Journal of Forensic Research and Criminal Investigation
JFRCI, 1(2): 35-42
www.scitcentral.com

Original Research Article: Open Access

A Taxonomy for Social Engineering Attacks in Twitter

Khalid Alissa[1*], Tarfah Al Sultan[2] and Nazar A

[1*]*Department of Computer Science, College of Computer Sciences & Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia.*

[2]*Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia.*

[3]*Saqib Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia.*

Received June 11, 2020; Accepted June 15, 2020; Published July 26, 2020

ABSTRACT

Social engineering is one of the biggest threats to information security. This avenue of attack is affecting the Saudi community in different ways. Recent studies show that social media in general and specifically Twitter has become one of the biggest internal threats to national security. Today, social engineers use Twitter to perform their attacks due to the large number of Twitter users in Saudi Arabia. This study proposes a taxonomy that identifies and categorises social engineering (SE) attacks in Saudi Twitter, based on an extensive review and analysis of the collected data. The proposed taxonomy defines (SE) attacks using five main entities: social engineer, victim, goal, attack method, and persuasion principles. As the taxonomy focuses on Social Engineering Attacks in Saudi Twitter (SEAST), it was named SEAST taxonomy. The paper also shows case studies using real life SE attack scenarios from Saudi Twitter that are identified and classified using SEAST taxonomy.

Keywords: Social engineering, Twitter, Taxonomy, Phishing, Information security

Abbreviations: SE: Social Engineering; KSA: Kingdom of Saudi Arabia; SEAST: Social Engineering Attacks in Saudi Twitter; SMS: Short Messaging Service; XSS: Cross-SiteScripting

INTRODUCTION

Recent studies on Saudi national security have found that social media has become one of the biggest internal threats for national security in Saudi Arabia [1]. In 2018, there were over 13.8 million Twitter users in KSA [2]. This large number raises questions about the security and privacy of Saudi Twitter users. Despite the great benefits provided by Twitter, it is an ideal place for attackers to perform social engineering attacks.

Nowadays, social engineering is one of the biggest threats to information security [3]. "It is the art of tricking people to gain information from them or persuade them to perform an action that will benefit the attacker in some way"[3]. In other words, by using social engineering, hackers can exploit users with deception to persuade them to accept the attack; it mainly depends on brain manipulation and deception.

Recent studies have focused on the different techniques and types of social engineering attacks in social networking sites. As far as this author knows, there is no study that focusses on social engineering in Saudi Twitter. It remains unclear how social engineers trick Saudi Twitter users and persuade them to perform some actions that might affect individuals or the whole country. The purpose of this study is to analyze social engineering attacks in Saudi Twitter feeds and propose a taxonomy of the issue.

This paper is organized as follows. The next section presents a review of the literature on social engineering attacks in social media sites. Section 3 presents the SEAST taxonomy and explains how it can be used. Section 4 shows case studies

Corresponding author: Khalid Alissa, Department of Computer Science, College of Computer Sciences & Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, E-mail: KaAlissa@iau.edu.sa

Citation: Alissa K, Sultan TA & Nazar A. (2020) A Taxonomy for Social Engineering Attacks in Twitter. J Forensic Res Criminal Investig, 1(2): 35-42.

using real social engineering attacks in Saudi Twitter. Section 5 states the future work. Section 6 is the conclusion of the paper.

LITERATURE REVIEW

Twitter is one of the major social networking platforms that provides free microblogging services and has attracted users around the globe to share their messages, their moments or posts, and to track comments of the other users and respond to their comments. At the same time, this platform has brought on several negative issues such as spamming [4,5], phishing [6], tampering [7], misinformation propagation [8], trend manipulation [9], online public shaming [10] and social engineering [11], to name few.

Social engineering is one of the biggest threats to individuals and organizations [11].

Social engineering attacks are difficult to detect and challenging to control [12]. It is about exploiting the weakest link in information security, which is people [13]. In a social engineering attack, the victim is deceived to share information with or perform some actions for the attacker [13].

In the literature, social engineering has been defined in different ways. Bezuidenhout [14]defines it as "breaking an organization's security by interactions with people". From another point of view, Algarni et al. [3] define it as "the art of deceiving people to help the attacker to reach his goal."

Researchers have tried to look at social engineering from different perspectives. Algarni et al. [3] believe there is a strong relation between social engineering and persuasion. They define social engineering attacks as "dishonest persuasion" [3]. They state that "social engineering is a form of manipulation that uses characteristics of persuasion to trick the victim" [3]. Simons [15] defines persuasion as "Human communication that is designed to influence others by modifying their beliefs, values, or attitudes". Manipulation and persuasion are similar to each other; they both use emotions to reach the goal [3]. Van Dijk [16] found that the main difference between manipulation and persuasion is that in persuasion the recipients have the freedom to believe the persuader or act as they want. Conversely, in manipulation the victim will act and follow the manipulator because the victim cannot understand the real intention of the manipulator and cannot see the consequences of believing them [16].

Researchers have presented different models and cycles to help understand social engineering attacks. The most commonly known social engineering attack model is Kevin Mitnick's model [17]. Mitnick's social engineering attack cycle starts with research, which is the information gathering phase where the attacker gathers information about the victim. The next phase is the development of rapport and trust with the target because once the trust exists, the victim is more likely to share information with the attacker. The third phase in the cycle is when the attacker exploits the trust and convinces the victim to share information with them or perform some action. At the end of the cycle, the attacker uses the outcome of the previous step and utilizes it. **Figure 1** illustrates Mitnick's model.

Figure 1. Kevin Mitnick's Social Engineering Attack Cycle.

Nowadays, social networking sites have become one of the most common means of social engineering attacks. Algarni [12] has discussed the entities and sub-entities that affect social engineering attacks in social networking sites. Based on the researcher's findings, there are four main entities that affect social engineering in social networking sites: environment, social engineer, plan and technique, and users **(Figure 2)**. Social networking sites are the environment and they help the attacker to easily reach the victim. The environment also helps the attacker to gather information about the user (the victim). Attackers can gather information about victims by accessing their public profiles if the victims do not change the privacy settings to make their profiles private, or by tricking victims by using psychological techniques to gain the victims' trust by establishing a friendship, and using embedded harmful contents like suspicious links in the published posts. The social engineer(attacker) is a critical entity that plays an essential role in the success of the social engineering attacks. The "plan" is the tactic and techniques used by the social engineer to trick the victim.

Studies on social engineering attacks show that social engineers use different techniques to persuade the victim. Cialdini's principle of persuasion is one of the most comprehensive principles used by social engineers. It contains six main persuasion techniques: authority, social proof, reciprocity, commitment, liking, and scarcity [18].

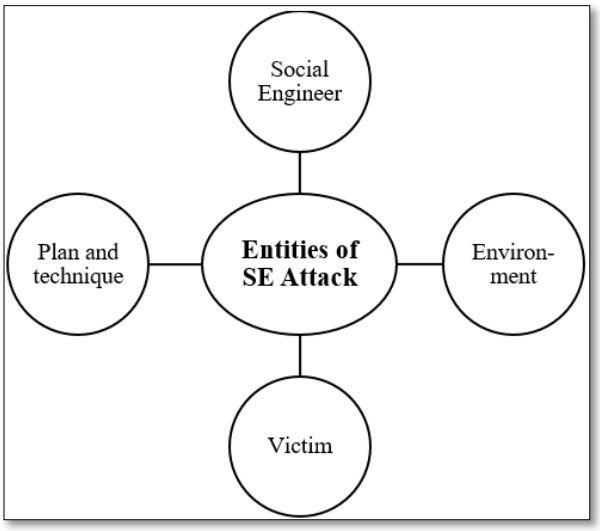

Figure 2. Entities of SE Attack.

Bullée et al. [13] presents the use of Cialdini's principles of persuasion by social engineers, and they define them as:

Authority: "the principle that describes people's tendency to comply with the request of authoritative figures" [13].

Social Proof: "the act of imitating the behavior of other people. Members of the in-group have a stronger feeling of group safety compared with members of the out-group" [13].

Reciprocity: "the giving of something in return"[13].

Commitment: "the likelihood of sticking to a cause or idea after making a promise or adhesion"[13].

Liking: "someone puts that person in a favorable position. People tend to like others who are similar in terms of interests, attitudes, and beliefs"[13].

Scarcity: "occurs when a product, service, or information has limited availability. People therefore perceive an increased value and attractiveness towards these products which makes them more desired than others" [13].

Ghafir et al. [19], believe social engineering attacks are classified into two main categories: physical and psychological. Hackers can use the physical location to perform the attack such as in the workplace, on the telephone, and online. In the psychological category, the hacker uses different techniques such as authority, neutral inclination to help, liking and similarity, commitment and consistency, reciprocation, and low involvement.

Mouton et al. [20] propose an ontological model of a social engineering attack. This model contains six main elements: social engineer, target, compliance principles, techniques, goal, and medium. In terms of the goal, the authors believe there are three main goals: financial gain, unauthorized access, and service disruption. For the techniques, they defined four main techniques: phishing, pretexting, baiting and quid pro quo. The model used Cialdini's principles of persuasion as the compliance principle. Based on the model, the medium of the attack can be e-mail, face-to-face, telephone, SMS, paper mail, storage media, webpage, and pamphlets. Finally, the target and the social engineer can be an individual or a group. They state that in order to define a type of social engineering attack, one must decide on one aspect of each element of the six main elements. For example, a social engineering attack could be: For financial gain by using phishing through email, and making use of the scarcity principle, where the attacker is individual, and the victim is also an individual.

To identify social engineering attacks that are specific to social networking sites, the method of defining them might be different. (The medium is always a social media site, so there is no need to include it in the definition.) There have been some efforts to try to define and categorize social engineering attacks on social networks. Algarni [12] identifies some specific types of social engineering attacks that happen by using social networking sites: phishing, spam, cross-site scripting (XSS), and defamation. In phishing, hackers publish stories or give offers to attract the victim to visit a link. Spam is another type that can be replaced in social networking sites with posts instead of e-mails. XSS can also be used in the social networking sites. Finally, defamation and ballot-stuffing are used in social networking sites to destroy someone's reputation [21].

Based on an another study [11], the previous categories might not be enough. They identified other forms of social engineering attacks in social networking sites [11]. These types are: identity theft, fake credentials, impersonation, copyright violation, content-based phishing, application-based phishing, interpersonal deception, dishonest and malicious contents [11].

It is evident from the literature that there is no study that defines a clear taxonomy of social engineering attacks in Twitter. There are many types of and techniques used in social engineering attacks in social networking sites, but to this author's knowledge in Saudi Twitter there are other goals and techniques used by social engineers that need to be investigated. Based on this fact, the following sections will present Social Engineering Attacks in Saudi Twitter (SEAST) taxonomy.

SOCIAL ENGINEERING ATTACKS IN SAUDI TWITTER (SEAST) TAXONOMY

Social engineering attacks in Saudi Twitter are affecting the Saudi community in different ways. This type of attack leads to spreading dangerous phenomena in the community, which in turn threatens Saudi national security. In a recent study in the national security, Alshehri [1] found that the new media and intellectual dissolution are classified as internal threats of Saudi national security [1].

Social media has become an ideal space for spreading rumours, and arousing the public opinion[1]. According to one statistic, in 2016 there were more than 10,000 fake Twitter accounts that were targeting Saudi national security [1]. Based on this fact, the authors have studied Saudi Twitter to analyse and categorize social engineering attacks in Saudi Twitter.

The study started by collecting data from Saudi Twitter. The first phase of data collection was done by manually collecting random tweets from the trending hashtags. The dataset reached 700 tweets, which were then analysed one by one. This initial data gave an indication of the size of social engineering attacks-it showed that 10% of the collected tweets contained a type of social engineering attack. Therefore, it was necessary to collect more data on a bigger scale.

The manual process of the data collection was time-consuming. For that reason, the second phase of the data collection was automated using Cloud services. The tools used in this phase were Google spreadsheets and Twitter Archiver add-on, which are provided by Google Drive to collect data automatically. The Twitter Archiver add-on tool accesses Twitter API and executes a specific rule created by the researcher to fetch tweets form the trending hashtags in Saudi, based on their geographical location. The add-on then fills in Google spreadsheets with the data. In the second phase, when the dataset reached more than 10,000 tweets, they were then analysed by the researcher manually. The analysis of 10,000 tweets confirmed that approximately 10% of the tweets contained a type of a social engineering attack.

When trying to categorise and analyse social engineering attacks, one of the most important aspects to look at is the goal of the attack [20]. Social engineers have different goals when they perform the attack. Mouton et al. [20]found that social engineers' goals can be financial gain, unauthorized access, or service disruption.

An analysis of the Saudi Twitter feed found that some attacks have goals beyond the three goals mentioned above. In Saudi Twitter, some of the social engineering attacks were aimed for political gain to threaten the national security with different techniques. One technique was to publish targeted fake news and pictures to spread fear among people. By publishing these rumors, people were convinced that their country was not a safe place and that national security was unstable. Another way was to target teenagers and inundate them with a devastating mindset such as: feminism and atheism to leave their homeland and families. These attackers were usually motivated by political reasons and had one goal, which was to affect and disrupt the whole country.

Another goal of social engineering attacks in Saudi Twitter was to discredit a specific person, where social engineers used their techniques to destroy someone's reputation. In 2018, a group of hackers published fake news about a Saudi deceased writer and they persuaded people to believe that she was killed by her father, thereby destroying her family's reputation. The group of hackers were publishing many tweets under a specific hashtag and letting people contribute with this story by the retweets and replies until that hashtag became a trend in Saudi during that time. In such attacks, the target is specific while the attacker could be one or more people.

Gaining fame is another goal that was a target of social engineering attacks in Saudi Twitter. This was done using different techniques, one of which was to raise controversial issues and let people contribute to them and have many replies, retweets and likes. Therefore, along with "financial gain, unauthorized access, or service disruption" [20] there are also three other goals: Political gain, fame gain, and discredit.

To achieve these goals, social engineers use different methods and techniques in social networking sites such as phishing and defamation [12]. Moreover, identity theft, impersonation, and establishing relationships are also other methods of social engineering attacks [11]. Based on the analysis of the Saudi Twitter feed, the researcher found that there are other attack methods used by social engineers, such as spreading rumours and fake news, taking advantages of social media influencers, and blackmailing. They use social media influencers to take the advantage of their fame and their effect on their followers. For example, they use them to spread hidden ideas indirectly to convince people to perform some actions to benefit the attacker. Therefore, there are seven main techniques that the attackers use: phishing, defamation, identity theft, impersonation, establishing relationships, taking advantages of social media influencers, spreading rumours and fake news, and blackmailing.

When social engineers use one of the attack methods, they take advantage of the persuasion principles that makes the victim comply with the attacker's request. Cialdini's persuasion principle [18] is the most common principle used by social engineers ,which contains six main persuasion principles: authority, social proof, reciprocity, commitment, liking, and scarcity [18].

From the dataset and based on the researcher's analysis, there are different persuasion principles used by social engineers in Saudi Twitter aside from those established in Cialdini's principles. One of their persuasion principles is to achieve their goal by gaining sympathy from others. They convince others that they are poor or sick to make people support them financially. Furthermore, social engineers reach their goal by exploiting people's curiosity, or by provoking people by publishing tweets that contain contents conflicting with their believes to gain their interaction. They also exploit people's fear of scandal by blackmailing them using their personal photos or words they said. So social engineering attacks in Saudi Twitter make use of the following principles: authority, social proof, reciprocity, commitment, liking, scarcity, sympathy, curiosity, provocation, and fear.

Social engineers and victims can be groups or individuals where a group of social engineers can target a specific person

or a group, and vice versa [20]. There is no simple way to define and categorize social engineering attacks in Saudi Twitter. The authors built a taxonomy that contains all the aspects specified earlier to categorize the attacks. Based on the literature review and the discussion above, the researcher introduced the Social Engineering Attacks in Saudi Twitter (SEAST) taxonomy that covers different social engineering attacks found in Saudi Twitter feed (**Figure 3** illustrates SEAST taxonomy). The taxonomy contains five main entities: social engineer, victim, goal, attack method, and persuasion principle. To categorize an attack, one has to define the goal of the attack, the attacking method, the persuasion principle, the type of the attacker, and the type of victim. The following section will show some examples on how to use SEAST taxonomy.

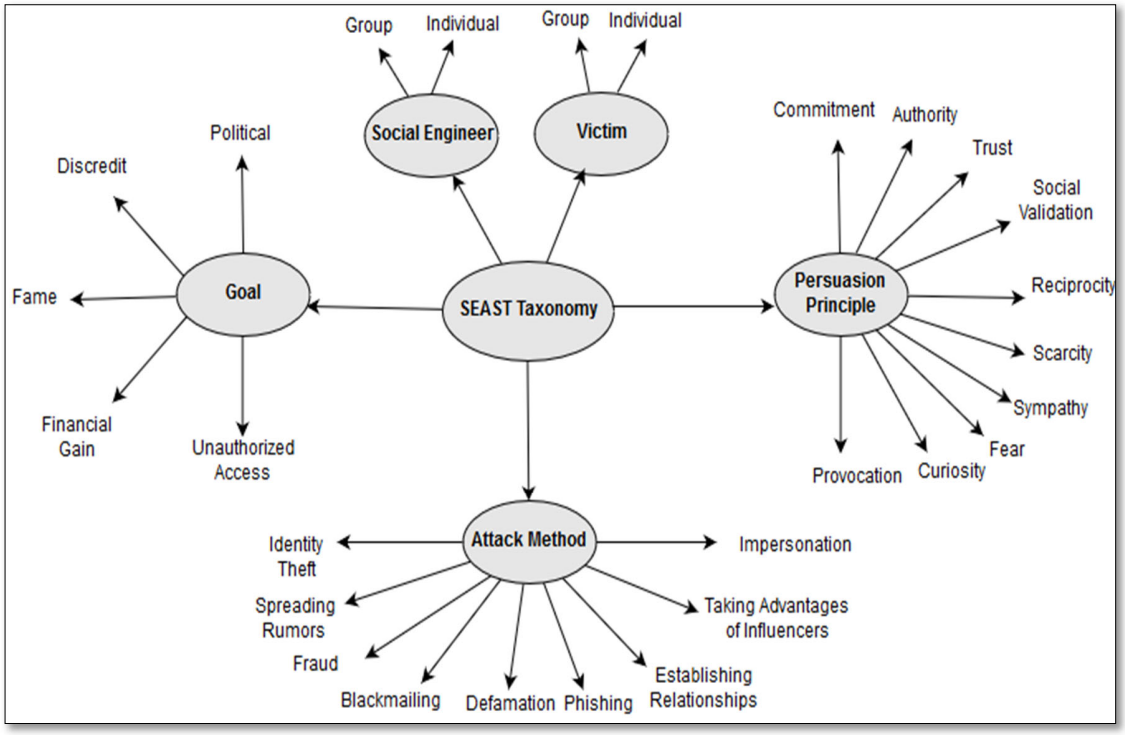

Figure 3. SEAST Taxonomy.

CASE STUDIES

This section identifies and categorizes different social engineering attacks in Saudi Twitter using SEAST taxonomy. This section will use real-life attacks that happened in Saudi Twitter and that used social engineering techniques. Using SEAST taxonomy, each social engineering attack is defined based on five entities: the social engineer, victim, goal, attack method, and persuasion principle. First, the scenario will be explained, then SEAST taxonomy will be applied to categorize the attack.

Scenario 1

A thief in Saudi Twitter used a fake account pretending to be a child who was fighting cancer, convincing others that he needed money for his cure and shared his bank account with people who wanted to help. Also, he posted pictures of his suffering as bed-ridden in a hospital and posted tweets that played on people's emotions such as, "I fight cancer with my smile," along with a picture of the cancerous child smiling. Based on SEAST taxonomy this social engineering attack can be defined as follows:

Social Engineer: Individual
Victim: Group
Goal: Financial
Attack method: Impersonation
Persuasion principle: Sympathy

Scenario 2

A thief in Saudi Twitter used a fake account pretending to be a manager in a company seeking employees. He was targeting female job seekers and contacted them to get their CVs and full information. Once he got this information, he contacted the victim to inform them about their acceptance in the job to gain their trust. After that, he asked for their personal photos in which they were not fully clothed for the purpose of completing the employment process. Once he got that, he blackmailed them by their photos and personal information

and asked for money. Based on SEAST taxonomy this social engineering attack can be defined as follows:

Social Engineer: individual
Victim: individual
Goal: financial
Attack method: blackmailing
Persuasion principle: fear

Scenario 3

A Saudi girl was convincing others via her Twitter account that she was violated by her family. She posted fake pictures of violence, and asked others for support. Some people were convinced by this idea and supported her by retweeting her tweets and mentioning the human rights commission to provide help for her. She was trying to manipulate people's brains and letting them believe that her country did not support her, so many people were convinced and her story was a trend in Saudi Twitter. After a while, the girl migrated to another country claiming that that country was a safer place and would provide her freedom and peace. After her migration, she was posting pictures of her new life, and she was describing how happy she was. Investigations showed that all of this scenario was supported by other entities who aimed to threaten Saudi national security by convincing teenagers that their country was not capable of providing them their rights and convinced them to migrate. Based on SEAST taxonomy this social engineering attack can be defined as follows:

Social Engineer: group
Victim: group
Goal: political
Attack method: spreading rumours
Persuasion principle: sympathy

Scenario 4

A Twitter user posted a tweet that contained a picture of a phone keypad along with words such as money, food, friends, travelling, shopping, coffee, laugh, love, and music, where each number on the keypad represented a word (as shown in **Figure 4**). He wrote, "Your mobile password specifies the most things you need in your life." Many people were curious so they replied to that tweet and shared their passwords using their needs to have fun. Based on SEAST taxonomy this social engineering attack can be defined as follows:

Social Engineer: individual
Victim: group
Goal: unauthorized access
Attack method: phishing
Persuasion principle: curiosity

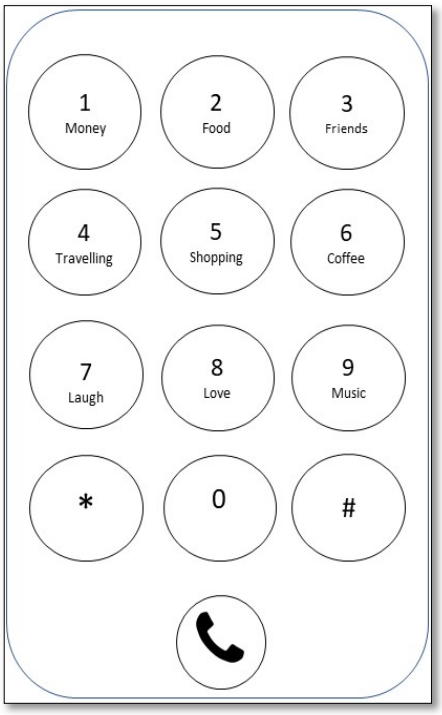

Figure 4. Phone keypad along with words.

Scenario 5

In Saudi trending hashtags, there were many tweets that contained fake news and a link. The tweet didn't contain the full news to let the user be curious and visit the link to read the more about the news. Instead, once the user visited the link, he could not find any useful information, rather he found an add or unrelated materials. By this way, the attacker reached his goal, which was making users click and visit the link where the attacker benefitted from this action financially due to the number of visits. Based on SEAST taxonomy this social engineering attack can be defined as follows:

Social Engineer: individual
Victim: group
Goal: financial
Attack method: phishing
Persuasion principle: curiosity

Scenario 6

A hacker ran a Twitter account as a bank customer service, and he disguised himself using the logo of the bank. When people contacted that account for support, the hacker claimed that he could not solve the issue for the user until he got the user's bank credentials. Once the hacker got the user's bank credentials, he deleted his account. Based on SEAST taxonomy this social engineering attack can be defined as follows:

Social Engineer: individual
Victim: group

Goal: financial
Attack method: impersonation
Persuasion principle: trust

Scenario 7

A hacker published a fake link in Twitter that redirected users to a web tool that asked for the user's Twitter account credentials for the purpose of giving the user information and statistics about who was visiting their profile every day. The hacker exploited people's curiosity about who visited their profile, read their tweets, and cared about them. In this way, the hacker stole the user's account. Based on SEAST taxonomy this social engineering attack can be defined as follows:

Social Engineer: individual
Victim: group
Goal: unauthorized access
Attack method: phishing
Persuasion principle: curiosity

FUTURE WORK

For future work, the proposed taxonomy can be used to automate the detection of social engineering attacks in Saudi Twitter feed. Using machine learning algorithms and SEAST taxonomy, a model can be developed to detect social engineering attacks in Saudi Twitter in an automated way. The developed model will make use of the categories and definitions identified by SEAST taxonomy.

CONCLUSION

Security issues are one of the most important issues that also takes the biggest attention of countries' leaders. Social media and especially Twitter are constituting a real threat on the national security level as it is being used for publishing extremist thoughts and terrorism. This study analyzed social engineering attacks in Saudi Twitter feed and proposed a taxonomy of social engineering attack in Saudi Twitter (SEAST).

As a result of this study, there are five main entities that can be used to classify a social engineering attack in Saudi Twitter: social engineer, victim, goal, attack method, and persuasion principle. The victim and the social engineer can be an individual or a group. The goal can be political, financial, gaining unauthorized access, discrediting, and gaining fame. For the attack method, it can be identity theft, spreading rumors, fraud, defamation, phishing, establishing relationships, taking advantages of influencers, impersonation, and blackmailing. There are ten persuasion principles used by the hacker to reach his goal: Commitment, authority, trust, social validation, reciprocity, scarcity, sympathy, curiosity, provocation, and fear. Finally, section 4 of this study showed real-life attack scenarios, and analyzed and categorized them using SEAST taxonomy.

AUTHORS' CONTRIBUTIONS

Authors have contributed equally to this paper. They all have participated in identifying the issue and developing the solution as well as the writing.

REFERENCES

1. Alshehri R (2018) An outlook on saudi national security under the 2030 visionرؤية استشرافية للأمن الوطني للمملكة العربية السعودية في ضوء رؤية 2030,

2. Leading countries based on number of Twitter users as of April 2018 (in millions) (2018). Accessed on: October 01, 2018. Available online at: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.

3. Algarni A, Xu Y, Chan T, Tian Y (2013) Toward understanding social engineering. Proc 8th Int Conf Leg Secur Priv Issues IT Law: 279-300.

4. Wu T, Wen S, Xiang Y, ZhouW (2018) Twitter spam detection: Survey of new approaches and comparative study. Comput Secur 76: 65-284.

5. Adewole KS, Han T, Wu W, Song H, et al. (2018) Twitter spam account detection based on clustering and classification methods. J Supercomput.

6. Liew SW, Mohd Sani NF, Abdullah MT, Yaakob R, et al. (2019) An effective security alert mechanism for real-time phishing tweet detection on Twitter. Comput Secur 83: 201-207.

7. Pfeffer J, Mayer K, Morstatte F (2018). EPJ Data Science 7, pp: 50.

8. Jin F, Wang W, Zhao L,Dougherty ET (2014) Misinformation Propagation in the Age of Twitter. Computer 47: 90-94.

9. Zhang Y, Ruan X, Wang H, Wang H, He S (2017) Twitter trends manipulation: A first look inside the security of twitter trending. IEEE Transact Info Forens Secur 12: 144-156.

10. Basak R, Sural S, Ganguly N, Ghosh SK (2019) Online Public Shaming on Twitter: Detection, Analysis, and Mitigation. IEEE Transact Comput Soc Sys 6: 208-220.

11. Algarni A, Xu Y, Chan T, Tian YC (2014) Social engineering in social networking sites: Affect-based model. ICITST 2013: 508-515.

12. Algarni A (2013) Social engineering in social networking sites: Phase-based and source-based models. Int J e-Education e-Business e-Management e-Learning: 508-515.

13. Bullée JH, Montoya L, Junger M, Hartel P (2018) On the anatomy of social engineering attacks - A literature-based dissection of successful attacks.

14. Bezuidenhout M (2010) Social engineering attack detection model: SEADM 2010. Inf Secur South Africa: 1–8.

15. Simons HW, Jones JG (2010) Persuasion in Society.

16. Teun Adrianus van Dijk (2008) Discourse and power. Palgrave Macmillan UK.

17. Mouton F, Malan MM, Leenen L, Venter HS (2014) Social engineering attack framework. Inf Secur South Africa 2014: 1-9.

18. Cialdini RB (2009) Influence.

19. Ghafir I (2016) Social engineering attack strategies and defence approaches. IEEE 4th Int Conf Futur Internet Things Cloud: 145-149,

20. Mouton M, Leenen L, Venter HS (2016) Social engineering attack examples, templates and scenarios. Comput Secur 59: 186-209.

21. Cutillo LA, Molva R, Strufe T (2009) Safebook: A privacy-preserving online social network leveraging on real-life trust. IEEE Commun Mag 47: 94-101.