THE POWER OF INTENTION IN DETECTING SOCIAL ENGINEERING ATTACKS

Ibrahim Mohammed Alseadoon *

Department of Information and Computer Science, College of Computer Science and Engineering, University of Ha'il, Hail
Kingdom of Saudi Arabia

* Corresponding Author, e-mail: i.alsedon@uoh.edu.sa

Abstract: Social engineering attacks are increasing over the years along with the growth in Internet users. Our study aims to find ways to improve users' protection against these attacks. Finding a suitable way to enhance users' detection will have the impact of reducing the number of victims. Our study uses quantitative methods including experiment (role play) and a questionnaire to collect data. The findings suggest that raising users' intention to validate the authenticity of received messages will increase their ability to make a robust decision about deceptive messages. The detection rate significantly increased when users have been alerted about social engineering attacks.

Key words: social engineering attacks, human behaviour, awareness, deception, detection.

1. INTRODUCTION

Social engineering attacks have been increasing over the years [1]. Additionally, the financial cost on users and organisations is enormous [2]. Social engineering attacks are designed to build a relationship with victims to lure them into giving information to attackers or performing an action that will result in financial loss or malicious impact [3]. Since protecting users result in protecting organisations, our study focuses on improving users' detection mechanism.

One approach to stop the harm from social engineering attacks is to prevent them from reaching users through technical tools. These tools have shown that they can detect the majority of social engineering attacks [4]. However, some attacks manage to escape detection and end up reaching users. In this situation, users will have to decide as to whether the received messages are legitimate or not. Users with low experience in detection will be more vulnerable to social engineering attacks.

Therefore, many studies have proposed cybersecurity education tools to improve users' detection ability.

Hence the second approach is educating users. Many education studies have shown that users are more resistant to social engineering attacks following a proposed cybersecurity education course [5-7]. Some have claimed that users were able to retain cybersecurity information for a long period of time to be able to detect and protect themselves from such attacks [7]. However, other studies have shown that cybersecurity education may not be a suitable defence mechanism against social engineering attacks [8]. Cybersecurity education courses teach users to identify certain types of social engineering attacks; but it is common knowledge that social engineering attacks are always adapting and improving their techniques. Users will then be vulnerable to new attack types. A study on users' ability to detect scams found that users were able to detect familiar scams but not unfamiliar ones [9]. Moreover, the same principle applies to cybersecurity tools; as attack techniques progress, these tools will no longer be able to identify them.

Our study claims that an understanding of the users themselves in the process of detection has received less attention. The users' personality plays an important role in making users into detectors [10, 11]. Supporting users with technical and educational tools will not prevent them from becoming victims if they are careless in making reliable decisions about fake messages. Our study claims that users' intention to detect deception is a main factor in making users more reluctant to respond to fake messages. Cybersecurity technical and educational tools will assist users' decisions, but the trigger to identify fake messages start from the users' attention.

Many studies in the area of detection have not studied the impact of warning in users' ability to detect [12]. The focus was instead on educating and training users to detect deception. The literature on deception has found conflicting results about the effect of education. For example, some studies [13-15] found that training did not have an important effect on users' ability to detect deception. On the other hand, other studies [6, 7, 16] found that education has a significant effect in improving users' ability to detect deception. Moreover, the same conflict in findings can be seen in demographic data; for example, some studies found that gender and age have an impact on users' detection ability [17, 18], while other studies have found no significance [19, 20]. There is something missing between these studies. The effectiveness of education in the realm of cybersecurity deception needs more investigation. Our study aims to resolve this conflict and finds the trigger that make users better detectors.

Our study shows the importance of making users more engaged in making a reliable decision about received social engineering attack messages. If users were careless in decision making, security and education tools are not sufficient enough to prevent attacks from being successful.

2. THE MODEL

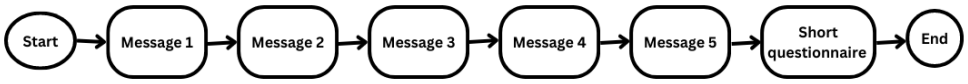
The Elaboration Likelihood Model (ELM) suggest that users have two main decision-making routes: peripheral decision-making (the peripheral route) and deep decision-making (the central route) [21]. It has been claimed that users are more likely to fall victim to social engineering attacks if they are using the peripheral decision-making route in judging messages [22, 23]. Our study claims that informing users about the potential risk of social engineering attacks will make them more likely to apply the deep decision-making route, which will increase their ability to detect fake messages. The hypothesis of our study is as follows:

H: Users' intention plays a key role in creating users' resistance in responding to social engineering messages.

Intention here means that users want to validate the authenticity of messages. Making users aware of the potential attack will lead them to apply their knowledge to better judge whether a message is legitimate or fake. Security education studies have found that educating users to identify fake messages increases their detection rate [5, 7, 24, 25]. However, education studies test users about fake messages which they already know how to recognise. In contrast, social engineering attacks are continuously evolving and will not use the same cues in their attacks. Users will rely on their current knowledge to determine whether the encountered message is fake or legitimate. Therefore, having the ability and the skills to authenticate messages is crucial.

3. EXPERIMENT

The experiment involved two groups: a control group with 34 participants, and an informed group with 28 participants. The informed group were told that some of the messages are social engineering attacks without specifying the type or severeness. The experiment is a role play in which the participants were shown five messages in the form of images as follows: two legitimate messages and three social engineering attack messages. The difficulty of discerning legitimacy went from easy to hard. The first message was a legitimate message from a legitimate organisation asking for participation in a survey. The second message was a fake message that impersonates a legitimate government organisation, informing the receiver that his\her status has been changed. The third message was a legitimate message from a legitimate charity organisation. The fourth message was a fake message that impersonates a legitimate government organisation, informing the receiver that s\he received a fine. The fifth message was a fake message that threats the receiver in a social media platform, which demands an immediate action to stop the threat (see Figure 1*)*.



1. *The experiment steps*

Participants were asked to rate the choice to click on a link provided in the message using a five-point Likert scale [26] whereby five is strongly disagree and one is strongly agree (see Table 1).

Table 1. Participants' response scale (click on included link)

No.	***Statement***
1	*Strongly agree*
2	*Agree*
3	*Neutral*
4	*Disagree*
5	*Strongly disagree*

There is a drawback in using the questionnaire in our study, which is the limited effect of perceived real damage. Participants cannot feel the danger of clicking on harmful links, and their assessment might be different in the real world. However, the first intention of response is the main target of our study. Users choosing whether or not to click on the links will make their decision based on the message cues that are presented in our study.

Participants were undergraduate students in their freshman year in computer studies. These participants were chosen because they are closer to average users who do not have advanced knowledge in computing, which would have the advantage of increasing users' perspective about cybersecurity. Participants were only male students. The experiment begins with the controlled participants then the informed ones, avoiding any chance of affecting participants perception or obtaining knowledge about detection techniques. The age group ranges from 18 to 25 as this is the normal undergraduate students as well as it has been reported to be more vulnerable to attacks [27].

4. ANALYSIS AND RESULTS

Preparing data for analysis is an essential step before conducting any analysis. Our study follows the preparation steps proposed by Fink [28]. SPSS software was used to test the hypothesis proposed by our study: "Users' intention plays a key role in creating users' resistance in responding to social engineering messages". The null hypothesis is that users' intention makes no difference to whether users become victims of social engineering attacks. This means that there will be no difference between controlled group and informed group. Users who do not have an intention to detect deceptive messages will not act accordingly. Therefore, deceptive messages that are not known or previously encountered will not be recognised as a deceptive

message. Users normally think that they are out of danger, or they are not targeted by attackers. Therefore, users' initial response to messages will be, "will I respond or not based on my desire" not by the authenticity of the message. Therefore, if users choose to ignore a deceptive message, they are safe not because of their detection abilities. Furthermore, deceptive messages that are ascribed a low level of danger (according to users) will not be avoided as a high-level perceived danger.

The first step done in our study was to reverse the response of two messages (i.e., legitimate messages). Then, factor analysis was conducted. The results obtained from factor analysis indicated that two factors were measured. Therefore, based on the analysis, one message response was removed (message number 2). Factor analysis was conducted again, and the results indicate that only one factor was measured. Subsequently, the reliability of the measure of the likelihood to click on a link in a message was tested using Cronbach's alpha. The acceptable value recommended by scholars is 0.7 [29]. The Cronbach's alpha achieved in our study was 0.750 for four items, which is acceptable (see Table 2).

Table 2. Descriptive statistics for messages

No.	***Message***	***Mean***	***Median***	***Skewness***
1	*Legitimate message 1*	*2.53*	*2.50*	*0.30*
2	*Fake message 2*	*2.55*	*2.00*	*0.47*
3	*Legitimate message 3*	*2.73*	*3*	*0.26*
4	*Fake message 4*	*4.32*	*5*	*-1.93*
5	*Fake message 5*	*4.60*	*5*	*-2.82*

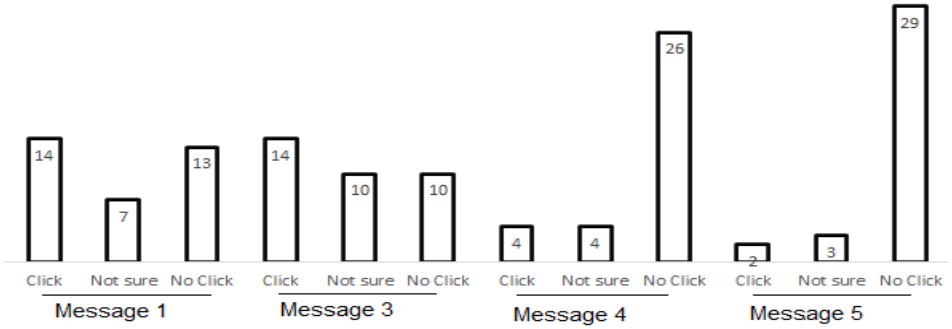
From Table 2, it can be seen that participants are more unlikely to click on links that are more suspicious and more likely to click on legitimate links, with the exception of message number two, which is a social engineering message. Surprisingly, participants were more willing to click on the link in message number two in both groups. It is not clear in our study why participants chose to click on the link. Message number two imitates a legitimate government organisation, which might be a reason for this outcome. This finding needs more investigation in a future work.

An additional goal of our study is to measure the impact of intention on accurate detection — i.e., whether users were able to identify fake messages as fake and legitimate messages as legitimate. To measure accurate detection, we firstly needed to reencode participants' responses to our experiment. As responses were measured in a five-point Likert scale, the encoding needed to be changed as follows: five and four were recoded as three; three was recoded as two; two and one were recoded as one. In the new encoding, number three means that participant will not click. Two means undecideds; therefore, these values were removed from the analysis. Lastly number one means the participant will click. Therefore, numbers three and one were included in calculating accurate detection (see Table 3).

Table 3. Reencoded messages for control group

Message	***1***			***3***			***4***			***5***		
Action	*Click*	*Not sure*	*No Click*	*Click*	*Not sure*	*No Click*	*Click*	*Not sure*	*No Click*	*Click*	*Not sure*	*No Click*
Code	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
No.	*14*	*7*	*13*	*14*	*10*	*10*	*4*	*4*	*26*	*2*	*3*	*29*

Table 3 shows that the number of undecided responses in the control group was 24, which represents around 18 precent of responses (see Figure 2).



The accuracy of detection was measured by applying the confusion matrix. Therefore, there are four categories: true positive, true negative, false positive and false negative. The results are shown below (see Table 4).

Table 4. Confusion matrix for control group

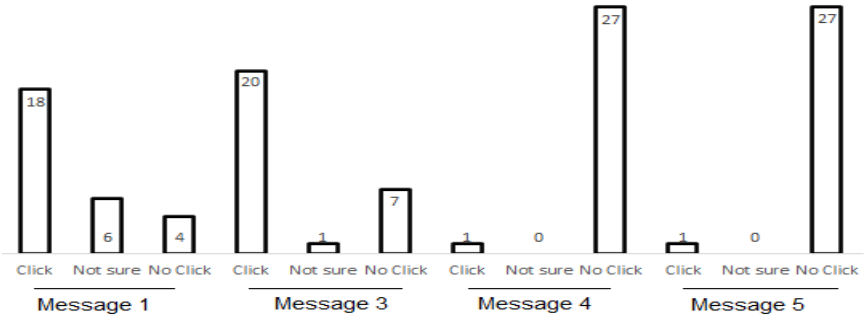
	positive	***Negative***
True	*28*	*23*
False	*6*	*55*

Table 4 shows that participants in the control group were able to accurately detect messages as legitimate and illegitimate at a rate of 74 percent and that 26 percent of participants made wrong decisions. Furthermore, the results show that participants are more willing to identify messages as legitimate than to identify them as fake (true negative 23). When participants are not sure about the authenticity of a message, they tend to classify them as legitimate (see Table 5).

Table 5. Reencoded messages for informed group

Message	**1**			**3**			**4**			**5**		
Action	*Click*	*Not sure*	*No Click*	*Click*	*Not sure*	*No Click*	*Click*	*Not sure*	*No Click*	*Click*	*Not sure*	*No Click*
Code	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
No.	*18*	*6*	*4*	*20*	*1*	*7*	*1*	*0*	*27*	*1*	*0*	*27*

Table 5 shows that the number of undecided responses in the informed group was seven responses, which represent around six percent of responses. It can be seen that participants are making robust decision about messages rather than leaving them without decision (see Figure 3).



Surprisingly, all participants except one decided not to respond to social engineering attack messages. Due to the anonymous nature of our study, it is not possible to follow-up with the participant who chose to respond. Additionally, the accuracy of detection for informed group was measured using the confusion matrix (see Table 6).

Table 6. Confusion matrix for informed group

	Positive	***Negative***
True	*38*	*2*
False	*26*	*54*

Table 6 shows that participants in the informed group were able to accurately detect messages as legitimate and illegitimate at a rate of 77 percent and found that 23 percent of participants made wrong decisions. Furthermore, the results show that participants were more willing to identify messages as fake than to identify them as legitimate. When participants are not sure about the authenticity of a message, they tend to classify them as fake. It can be said that participants were more cautious in

making decisions about undecided messages, in order to be safer than trusting unknown messages.

Finally, to test our study hypothesis, a t-test measure was applied between the two groups of our study. The results of the t*ests are shown in Table 7.

Table 7. Results from t-test

Message	***t-test***	***Sig.***
Legitimate message 1	*-2.24*	*0.029*
Legitimate message 3	*-3.49*	*0.001*
Fake message 4	*-2.64*	*0.011*
Fake message 5	*-1.14*	*0.260*

Table 7 shows that the results indicate a significant difference between the two groups in three messages. The last message (i.e., message 5), which is a threatening attack message, did not have any significant differences between the two groups. The reason might be related to the fact that the majority of participants identify message 5 as an attack message, since it uses direct threats toward users, which is an obvious cue for a social engineering attack message. It can be seen that participants have a common understanding that this kind of message is dangerous. Therefore, informing participants about the potential risk of social engineering attacks confirms their suspicious about the message.

5. DISCUSSION

Many studies have concluded that training users or educating them or providing security tools will increase users' defence against social engineering attacks [25]. Therefore, numerous social engineering attacks have been prevented with these measures. However, social engineering attacks have increased in numbers as well as progressed in applying new techniques to lure victims [1]. It can be said that cybersecurity tools can prevent social engineering attacks from reaching users but cannot prevent all attacks. Additionally, cybersecurity education can inform users about ways to detect attacks, but not all types of attacks. For example, non-familiar attacks may still not be detected by users. Therefore, there is a need to encourage users to pay more attention in detecting attacks.

Previous studies show that users can detect familiar attacks. However, non-familiar attacks will not be detected easily. Most attacks mimic legitimate messages, which is the main reason users are lured into believing that attack messages are legitimate — for example asking users to click on links in emails, to update their account or open attached files. These requests will not be suspected by users as they are common actions in legitimate emails; but coming from a malicious entity the consequences will not be beneficial for users or organisations.

The theory of ELM suggest that users have two types of decision making, peripheral and central. The study by Vishwanath et al. [23] suggests that users become victims because of the use of the peripheral decision making scheme in

deciding whether to respond to a message or not. therefore, the trigger to make users switch their decision mood is by making them highly alerted about responding to messages. Then, the impact of recognising or identifying deceptive cues will trigger users to be suspicious about suspected messages. The shift in users' detection behaviour will determine the users' response.

Our study supports the idea that intention plays an important role in preventing users from falling victims to social engineering attacks. The role of intention will make users always be alert for fake messages. Users who think that fake messages will not come to them or will not harm them are more vulnerable to social engineering attacks. Therefore, organisations should play a role in making users more vigilant in their decision-making about received messages, as some of them might be social engineering messages.

The results obtained from our study support the argument that intention plays an important role in increasing users' defences against social engineering attacks. Our findings show that just informing users that social engineering messages can reach them shifts their mood from normal behaviour to a detection mood. Only just warning users triggers their attention to make robust decisions about messages, even when they encounter new social engineering attacks that they are not familiar with. Our participants were more cautious in making decisions about these messages. Therefore, users' decision-making shift from assuming that undecided messages are not harmful to interact with, to avoid interacting with undecided messages. In addition, this shift did not cause the avoidance of legitimate messages. Our results show that participants were able to identify legitimate messages and respond to them. The main danger lies in interacting with harmful messages.

Our study tackles the conflict in the reported results, namely that some studies find that some cybersecurity education programs and users' traits affect users' ability to detect social engineering messages, while other studies did not find any effect or sometime find the opposite results. Our results indicate that the main shift between detecting and not detecting social engineering messages is the users' mental state. If users are careless about the legitimacy of messages, they will rely on weak decision-making tools to judge messages. On the other hand, if users are warned or triggered about social engineering messages, they will use their already obtained knowledge or sometimes advance their knowledge and make a robust decision about received messages. Increasing users' attention to security will improve their defences against cybersecurity attacks. Therefore, our advice to organisations is to always make users alert about the possibility of social engineering attack messages.

6. CONCLUSION

The aim of our study was to find new ways to make users less susceptible to social engineering attacks. In the literature, it has been found that detection is affected by several factors. Less attention was given to what trigger detection. Our study found that users' intention to evaluate the authenticity of received message has a significant impact on making them less susceptible to social engineering attacks.

Warning users has improved users' detection accuracy from 74% to 77% and reduces wrong decisions from 26% to 23%. Additionally, users' hesitation was reduced dramatically. Training and awareness can improve users' detection for known attacks. However, zero-day attacks will still a threat to users. Therefore, making users always careful about responding or sometimes opening fake messages will increase their cyber-immune system against social engineering attacks. Organisations cannot train all their employees on all cybersecurity techniques available. Further research is needed to validate these findings in our study to the general users as our findings are not generalisable to the public users.

REFERENCES

[1] APWG. *Phishing Attack Trends Report – 3Q 2022* <https://apwg.org/>. Phishing Attack Trends Report 12 December 2022.

[2] F.B.I., *2021 IC3 Annual Report*, in *Internet Crime Report*, Abbate, P., Editor. 2022, Federal Bureau of Investigation: Federal Bureau of Investigation website. pp. 1-33.

[3] Krombholz, K., Hobel, H., Huber, M. and Weippl, E., Advanced social engineering attacks. *Journal of Information Security and Applications*, vol. 22, 2015, pp. 113-122.

[4] Syafitri, W., Shukur, Z., Mokhtar, U.A., Sulaiman, R. and Ibrahim, M.A., Social engineering attacks prevention: A systematic literature review. *IEEE Access*, vol. 10, 2022, pp. 39325-39343. DOI: https://doi.org/10.1109/ACCESS.2022.3162594.

[5] Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., et al. Getting users to pay attention to anti-phishing education: evaluation of retention and transfer. in *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*. 2007. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery.

[6] Sun, J.C.-Y., Kuo, C.-Y., Hou, H.-T. and Lin, Y.-Y., Exploring learners' sequential behavioral patterns, flow experience, and learning performance in an anti-phishing educational game. *Journal of Educational Technology & Society*, vol. 20, no. 1, 2017, pp. 45-60.

[7] Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., et al. School of phish: a real-world evaluation of anti-phishing training. in *Proceedings of the 5th Symposium on Usable Privacy and Security*. 2009. Mountain View, California, USA: ACM.

[8] Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., Jerram, C. Phishing for the truth: A scenario-based experiment of users' behavioural response to emails. in *Security and Privacy Protection in Information Processing Systems*. 2013. Berlin, Heidelberg: Springer Berlin Heidelberg.

[9] Downs, J.S., Holbrook, M.B. and Cranor, L.F. Decision strategies and susceptibility to phishing. in *Proceedings of the second symposium on Usable privacy and security*. 2006. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery.

[10] Alseadoon, I., Chan, T., Foo, E. and Gonzales Nieto, J. Who is more susceptible to phishing emails?: A Saudi Arabian study. in *ACIS 2012: Location, location, location:*

Proceedings of the 23rd Australasian Conference on Information Systems 2012. 2012. Geelong, VIC, Australia: ACIS.

[11] Alseadoon, I., Othman, M. and Chan, T., *What Is the Influence of Users' Characteristics on Their Ability to Detect Phishing Emails?*, in *Advanced Computer and Communication Engineering Technology*, Hamzah, A.S., et al., Editors. 2015, Springer: Switzerland.

[12] Baki, S. and Verma, R., Sixteen years of phishing user studies: What have we learned? *IEEE Transactions on Dependable Secure Computing*, vol. 14, no. 8, 2022, pp. 1-13.

[13] Burns, A., Johnson, M.E. and Caputo, D.D., Spear phishing in a barrel: Insights from a targeted phishing campaign. *Journal of Organizational Computing and Electronic Commerce*, vol. 29, no. 1, 2019, pp. 24-39.

[14] Goel, D. and Jain, A.K., Mobile phishing attacks and defence mechanisms: State of art and open research challenges. *Computers & Security*, vol. 73, 2018, pp. 519-544.

[15] Junger, M., Montoya, L. and Overink, F.-J., Priming and warnings are not effective to prevent social engineering attacks. *Computers in human behavior*, vol. 66, 2017, pp. 75-87.

[16] Volkamer, M., Renaud, K., Reinheimer, B., Rack, P., Ghiglieri, M., et al. Developing and evaluating a five minute phishing awareness video. in *International Conference on Trust, Privacy and Security in Digital Business*. 2018. Cham: Springer International Publishing.

[17] Valipe, V., *Impact of phishing on businesses: User awareness and response triggers to phishing emails*, in *Department of Information Systems*. 2018, University of Nebraska at Omaha: 789 East Eisenhower Parkway. pp. 80.

[18] Lin, T., Capecci, D.E., Ellis, D.M., Rocha, H.A., Dommaraju, S., et al., Susceptibility to spear-phishing emails: Effects of internet user demographics and email content. *ACM Transactions on Computer-Human Interaction*, vol. 26, no. 5, 2019, pp. 1-28.

[19] Parsons, K., Butavicius, M., Delfabbro, P. and Lillie, M., Predicting susceptibility to social influence in phishing emails. *International Journal of Human-Computer Studies*, vol. 128, 2019, pp. 17-26. DOI: https://doi.org/10.1016/j.ijhcs.2019.02.007.

[20] Butavicius, M.A., Parsons, K., Pattinson, M.R., McCormac, A., Calic, D., et al. Understanding susceptibility to phishing emails: Assessing the impact of individual differences and culture. in *International Symposium on Human Aspects of Information Security and Assurance*. 2017. Adelaide, Australia: University of Plymouth.

[21] Petty, R.E. and Cacioppo, J.T., *The elaboration likelihood model of persuasion*, in *Communication and persuasion*, Berkowitz, L., Editor. 1986, Academic Press.

[22] Dhamija, R., Tygar, J.D. and Hearst, M. Why phishing works. in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006. Montreal, Quebec, Canada: Association for Computing Machinery.

[23] Vishwanath, A., Herath, T., Chen, R., Wang, J. and Rao, H.R.J., Why do people get phished? Testing individual differences in phishing vulnerability within an integrated,

information processing model. *Decision Support Systems*, vol. 51, no. 3, 2011, pp. 576-586.

[24] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., et al. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. in *Proceedings of the 3rd symposium on Usable privacy and security*. 2007. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery.

[25] Alotaibi, F., Furnell, S., Stenge, I., Papadaki, M., Design and evaluation of mobile games for enhancing cyber security awareness. *Journal of Internet Technology and Secured Transactions*, vol. 6, no. 2, 2018, pp. 569-578.

[26] Likert, R., A technique for the measurement of attitudes. *Archives of psychology*, vol. 22, no 140, 1932, pp. 55.

[27] Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L.F. and Downs, J. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2010. ACM.

[28] Fink, A., *How to conduct surveys: A step-by-step guide*. 2015, UCLA, Los Angeles, USA: Sage Publications.

[29] Hair, J.F., Black, W.C. and Babin, B.J., *Multivariate Data Analysis: A Global Perspective*. Global Edition, ed. edition, t. 2010, Upper Saddle River, N.J., USA: Pearson Education. 800.

Information about the author:

Ibrahim Alseadoon – received his master's degree from University of Wollongong (UOW), Australia in 2008 and PhD from Queensland University of Technology (QUT), Australia in 2014. He is currently an associate professor in Information and Computer Science (ICS) department at University of Ha'il (UOH), Ha'il, KSA. He is an author of more than 6 articles in the field of Computer Security and Users' Behaviour.

Manuscript received on 13 June 2023