**Journal of Organizational and End User Computing**
Volume 32 • Issue 3 • July-September 2020

# A Risk Analysis Framework for Social Engineering Attack Based on User Profiling

Ziwei Ye, State Key Laboratory of Mathematical Engineering and Advanced Computing, China
Yuanbo Guo, State Key Laboratory of Mathematical Engineering and Advanced Computing, China
Ankang Ju, State Key Laboratory of Mathematical Engineering and Advanced Computing, China
Fushan Wei, State Key Laboratory of Mathematical Engineering and Advanced Computing, China
Ruijie Zhang, State Key Laboratory of Mathematical Engineering and Advanced Computing, China
Jun Ma, State Key Laboratory of Mathematical Engineering and Advanced Computing, China

## ABSTRACT

Social engineering attacks are becoming serious threats to cloud service. Social engineering attackers could get Cloud service custom privacy information or attack virtual machine images directly. Existing security analysis instruments are difficult to quantify the social engineering attack risk, resulting in invalid defense guidance for social engineering attacks. In this article, a risk analysis framework for social engineering attack is proposed based on user profiling. The framework provides a pathway to quantitatively calculate the possibility of being compromised by social engineering attack and potential loss, so as to effectively complement current security assessment instruments. The frequency of related operations is used to profile and group users for respective risk calculation, and other features such as security awareness and capability of protection mechanism are also considered. Finally, examples are given to illustrate how to use the framework in actual scenario and apply it to security assessment.

## KEYWORDS

Authority, Cloud Security, Network Security Assessment, Operating Frequency, Risk Analysis, Social Engineering, User Profiling, Vulnerability

## INTRODUCTION

With the development of cloud security, social engineering attacks have been paid more and more attention because of the outstanding capability of penetrating cloud service which is difficult to the conventional techniques. However, the awareness for social engineering of cloud service customs and providers is relatively low (Krombholz et al., 2015; Kuyoro et al., 2011). Improving awareness and ability of prevention to social engineering attacks is of great significance for the accuracy of cloud security assessment and anti-attack capability.

Current researches on social engineering attacks are mainly of classification and qualitative analysis based on known cases. Chandra et al., (2015) treats social engineering attack as a feature of APT and discuss how to deploy defense system in cloud. A case is given to show how to run a malicious

DOI: 10.4018/JOEUC.2020070104

virtual machine in Amazon EC2 by social engineering (Meer et al., 2009). Some researchers summed up the existing social engineering attack classification achievements (Fooxy et al., 2011) and frequently used social engineering malwares (Abraham & Chengalur-Smith, 2010). In Mouton et al. (2014) the writers extracted a social engineering attack framework based on Kevin Mitnick's book "The art of deception: Controlling the human element of security" (Mitnick & Simon, 2001). A multi-layered model was presented for assessing possible social engineering exploits (Jaafor & Birregah, 2016). How educational and technological means can be used to reduce social engineering risk which the social media users faced was discussed (Tayouri 2015). Sheng et al. (2010) used a roleplay survey instrument to assess user's vulnerability of phishing. However, there is little quantitative evaluation of possibility of being compromised by social engineer attacker and potential loss expect (Sheng et al., 2010), so that it is impossible to compare risks between social engineering attacks and technical attacks without social factors. Social engineering attack risk couldn't be taken into account in network security assessment, resulting in negative effects on network reinforcement.

User profiling is an important application of big data. By adding descriptive tags to users, it can depict users from multiple dimensions, and reflect users' behaviors, hobbies, jobs, etc. In recent years, user profiling has been widely used in network security research. An automated insider threat detection system was realized user profiling (Legg et al., 2017). Nurse et al. (2016) the writers studied how to distinguish online identify falsification. A framework was proposed to analyze users' attributes and recognize accounts which belong to the same user from different social networks (Monika et al., 2016). As social engineering attack is the exploitation of human weaknesses, and the object of user profiling is also human, user model can be built by user profiling to find possible social engineering attack surface. It will be more pertinence comparing with techniques whose objects are networks or computers.

Aiming at the problem that existing network security assessment techniques cannot quantify the risk of social engineering attack, the authors present a risk analysis framework for social engineering attack based on user profiling. By extracting the relevant features, the possibility of being compromised and potential loss caused by social engineering attack could be quantified. The framework can be applied to various network security assessment instruments to optimize results and guide the employment of protection mechanisms against social engineering attack.

## SOCIAL ENGINEERING VULNERABILITY AND RISK

Social engineering attacks are becoming more and more complicated with the development of information technology. According to the classification summary on social engineering attacks (Foozy et al., 2011), it is recognized that social engineering attacks can be classified into two categories according to their correlation with computer and network: human-based attacks and technical-based attacks. Aiming at introducing social engineering risk analysis into network security assessment, the framework mainly directs at the technical-based social engineering attacks which are more dependent on computer and network. Within the instances of technical-based social engineering attacks, related operation frequencies of phishing, waterholing, malware and pop-up windows are obvious and easy to extract by user profiling. On this account the authors choose the above four attack types as objects. In future works, our work would be extended to human-based attacks, and other emerging attacks.

### Typical Social Engineering Attacks

#### *Phishing*

Phishing is one of the most common forms of cyber-attacks. Usually the attacker sends e-mails or SMS with malicious link or attachment containing malicious code, to induce victims to open attachment or click on the link, so as to realize remote control or get information.

The possibility of being attacked is affected by the frequency of e-mailing. The higher frequency, the more likely it is that user receives a phishing e-mail as a normal one and clicks on the malicious link or performs the attachment. Therefore, when analyze the vulnerability of phishing, the frequency of e-mailing can be considered as the main reference index.

### *Waterholing*

Waterholing is such a type of social engineering attack that attacker embeds malicious code in particular websites which is frequently accessed by targeted individuals. Nowadays more and more attackers choose waterholing to attack specific organization or enterprise employees. The targeted websites are usually used for exchanging news or technology of industry, and the users must be industry practitioners, so that the directivity of waterholing is extremely strong.

The higher frequency of browsing a specific website, the more likely to be attacked by waterholing. Therefore, when analyze the vulnerability of waterholing, the frequency of accessing a particular website can be considered as a major reference indicator.

### *Malware*

Malware in the field of social engineering refers to malicious software which is disguised as secure software, and tricks users to download and execute it. The main means is to redistribute the legitimate software before releasing it. In order to avoid the prevention by security software, attackers often release malwares under the name of "cracked version" or "register machine", so as to induce users to shut down security software before executing malware.

Obviously, the more software users download from Internet, the more likely they are attacked by malware. Therefore, when analyze the vulnerability of malware, the frequency of downloading software can be considered as the main reference index.

### *Pop-Up Windows*

Pop-up windows refers to malicious code embedded in a web page. When the victims access the page, malicious code can be disguised as browser plug-in, and the victims will be induced to download and execute the malicious code as a legitimate software. The difference between pop-up windows and malware is that malwares compromise victims when they actively search for needed software, while pop-up windows compel victims to accept malicious code passively.

Similar to waterholing, the higher the frequency of accessing a specific website, the more likely it is to be attacked by pop-up windows. Therefore, when analyze the vulnerability of pop-up windows, the frequency of accessing a specific website can also be taken as the main reference index.

## Definitions of Social Engineering Vulnerability and Risk

The concept of social engineering vulnerability is similar to the vulnerability of hardware and software, reflecting whether the user is easy to be compromised by social engineering attack. It's decided by the frequency of user operation, user's security awareness, and protection mechanism in the network related to specific social engineering attack. The higher frequency there is, the greater attack surface is exposed. And the less security awareness, the more lack of defense ability. The better protective effect of protection mechanism, the less possibility of compromised. Social engineering risk forecasts potential loss caused by social engineering attack, determined by the user's social engineering vulnerability and authority. The more vulnerability and greater user authority there is, the more possibility that attacker compromises victim and gets a greater authority. User profiling implements the extraction of related features by analyzing the behaviors of network users, and provides a data base for the quantitative calculation of social engineering vulnerability and risk.

Table 1 shows the symbols and corresponding descriptions used in this article, in which the upper case ones represents the set or tuple, and the lower case ones represents the variable.

**Table 1. Summary of notations**

| Symbols | Definitions |
|---|---|
| ***OF*** | Set of operating frequencies related to social engineering attack |
| ***SA*** | Set of security awareness in social engineering attack |
| ***DD*** | Set of defense degrees of protection mechanism |
| ***UP(ua,OF,SA)*** | User profiling: a triple composed by user authority, set of operation frequency and set of security awareness |
| ***ua*** | The current user authority |
| ***ga*** | The authority which attacker may gain by a successful attack |
| ***Wei*** | Set of weighting factors |
| ***AP*** | Set of possibilities of attacked by social engineering attack |
| ***CP*** | Set of possibilities of compromised when attacked |
| ***SoEV*** | Set of user's social engineering vulnerabilities |
| ***cv*** | Composite social engineering vulnerability |
| ***SoER*** | Set of social engineering risks |
| ***cr*** | Composite social engineering risk |
| ***x*** | $x \in \{phish, waterholing, malware, pop\text{-}up\}$: the types of social engineering attacks user may face |
| ***y*** | $y \in \{e\text{-}mail, browse, download\}$: user's operations related social engineering attacks |

Because the vulnerability and risk are different for each type of social engineering attack, the definition of user features, vulnerability and risk should be differentiated. The authors use $x \in \{phish, waterholing, malware, pop\text{-}up\}$ to express the types of attacks users may face, and make the following definitions:

**Definition 1:** Let *OF* be a set of frequency of user operation related to social engineering attacks $OF=\{of_{e\text{-}mail}, of_{browse}, of_{download}\}$. Elements in the set express frequencies of e-mailing, browsing specific websites and downloading software in a unit period in turn. Taking into account general users' work habits, one week could be taken as the unit period. Too short cycle selection will lead to user profiling error with the actual situation, and too long to too much data to deal with and negative influence of analysis efficiency.

**Definition 2:** Let *SA* be a set of security awareness of users in social engineering attack $SA=\{sa_{phish}, sa_{waterholing}, sa_{malware}, sa_{pop\text{-}up}\}$.Elements in the set in turn are users' security awareness of phishing, waterholing, malware and pop-up windows.

**Definition 3:** Let *DD* be a set of defense degree of protection mechanism of enterprise $DD=\{dd_{phish}, dd_{waterholing}, dd_{malware}, dd_{pop\text{-}up}\}$, reflecting the protective effects against specific social engineering attack. The values are given based on the probabilities of protection mechanism to prevent social engineering attack in the actual scene.

**Definition 4:** Let *AP* be a set of attacked possibility $AP = \left\{ap_x \mid ap_x = g\left(of_y\right)\right\}$. It reflects the possibility of attacked by a social engineering attacker. $g(of_y)$ is a function of positive correlation with $of_y$, and with $y \in \{e\text{-}mail, browse, download\}$ reflecting the operations of the user. The matchup relationship of $of_y$ and $ap_x$ has been explained in section 2.1.

**Definition 5:** Let *CP* be a set of compromised possibility $CP = \left\{cp_x \mid cp_x = h\left(sa_x, dd_x\right)\right\}$. It reflects the possibilities of compromised after attacked by a social engineering attacker. $h\left(sa_x, dd_x\right)$ is a function of negative correlation with $sa_x$ and $dd_x$.

**Definition 6:** Let *SoEV* be a set of social engineering vulnerabilities *SoEV*= {*soevx*|*soevx*=*apx* * *cpx*}. It reflects the likelihood that users will be compromised by specific means of social engineering attack. Vulnerability is the product of attacked possibility and compromised possibility.

**Definition 7:** Let *SoER* be a set of social engineering risk $SoER = \left\{soer_x \mid soer_x = k\left(ga, soev_x\right)\right\}$. It reflects the potential loss caused by a social engineering attacker. *ga* means authority the attacker may gain after a successful attack, and $k(ga,soev_x)$ is a function of positive correlation with *ga* and $soev_x$.

**Definition 8:** Let *Wei* be a set of weighting factors *Wei* = {$wei_{phish}$, $wei_{waterholing}$, $wei_{malware}$, *and* $wei_{pop\text{-}up}$}. The weighting factors are used to assign different weights to each type of social engineering attack for specific needs. The values may be given according to the historical records of enterprise attacked by social engineering attackers, and the importance of each type of social engineering attack.

**Definition 9:** Let *cv* be composite social engineering vulnerability. It indicates the possibility of compromised by all types of social engineering attack. *cv* is derived from the cumulative vulnerability of social engineering vulnerabilities according to the weights, and the formula is:

$$cv = \sum wei_x * soev_x \tag{1}$$

**Definition 10:** Let *cr* be composite social engineering risk. It indicates the composite loss caused by all types of social engineering attack. *cr* is derived from the cumulative risk of social engineering risks according to the weights,and the formula is:

$$Cr = \sum wei_x * soer_x \tag{2}$$

In the above definitions, the authors just give the parameters of function *g(), h()* and *k()*, and the correlation between parameters and results. The selection of these functions can be determined according to the actual situation of specific application scenarios.

## USER PROFILING

To analyze social engineering vulnerabilities, a comprehensive, accurate and multi-dimensional description of users is needed. User profiling has been widely applied in many fields in recent years. By analyzing user's network behaviors, business behaviors and so on, user profiling establishes formalized representation of the user in each feature, so as to build user model to achieve precise commodities or services pushing, or assistance to network security analysis.

User profiling is closely linked to social engineering. Social engineering is a discipline to study human vulnerability, and the attacker exploit human vulnerability to manipulate victims to take specific actions. User profiling describes the outline of user, and is able to discover human vulnerabilities which could be exploited by attackers. Social engineering can be regarded as human vulnerability exploiting, while user profiling is the mining of human vulnerabilities. Therefore, user profiling is more targeted compared with the analysis techniques whose objects are network or computer in analysis of social engineering vulnerability.

### User Profiling for Social Engineering Risk Analysis

The process of user profiling is as follows. First, collected behavior data is preprocessed, and then the individual behavioral characteristics are extracted. Classification techniques can be used to classify

users into groups and describe each group according to the behavioral characteristics. Clustering techniques are helpful to put new users into known groups, and make reasonable explanations to the groups, so as to find out valuable user groups and predict individual behaviors.

For the four types of social engineering attack aimed at, users could be grouped based on operating frequency characteristics, and the vulnerability and risk of corresponding attack pattern would be calculated. At present the relevant academic achievements have proved that with the help of network traffic analysis, operating frequency characteristics can be obtained by association rules mining and sequence pattern mining to find out the most frequently used type of service or website page. In [15] the writer extracts the e-mail using frequency and text features to improve intelligent recommendation system. Gottlieb and Lorimor (2017) the writer extracts frequency of browsing specific website and type of advertisement which is viewed mostly through cookie data collection for the establishment of user portrait. (Ma et al., 2016) presents an algorithm named LED to discovery overlapping community in complex network based on structural clustering. These results show that it is feasible to extract the features mentioned above and make user models by existing techniques. Therefore, specific technical scheme of user profiling wouldn't be discussed in detail in this paper.

### User Profiling Definition

User profiling is defined as follow:

**Definition 11:** Let *UP (ua, OF, SA)* be a user profiling, which consists of three elements, specifically user authority *ua*, related operating frequency *OF*, and security awareness *SA*. $ua \in (0,1]$ is user authority, whose higher value indicates that user's access and management level to the network resources is greater. *ua=1* indicates that user has full authority. In practice, *ua* can be assigned by the ratio of resources which user can access and manage to total resources. Since user authorities are difficult to identify as features when performing user profiling, manual tagging is required after completion of the grouping.

## CALCULATION OF SOCIAL ENGINEERING VULNERABILITY AND RISK BASED ON USER PROFILING

According to the above definitions, inputs of the framework is user profiling *UP*(*ua, OF, SA*), and outputs are possibility of being attacked *AP*, possibility of being compromised *CP*, social engineering vulnerability *SoEV*, social engineering risk *SoER*, composite vulnerability *cv* and composite risk *cr*.

Each parameter and function is selected as follows.

### Operating Frequency *OF*

In order to avoid large numerical difference of operating frequency among different user groups, the ratio method is applied to value assignment. First take the average operating frequency of all users in the group, then set the parameter value of group whose arithmetic mean value is the highest to 1, and the rest groups' parameter values are set to the ratio of its arithmetic mean value to the highest value. The parameter values are used to calculate *AP*.

### Security Awareness SA

Security awareness reflects the possibility of recognizing and making efforts to prevent social engineering attacks. User's attributes such as profession and training times of social engineering can be used to assess *SA*. Obviously, the more relevancy of profession to network security, and the more training times, the greater security awareness there will be. Furthermore, users with elevated authority would be more careful and less susceptible to social engineering attacks. *sa* should be limited as *sa*∈*[0,1]*. *sa=0* means that user has no idea about specific type of social engineering attack, while

*sa=1* means that user can recognize most of attacks. In practice, the actual value of *sa* can be decided by instruments such as questionnaire survey, induction training or social engineering testing. While profiling, the average *sa* value of group members can be set as the *sa* value of the group.

### Defense Degree *DD*

The defense degree values are given based on the probability that the protection mechanism prevents the attack when the user is attacked by social engineering attacker in the practical application scenario. Theoretically there is a value range that $dd_x \in [0,1]$. When $dd_x = 1$, it indicates that the protection mechanism has absolute protection against specific type of social engineering attack, and the probability of compromised is zero. But because for any protective mechanism there will be false negative, so the defense degree $dd_x \in [0,1)$ in fact. When $dd_x = 0$, it shows that there is no protective mechanism against such type of attack. In practice $dd_x$ can be determined by historical data, e.g. the proportion of times of successful prevention and total attacks in a given period (e.g. one mouth), or use the data provided by the manufacturer.

### The Authority That an Attacker May Gain ga

The value range of *ga* is $ga \in [ua,1]$, and the definite value is determined by the authority of the attacked user. For example, if the attacked user has the highest authority, the attacker could directly gain the highest authority after a successful attack, at the moment $ga = ua = 1$. Otherwise, the attacker would first gain the same authority with the attacked user, but because of potential privilege escalation exploit, the attacker may be promoted to a higher authority. The probability of successful privilege escalation is unknown, so at this time $ga \in (ua,1]$. Set $ga = \mathrm{l}(ua)$, and $l(ua)$ is of positive correlation with *ua*.

### Weighting Coefficient Wei

Wei is used to distribute weightings for all types of social engineering attack according to the actual needs when calculate the composite vulnerability cv and composite risk cr. The constraint condition of *Wei* is assigned as follows: $0 < wei_x < 1$, $\sum wei_x = 1$:

Function $ap_x = g(of_y)$, $cp_x = h(sa_x, d_x)$, $soer_x = k(ga, soev_x)$ and $ga = \mathrm{l}(ua)$

The four functions are used to calculate *AP*, *CP*, *SoER* and *ga*. The correlations with their respective parameters have been defined above. With the purpose of introducing risk analysis of social engineering attack into network security assessment, the four functions can be selected according to the network security assessment technique used in practice. In particular, when calculate the possibility of compromised, if $dd_x=0$, $cp_x$ is only related to the security awareness $sa_x$. And if $sa_x = 0$, which means the user has no security awareness at all, it's regarded that any false negatives from protection mechanism will lead to be compromised, so as to set $cp_x$ to $\infty$.

## SIGNIFICANCE AND USE CASES

In recent years, exposed APT incidents show that enterprises, institutions and government departments lack necessary awareness and means to prevent social engineering attacks. Nowadays, security mechanisms such as intrusion detection system and e-mail filter system have been popularized in large scale, and network security assessment techniques such as attack tree and attack graph have been widely used. However, currently there is no standardized technique system that can predict and prevent social engineering attacks. Our framework provides a way to quantitative analysis the

possibility of being compromised and potential loss caused by proposed social engineering attacks. By this way decision makers could regard users as network nodes, and social engineering attack surface as a new type of exploit when evaluate network security. Social engineering exploits and technical exploits can be processed equally, in order to achieve prediction of social engineering attack in the assessment phase.

The following two cases are given to illustrate how the framework is used for specific scenario and its application in network security assessment.

## Use Case 1: Usage for Specific Enterprise Scenario

In order to illustrate influence of parameters on calculation results and significance of calculation results, the authors give the first case as follow. For portable calculation, comparison and explanation of results, functions are taken as follows:

$g(x) = x$
$h(x, y) = (1-y) / x$
$f(x, y) = x * y$
$l(x)=(1+x)/2$

Assume all roles of the enterprise include executive administrator, network administrator, and staff. By user profiling, all personnel can be divided into five groups, and specific classification and profiles are shown in Table 2.

As seen from Table 2, in this enterprise, executive administrators and network administrators have entire user authority, meaning $ua = 1$, and the best security awareness. Executive administrators often need to send and receive e-mail and access to industry related websites, and install software from network administrators without need of downloading software from the Internet. Network administrators rarely e-mail, but often need to access related websites and download kinds of work or security software for availability and security of the network. Non-management staff can be divided into three groups according to entry time, in the order of older employees, employees with new work experience and new employees.

Staff group 1 compared to the other two groups with higher network authority, often visit industry related websites, and be of medium security awareness. Staff group 2 have less network access, and compared to staff group 1 are lack of awareness of malware and pop-up windows. Staff group 3 are in less familiar with the business stage, so the members less e-mail and access industry websites, and more need to install software without any security awareness.

Assuming that the enterprise deploys an e-mail filtering system whose phishing filtering probability is 0.6, and no other protection mechanisms. The social engineering vulnerabilities of each group is calculated according to Table 2, as shown in Table 3.

**Table 2. User profiling**

| User Groups | $ua$ | $of_{e\text{-}mail}$ | $of_{browse}$ | $of_{download}$ | $sa_{phish}$ | $sa_{waterholing}$ | $sa_{malware}$ | $sa_{pop\text{-}up}$ |
|---|---|---|---|---|---|---|---|---|
| Executive administrator | 1 | 1 | 1 | 0 | 1 | 1 | 0.8 | 1 |
| Network administrator | 1 | 0.1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Staff group 1 | 0.8 | 0.8 | 1 | 0.2 | 0.8 | 0.8 | 0.8 | 0.8 |
| Staff group 2 | 0.6 | 0.8 | 0.8 | 0.2 | 0.8 | 0.8 | 0.5 | 0.5 |
| Staff group 3 | 0.6 | 0.5 | 0.5 | 0.8 | 0 | 0 | 0 | 0 |

**Table 3. Social engineering vulnerabilities and risks**

| User Groups | $ga$ | $soev_{phish}$ | $soev_{waterholing}$ | $soev_{malware}$ | $soev_{pop\text{-}up}$ | $soer_{phish}$ | $soer_{waterholing}$ | $soer_{malware}$ | $soer_{pop\text{-}up}$ |
|---|---|---|---|---|---|---|---|---|---|
| Executive administrators | 1 | 0.6 | 1 | 0 | 1 | 0.6 | 1 | 0 | 1 |
| Network administrators | 1 | 0.06 | 1 | 1 | 1 | 0.06 | 1 | 1 | 1 |
| Staff group 1 | 0.9 | 0.6 | 1.25 | 0.25 | 1.25 | 0.54 | 1.125 | 0.225 | 1.125 |
| Staff group 2 | 0.8 | 0.6 | 1 | 0.4 | 1.6 | 0.48 | 0.8 | 0.32 | 1.28 |
| Staff group 3 | 0.8 | ∞ | ∞ | ∞ | ∞ | ∞ | ∞ | ∞ | ∞ |

Surveying the data in Table 3, executive administrators wouldn't be attacked by malware because of no downloading. Network administrators less e-mail, which makes them be not susceptible to phishing. Staff group 1 and group staff 2 are less likely to be attacked by malware, but easy to be attacked by pop-up windows when browse specific websites. Staff group 3 are likely to be attacked by any form of social engineering attack without any security awareness. The data in Table 2 is only suitable for the longitudinal comparison, which means the comparison of multiple user groups for vulnerabilities and risks of the same attack type, but is not suitable for the comparison of the same user group for vulnerabilities and risks of different types of attacks. For example, relative to phishing, the employees are more vulnerable to waterholing and pop-up windows. However, in fact, possibility of being attacked by phishing may be greater than by pop-up windows due that the frequency of e-mailing are far greater than the frequency of access to industry websites. The premise of horizontal comparison is to set up weighting coefficient *Wei*.

Assume that the enterprise employees often need to communicate with customers via e-mail, and a large number of employees often visit industry news and technology websites, but less download software from the Internet. We can rank attack types from high threat level to low as phishing, waterholing, pop-up windows and malware, which means $wei_{phish} \geq wei_{waterholing} \geq wei_{pop\text{-}up} \geq wei_{malware}$. Setting $wei_{phish} = 0.5$, $wei_{waterholing} = 0.4$, $wei_{pop\text{-}up} = 0.05$, and $wei_{malware} = 0.05$, the composite vulnerability and composite risk of social engineering attack are calculated for each group. The results are shown in Table 4.

**Table 4. Composite vulnerability and composite risk**

| User Group | $cv$ | $cr$ |
|---|---|---|
| Executive administrators | 0.75 | 0.75 |
| Network administrators | 0.53 | 0.53 |
| Staff group 1 | 0.88 | 0.79 |
| Staff group 2 | 0.80 | 0.72 |
| Staff group 3 | ∞ | ∞ |

The results in Table 4 show that composite vulnerability and composite risk of network administrators are minimum due that phishing is the greatest threat, and network administrators least e-mail. Although staff group 1 e-mail and access industry websites less frequently than the two groups of administrators, their composite vulnerability and composite risk are greater because of less awareness. Composite vulnerability of staff group 2 is higher than that of executive

administrators. However, composite risk is less than that of executive administrators and staff group 1 due to lower authority.

The outputs of this framework can guide enterprises to strengthen the security measures against social engineering attack by the following ways:

1. Operating frequency *OF* is determined by the nature of user group objectively. If any vulnerability of a group is significantly higher than other groups, education of this group should be strengthened to improve the security awareness *SA*;
2. Strengthening authority management is an effective way to reduce the social engineering risks. With the premise of social engineering vulnerabilities invariability, set user authority *UA* minimum, and attacker may gain least *ga*, thus effectively reducing the risks *SoER*;
3. For user group whose composite vulnerability *cv* and composite risk *cr* are too high, safety audit should be regularly conducted to check whether members of the group have already been attacked. If the enterprise deploys intrusion detection or other security systems, *cv* can also be applied to these systems in order to reduce false negatives and false positives.

## Use Case 2: Application in Network Security Assessment

The analysis results can be applied to a variety of network security assessment techniques. Attack graph is demonstrated as an example.

Attack graph is a graph-based assessment technique of network security based on network topology, whose inputs are hardware and software exploits, and other relevant information. Attack graph shows attack paths which may be used by the attacker to reach target node from an outside node in a graphical way, and can calculate occurrence probability, cost and profit of each attack path. Our framework provides a pathway to constructing hybrid attack graph (Beckers et al., 2015) containing social engineering exploits and technical exploits by adding social engineering vulnerability and risk as inputs. Treating users as nodes of network, and social engineering attack surface as social engineering exploits, it is realized to construct hybrid attack graph to optimize the assessment results. Attribute attack graph is the mostly used type of attack graph, and usually can be defined as $AG = (C, T, E)$:

1. $C$ is a condition vertex set that contains pre-conditions $C_{pre}$, and post-condition $C_{post}$, indicating the permissions required to perform an exploit and the privileges that the attacker can obtain after a successful exploit;
2. $T$ is a set of exploit vertexes (or a set of atomic attack nodes) that represents exploits may be exploited by attacker. Each exploit vertex contains two attributes: attack difficulty *Dif* and attack gains *Gain*. The former is the probability of exploited, which can be used for the occurrence probability calculation of attack path. The latter reflects the gains, and can be used to calculate the profits and judge the attack tendency;
3. $E$ is an edge set that reflects the connection between condition vertexes and exploit vertexes.

With the help of outputs of our framework, attribute attack graph *AG* can be extended as follows:

1. Social engineering attack surface can be treated as social engineering exploit vertexes, and added to inputs of attack graph generation algorithm. At this time, exploit vertex set T contains social engineering exploit vertexes $T_{se}$ and technical exploit vertexes $T_{tech}$. Composite vulnerability *CV* can be used as the *Dif* attribute of social exploit vertex $T_{se}$ for the occurrence probability calculation of attack path. Composite risk *CR* can be used as *Gain* attribute of $T_{se}$ for attack proceeds calculation and judging attack tendency;

2. Social engineering attacks are often used to break through boundary of network in the condition that attacker is in the absence of target network access. So that the social engineering attacks can be regarded as without pre-conditions, and their post-conditions is determined by *ga*;
3. Positions of social engineering exploit vertexes $T_{se}$ in attack graph and their connections with technical exploit vertexes $T_{tech}$ and condition vertexes $C$ are determined by network devices operated by users.

Adding new type of exploit vertexes, some unreachable exploit vertexes in the original attack graph may become reachable, so as to find new attack paths. Occurrence probability of attack path may be corrected according to *Dif* and *Gain* of social engineering exploit vertexes, which makes security assessment results more perfect. In a similar fashion, social engineering vulnerability and risk can be applied to other network security assessment techniques, such as privilege graph, attack tree, fault tree, etc.

## CONCLUSION

The authors contribute a risk analysis framework for social engineering attack based on user profiling for quantization calculation of social engineering attack vulnerability and risk. Building user profile by extracting features related to social engineering attack, it's feasible to calculate the respectively vulnerability and risk of specific type of social engineering attack, the composite vulnerability and composite risk. Examples are given to illustrate usage in actual scenario, and application in network security assessment. The presented framework could be applied to kinds of network, such as cloud service and social network.

The framework is extensible. The authors investigate four types of social engineering attack, three profiling features and one defense factor. Within the framework more attack types, profiling features and defense factors can be extended to optimize the results or calculate unmentioned indicators. For example, women and users between the ages of 18 and 25 are more susceptible to phishing, so that gender and age can be considered while doing user profiling. Character may have an impact, e.g. people with more greed and curiosity are easier to be compromised (Hadnagy et al., 2015). The deployment of protection mechanism may drop user's vigilance, which appears as lower security awareness in the framework. In future works, the authors will further study the user features related to other types of social engineering attacks, and optimize the framework for specific security assessment or protection techniques.

## ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China [grant number 61602515, 61501515], and the Foundation of Science and Technology on Information Assurance Laboratory [grant number KJ-17-001].

## REFERENCES

Abraham, S., & Chengalur-Smith, I. S. (2010). An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society*, *32*(3), 183–196. doi:10.1016/j.techsoc.2010.07.001

Beckers, K., Krautsevich, L., & Yautsiukhin, A. (2015). Analysis of social engineering threats with attack graphs. *Proceedings of the International Workshop on Quantitative Aspects in Security Assurance* (pp. 67-73). Springer, doi:10.1007/978-3-319-17016-9_14

Chandra, J. V., Challa, N., & Pasupuleti, S. K. (2015). Intelligence based defense system to protect from advanced persistent threat by means of social engineering on social cloud platform. *Indian Journal of Science and Technology*, *8*(28), 1.

Foozy, F. M., Ahmad, R., Abdollah, M., Yusof, R., & Mas'Ud, M. (2011). Generic taxonomy of social engineering attack. *Proceedings of the Malaysian Technical Universities International Conference on Engineering & Technology,* Batu Pahat, Johor (pp. 527-533). Academic Press.

Gelfand, M., Michel, Z., & Polonsky, N. (2016). U.S. Patent Application No. 14/996,329.

Gottlieb, G., & Lorimor, T. G. (2017). U.S. Patent No. 9,646,095. Washington, DC: U.S. Patent and Trademark Office.

Hadnagy, C., Fincher, M., & Dreeke, R. (2015). *Phishing dark waters: the offensive and defensive sides of malicious emails*. John Wiley & Sons. doi:10.1002/9781119183624

Jaafor, O., & Birregah, B. (2016). Multi-layered graph-based model for social engineering vulnerability assessment. *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp.1480-1488). IEEE.

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security & Applications*, *22*(C), 113–122. doi:10.1016/j.jisa.2014.09.005

Kuyoro, S. O., Ibikunle, F., & Awodele, O. (2011). Cloud computing security issues and challenges. *International Journal of Computer Networks*, *3*(5), 247–255.

Legg, P. A., Buckley, O., Goldsmith, M., & Creese, S. (2017). Automated insider threat detection system using user and role-based profile assessment. *IEEE Systems Journal*, *11*(2), 503–512. doi:10.1109/JSYST.2015.2438442

Ma, T., Wang, Y., Tang, M., Cao, J., Tian, Y., Al-Dhelaan, A., & Al-Rodhaan, M. (2016). Led: A fast overlapping communities detection algorithm based on structural clustering. *Neurocomputing*, *207*, 488–500. doi:10.1016/j.neucom.2016.05.020

Meer, H., Arvanitis, N., & Slaviero, M. (2009). *Clobbering the cloud*. Black Hat, USA.

Mitnick, K. D., & Simon, W. L. (2001). *The Art of Deception: Controlling the Human Element of Security*. John Wiley & Sons, Inc.

Monika, S., Anand, C., & Gnanamurthy, R. K. (2016). Analyzing the User Profile Linkage across Different Social Network Platforms. *International Journal on Computer Science and Engineering*, *4*(2), 1378–1383.

Mouton, F., Malan, M. M., Leenen, L., & Venter, H. S. (2014, August). Social engineering attack framework. *Proceedings of the 2014 Information Security for South Africa* (pp. 1-9). IEEE.

Nurse, J. R. C., Erola, A., Gibson-Robinson, T., Goldsmith, M., & Creese, S. (2016). Analytics for characterising and measuring the naturalness of online personae. *Security Informatics*, *5*(1), 3. doi:10.1186/s13388-016-0028-1

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 373-382). ACM. doi:10.1145/1753326.1753383

Tayouri, D. (2015). The human factor in the social media security – combining education and technology to reduce social engineering risks and damages. *Procedia Manufacturing*, *3*, 1096–1100. doi:10.1016/j.promfg.2015.07.181

*Ziwei Ye is currently a PhD candidate at State Key Laboratory of Mathematical Engineering and Advanced Computing, China. His research interests include network security assessment and social engineering.*

*Yuanbo Guo received a PhD degree in Computer Science from Xidian University, China, in 2005. He is a Member of the IEEE, Senior Member of the Chinese Institute of Electronics and a senior member of the Chinese Computer Federation. His research interests include insider threat detection, security analytics and Security architectures. He has published over 50 refereed articles in these areas and coauthored four books.*

*Ankang Ju studies cognitive security and network intrusion detection.*

*Fushan Wei is an associate professor at State Key Laboratory of Mathematical Engineering and Advanced Computing. His main research interests include cryptography, protocol analysis, and information security.*

*Ruijie Zhang is a lecturer at State Key Laboratory of Mathematical Engineering and Advanced Computing. Her main research interests include artificial intelligence, computer vision, and signal processing.*

*Jun Ma is a PhD candidate at Xidian University His main research interests include cryptography, computer network, and information security.*