Making Warning Messages Personal: A Big 5 Personality Trait Persuasion Approach

Joseph Aneke, Carmelo Ardito, Giuseppe Desolda

Making Warning Messages Personal: A Big 5 Personality Trait Persuasion Approach

Joseph Aneke[1] [0000-0001-9544-8972], Carmelo Ardito[2] [0000-0001-8993-9855], Giuseppe Desolda[1] [0000-0001-9894-2116]

[1]Dipartimento di Informatica, Università degli Studi di Bari, Italy
[2] Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, Italy.

joseph.aneke@uniba.it, carmelo.ardito@poliba.it, giuseppe.desolda@uniba.it

Abstract. Several mitigation strategies in form of warning messages against phishing attacks have continued to fail largely due to user negligence. Thus, it is important for researchers to focus not only on the accuracy of the provided recommendations but also on other factors that influence the acceptance of recommendations and the extent to which these recommendations are convincing or persuasive. In this paper, we present our ongoing approach that leverages on the *Big 5 Personality trait* model and users digital traces harvested from their social networks, thereafter, transformed into a personalized warning message. We argue that stimulating users through personal recommendations evokes an understanding of the implications of their actions or inaction.

Keywords: Personalized messages, Big 5 personality trait, Social Network Sites (SNSs)

1 Introduction

Recipients of a warning message in the event of a phishing attack are more likely to heed a warning if they believe they are the intended recipient [1] and perceive the content as applicable to their unique personality. In light of this, the issue of personalization becomes key in presenting warnings. In this work we are guided by the following research questions:

RQ1: How can we deploy human-related theories to enhance users' acceptance of warning recommendations?

RQ2: How can unique digital traces combined with persuasiveness as intrinsic characteristics be integrated into the design of a warning mechanism system to increase users' acceptance of the warning recommendations?

The impact of the Big 5 personality traits with reference to digital traces on social networks has been a subject of interest to researchers with a possible positive correlation been envisaged in recommendations [2]. For instance, authors in [1] through Facebook likes by users demonstrated that Computer-based personality judgments were more accurate and reliable than those made by humans or close friends.

In developing our personalized warning messages, we developed a framework consisting of two key approaches. First, there was the refinement of the warning message text to appeal to recommended individual personality trait preferences. Secondly, we exploited harvesting unique digital traces from already existing users' relationships in SNS to reinforce the acceptance of the warning advice. Thus, a variant of the warning text tailored to a user's personality trait, combined with inputs (e.g., profile picture) of a Phishing attack expert, makes the message personalized and easier to accept.

The rest of the paper is structured as follows: in Section 2 we present a background to related work, Section 3 discusses our conceptual framework. Conclusions and future work are provided in Section 4.

2 Background of related work

Personalization or customization of messages is a term used to describe different styles of tailoring strategies [3]. Its applications could be found in websites, emails, health campaigns, etc. [4]. The research community has tried to understand, among others, its structure and user interconnection, as well as interactions with Big 5 personality traits, in keeping persons glued for so long to messages [5].

The Big 5 Personality Traits model was developed by the psychologists Paul Costa and Robert McCray and it is one of the most common models used in modern research into information security behavioral analysis [6]. It consists of five factors that represent personality traits: conscientiousness, agreeableness, extroversion, neuroticism, openness to experiences [7].

With respect to information security, studies have shown that there is a significant relationship between conscientiousness personality and agreeableness personality as it pertains to security policy [8].

3 Design of Warning Text

In order to address RQ1, we adopted the Big 5 personality trait spectrum. In this spectrum, each of the individual personalities may fall anywhere either from Low (Negative), Neutral, or High (positive) [9], as depicted in **Fig. 1**. Hence, our designed warning texts were done to represent several variants that returned all possible emotional values within the spectrum range.

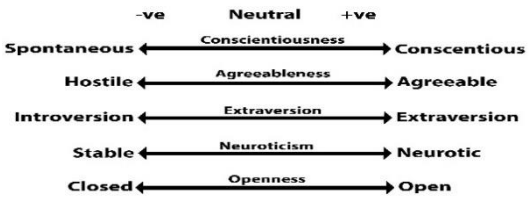

Fig. 1. Big 5 Personality Trait Spectrum

The generation of the explanations proposed in the text of our warning messages is based on a generic pattern purposely defined in [10-12] to instantiate each warning message. For example, for the Time life security indicator, a phishing warning message text reads as:

"This website was created recently. This is typical of fraudulent websites. Do not disclose private information on it.".

According to the warning guidelines indicated in [13], we produced five variants of the above text message (see **Table 1**), which had the same objective but were expressed in different ways. We then subjected these messages to sentiment analysis and readability measures evaluations. The messages returned all possible spectrum values (-ve, Neutral and +ve), as reported in the sentiment analysis column.

Table 1. Warning message variant across spectrum range

Message Variant	Sentiment Analysis	Reach (%)	Word count	Smog Index
This website was created recently. This is typical of fraudulent websites. Do not disclose private information on it.	-0.93	100	18	6.8
This website is young. It is typical of fake websites. Do not disclose your private information on it.	-0.85	100	17	5
Young (New) websites are famous for criminal activities. There is a potential risk if you proceed.	+0.39	100	17	7.2
Young websites are famous for criminal activities. It may be risky if you proceed.	+0.46	100	14	6
Young websites are famous for criminal activities. There may be a potential risk if you proceed.	0.0	100	15	7.2

3.1 Integration of Digital Traces in Warning Messages

In order to address RQ2, we adopted Cialdini's persuasion principles [14]. It states that when people are faced with a lifeline decision, they defer to experts or similar others, they are most likely inclined to accept requests made by somebody they like in a way, and usually have the tendency to commit to their previous or reported opinion or behavior.

The conceptual framework, presented in **Fig. 2**, comprises of two stages: Identification and pre-processing and customization. At the *Identification and pre-processing phase*, three key activities take place: 1) the defiled security indicator is identified from the pool; 2) the personality traits of the user and the spectrum rage values for the warning text are marked; 3) the list of social networks that the user has subscribed to is also identified and ranked. At the *Customisation phase*, the best fit warning text and the most persuasive image (e.g., profile picture) selected from users frequently used social network is selected. The most persuasive image is determined through a possible relationship to a friend or friend of friends. Once a relationship graph has been derived, the

shortest path is selected, and the resultant friends' picture is selected and uploaded on the tray for the warning message.

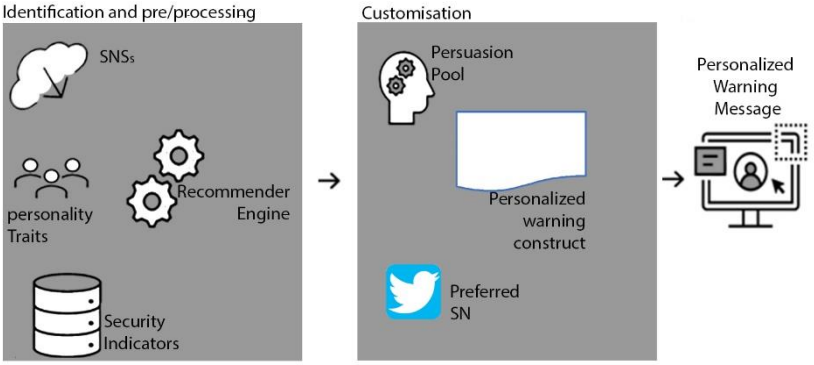

Fig. 2. Conceptualization of the proposed framework

Fig. 3 shows how the mechanism has been implemented as a splash-screen displayed as soon as the user tries to access a deceptive website. On the left side, one of the variants of the warning message is shown, while the right panel aims at persuading the user to not proceed since the users' friends Joseph, Lucia, and Andrea strongly recommend that he should not proceed to the fake URL (www.bank.0f.america.com).

Fig. 3. Implementation of the personalized warning message mechanism

4 Conclusion

In this paper, we presented a mechanism that aims at drawing users' attention to phishing warning messages and help them to take the right decision about proceedings towards a website or not. Our contribution is novel in that we proposed an approach that tailors the warning text to suit the Big 5 personality trait. Also, the personalized warnings included referrals and pictures harvested from a user's SNS and recommendations from experts in security domains implicit or explicit to them. As a next step, we would evaluate our design with real users. Also, it could be extremely interesting to study the possible use of the multiple SNSs which users do not subscribe to.

References

1. E. J. Baker, "Hurricane evacuation behavior," *International journal of mass emergencies and disasters*, vol. 9, pp. 287-310, 1991.
2. S. Winter, E. Maslowska, and A. L. Vos, "The effects of trait-based personalization in social media advertising," *Computers in Human Behavior*, vol. 114, p. 106525, 2021.
3. D. Houghton, A. Pressey, and D. Istanbulluoglu, "Who needs social networking? An empirical enquiry into the capability of Facebook to meet human needs and satisfaction with life," *Computers in Human Behavior*, vol. 104, p. 106153, 2020.
4. A. B. Kocaballi, S. Berkovsky, J. C. Quiroz, and L. Laranjo, "The Personalization of Conversational Agents in Health Care: Systematic Review," *J Med Internet Res*, vol. 21, p. e15360, 2019.
5. B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in *Proceedings of the 2nd ACM workshop on Online social networks*, 2009, pp. 37-42.
6. J.-L. Wang, L. A. Jackson, D.-J. Zhang, and Z.-Q. Su, "The relationships among the Big Five Personality factors, self-esteem, narcissism, and sensation-seeking to Chinese University students' uses of social networking sites (SNSs)," *Computers in Human Behavior*, vol. 28, pp. 2313-2319, 2012.
7. O. P. John and S. Srivastava, *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives* vol. 2: University of California Berkeley, 1999.
8. U. Franke and J. Brynielsson, "Cyber situational awareness–a systematic review of the literature," *Computers & security*, vol. 46, pp. 18-31, 2014.
9. Lim. *https://www.simplypsychology.org/ Retrived 15/04/2021*.
10. J. Aneke, C. Ardito, and G. Desolda, "Help the user recognize a phishing scam: The design of explanation messages in warning interfaces for phishing attacks," presented at the 23rd International Conference on Human-Computer Interaction., Washington DC, USA, 2021.
11. J. Aneke, C. Ardito, and G. Desolda, "Designing an Intelligent User Interface for Preventing Phishing Attacks," in *IFIP Conference on Human-Computer Interaction*, 2019, pp. 97-106.
12. A. C. Aneke J., Desolda G., " Unpacking Warning Messages: Towards Mitigating Phishing Attacks. ," presented at the In proc. of Italian Conference on ICT for Smart Cities And Communities (iCities '19). 18-20 September, 2019. Pisa, Italy, 2019.
13. C. Bravo-Lillo, L. F. Cranor, J. Downs, S. Komanduri, and M. Sleeper,"Improving computer security dialogs," in *IFIP Conference on Human-Computer Interaction*, 2011, pp. 18-35.
14. R. B. Cialdini, "The science of persuasion," *Scientific American*, vol. 284, pp. 76-81, 2001.