evaluation of ROA.

Attack	GE	Cost <sub>A</sub>	Countermeasures	Cost <sub>AC</sub>	ROA
M <sub>3</sub>	8	6	C <sub>2</sub>	4	-0.4
			C <sub>3</sub>	6	-0.833
			C <sub>1</sub> , C <sub>4</sub>	0	0.33
			C <sub>5</sub>	4	-0.8
M <sub>4</sub>	10	8	C <sub>1</sub>	8	-0.8438
			C <sub>2</sub> , C <sub>3</sub> , C <sub>4</sub>	0	0.25
			C <sub>5</sub>	4	-0.5833
M <sub>5</sub>	6	4	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub>	0	0.5
			C <sub>4</sub>	2	-0.25
			C <sub>5</sub>	4	-0.8125
M <sub>6</sub>	2	4	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub>	0	-0.5
			C <sub>4</sub> , C <sub>5</sub>	4	-0.875

- (A) Location
- (B) Photo
- (C) Hobbies
- (D) Friends' personal information (location, hobbies, etc.)

20) Which of the following information leaks do you think is related to privacy disclosure? (Can have multiple choices)

- (A) Location information
- (B) Hobbies
- (C) Daily arrangements (travel, shopping, etc.)
- (D) Regular activities

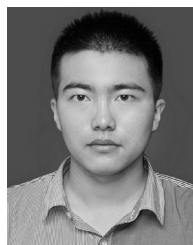
APPENDIX B

See Tables VIII–X.

REFERENCES

- [1] (2017). *Number of Global Social Media Users Exceeds Three Billion, With Facebook the Number One Platform*. [Online]. Available: <https://www.b2bmarketing.net/en-gb/resources/news/number-global-social-media-users-exceeds-three-billion-facebook-number-one-platform>
- [2] D. Parisi. (2017). *200 Mobile Apps, Sites Leaked Personal Information Last Year: Report*. [Online]. Available: <https://www.retaildive.com/ex/mobilecommercedaily/more-than-200-mobile-apps-and-sites-leaked-personal-information-last-year-report>
- [3] J. Jiang. (2017). *Personal Information Leakage on The Rise in China: Report*. [Online]. Available: <http://en.people.cn/n3/2017/03/31/c90000-9197748.html>
- [4] D. Guarini. (2013). *Experts Say Facebook Leak Of 6 Million Users' Data Might Be Bigger Than We Thought*. [Online]. Available: [https://www.huffingtonpost.com/2013/06/27/facebook-leak-data\\_n\\_3510100.html](https://www.huffingtonpost.com/2013/06/27/facebook-leak-data_n_3510100.html)
- [5] E. Bott. (2012). *Why Do Not Track is Worse Than a Miserable Failure*. [Online]. Available: <http://www.zdnet.com/article/why-do-not-track-is-worse-than-a-miserable-failure>
- [6] H. Li, Q. Chen, H. Zhu, D. Ma, H. Wen, and X. S. Shen, "Privacy leakage via de-anonymization and aggregation in heterogeneous social networks," *IEEE Trans. Depend. Sec. Comput.*, to be published.
- [7] H. Li, Q. Chen, H. Zhu, and D. Ma, "Hybrid de-anonymization across real-world heterogeneous social networks," in *Proc. ACM Turing 50th Celebration Conf. China*, 2017, p. 33.
- [8] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Comput.*, vol. 2, no. 1, pp. 46–55, Jan. 2003.
- [9] J. Krumm, "Inference attacks on location tracks," in *Proc. Int. Conf. Pervasive Comput.*, 2007, pp. 127–143.
- [10] A.-M. Olteanu, K. Huguenin, R. Shokri, M. Humbert, and J.-P. Hubaux, "Quantifying interdependent privacy risks with location data," *IEEE Trans. Mobile Comput.*, vol. 16, no. 3, pp. 829–842, Mar. 2017.
- [11] L. L. Zhang, C.-J. M. Liang, Z. L. Li, Y. Liu, F. Zhao, and E. Chen, "Characterizing privacy risks of mobile apps with sensitivity analysis," *IEEE Trans. Mobile Comput.*, vol. 17, no. 2, pp. 279–292, Feb. 2017.
- [12] N. Korula and S. Lattanzi, "An efficient reconciliation algorithm for social networks," *VLDB Endowment*, vol. 7, no. 5, pp. 377–388, 2014.
- [13] Z. Zhang, Q. Gu, T. Yue, and S. Su, "Identifying the same person across two similar social networks in a unified way: Globally and locally," *Inf. Sci.*, vol. 394, pp. 53–67, Jul. 2017.
- [14] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff, "Identifying users across social tagging systems," in *Proc. ICWSM*, 2011, pp. 522–525.
- [15] S. Lai, H. Li, H. Zhu, and N. Ruan, "De-anonymizing social networks: Using user interest as a side-channel," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Nov. 2015, pp. 1–5.
- [16] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2011, pp. 247–262.
- [17] N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Optimal geo-indistinguishable mechanisms for location privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2014, pp. 251–262.
- [18] A. Narayanan, N. Thiagarajan, M. Lakhani, and D. Boneh, "Location privacy via private proximity testing," in *Proc. NDSS*, 2011, pp. 1–12.
- [19] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *Proc. ACM Conf. Comput. Commun. Secur.*, 2012, pp. 617–627.
- [20] Q. Xiao et al., "POSTER: LocMask: A location privacy protection framework in Android system," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2014, pp. 1526–1528.
- [21] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Trans. Depend. Sec. Comput.*, to be published.
- [22] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. 1st Int. Conf. Mobile Syst., Appl. Services*, 2003, pp. 31–42.
- [23] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in GPS traces via uncertainty-aware path cloaking," in *Proc. 14th ACM Conf. Comput. Commun. Secur.*, 2007, pp. 161–171.

- [24] A. R. Beresford, A. Rice, N. Skehin, and R. Sohan, "MockDroid: Trading privacy for application functionality on smartphones," in *Proc. 12th Workshop Mobile Comput. Syst. Appl.*, 2011, pp. 49–54.
- [25] M. Li et al., "All your location are belong to us: Breaking mobile social networks for automated user location tracking," in *Proc. 15th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2014, pp. 43–52.
- [26] I. Polakis, G. Argyros, T. Petsios, S. Sivakorn, and A. D. Keromytis, "Where's Wally?: Precise user discovery attacks in location proximity services," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, 2015, pp. 817–828.
- [27] T. Moore, M. Raya, J. Clulow, P. Papadimitratos, R. Anderson, and J.-P. Hubaux, "Fast exclusion of errant devices from vehicular networks," in *Proc. 5th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw. (SECON)*, Jun. 2008, pp. 135–143.
- [28] S. Du, X. Li, J. Du, and H. Zhu, "An attack-and-defence game for security assessment in vehicular ad hoc networks," *Peer-to-peer Netw. Appl.*, vol. 7, no. 3, pp. 215–228, 2014.
- [29] S. Bistarelli, M. Dall'Aglia, and P. Peretti, "Strategic games on defense trees," in *Proc. Int. Workshop Formal Aspects Secur. Trust*, 2006, pp. 1–15.
- [30] Y. Zhang, J. Tang, Z. Yang, J. Pei, and P. S. Yu, "Cosnet: Connecting heterogeneous social networks with local and global consistency," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 1485–1494.
- [31] P. Hu, R. Yang, Y. Li, and W. C. Lau, "Application impersonation: Problems of OAuth and API design in online social networks," in *Proc. 2nd ACM Conf. Online Social Netw.*, 2014, pp. 271–278.



**LU ZHOU** received the B.Eng. degree in computer science and technology from Sichuan University, China, in 2015. He is currently pursuing the Ph.D. degree in computer science and technology with Shanghai Jiao Tong University, China. His research interests include security and privacy in vehicular network and cognitive radio networks.



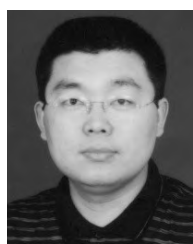
current research interests include the data driven analysis of online social networks and privacy.

**MINHUI XUE** received the B.Sc. degree in fundamental mathematics from East China Normal University in 2013, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Software Engineering. He is also serving as a Visiting Scholar with the Courant Institute of Mathematical Sciences and the Tandon School of Engineering, New York University, as well as a Research Assistant with New York University Shanghai, advised by Prof. K. W. Ross (NYU). His

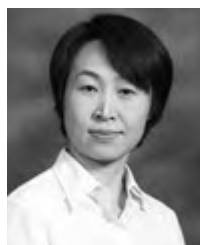


published 35 international journal papers, including the JSAC, TDSC, TPDS, TMC, TWC, and TVT, and 60 international conference papers, including the ACM CCS, ACM MOBIKOM, ACM MOBIHOC, IEEE INFOCOM, and IEEE ICDCS. His current research interests include network security and privacy enhancing technologies. His papers were top 100 most cited Chinese papers published in international journals in 2014. He was a Distinguished Member of the IEEE INFOCOM Technical Program Committee in 2015. He received a number of awards including the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2014, the Supervisor of Shanghai Excellent Master Thesis Award in 2014, the Outstanding Youth Post Expert Award for Shanghai Jiao Tong University in 2014, the SMC Young Research Award of Shanghai Jiao Tong University in 2011, and the Young Scholar Award of Changjiang Scholar Program by Ministry of Education, China, in 2016. He was a co-recipient of best paper awards of the IEEE ICC (2007) and Chinacom (2008) as well as the IEEE GLOBECOM Best Paper Nomination in 2014.

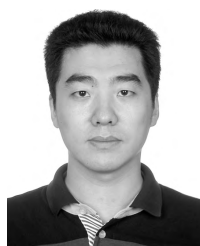
**HAOJIN ZHU** (M'09–SM'16) received the B.Sc. degree in computer science from Wuhan University, China, in 2002, the M.Sc. degree in computer science from Shanghai Jiao Tong University, China, in 2005, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Canada, in 2009. Since 2017, he has been a Full Professor with the Computer Science Department, Shanghai Jiao Tong University.



received the B.S., M.S., and D.Sc. degrees from the College of Computer, National University of Defense Technology, China, in 1988, 1995, and 1998, respectively. He is currently a Professor with the Institute of Information Engineering, Chinese Academy of Sciences. His research interests include wireless sensor networks, Internet of Things, and intelligent transportation systems. He is a Senior Member of the China Computer Federation. He was a Guest Editor of special issues in EURASIP, the *Journal of Wireless Communications and Networking* and the *International Journal of Distributed Sensor Networks*, respectively. He is an Editor of the *Journal of Computer Science* and the *Journal Computer Applications* (Chinese journals).



**SUGUO DU** received the Ph.D. degree from the School of Mathematical and Information Sciences, Coventry University, U.K., in 2002. She is currently an Associate Professor with the Department of Management Science, Shanghai Jiao Tong University, China. Her research has been supported by the National Science Foundation of China. Her current research interests include risk and reliability assessment, vehicular networks security and privacy protection, and social networks security management.



**XIAOLONG LI** received the B.Eng. degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2009, and the M.Sc. degree from Shanghai Jiao Tong University in 2013, where he is currently pursuing the Ph.D. degree with the Antai College of Economics and Management. His research interests include risk and reliability assessment, robust optimization, revenue management, and other areas of management science and operations management.



**JINLI ZHONG** received the B.Sc. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China, in 2012, where she is currently pursuing the M.Sc. degree. Her research interests include social networks privacy, vehicular network security, and network security and privacy.