Help the user recognize a phishing scam: design of explanation messages in warning interfaces for phishing attacks

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] Department of Electrical and Information Engineering, Politecnico di Bari, Italy.

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

Abstract. Remote work due to the COVID-19 pandemic is expected to be the new normal, suggesting a situation where people use their personal computers at home for several activities like reading emails, surfing the web, chatting with friends. While doing this, users are not focused on securing their systems and they often do not have the skills and knowledge to defend against cybercrime. In this paper, we present the design and the evaluation of a novel interface that warns users against phishing attacks. This interface looks like the ones shown by browsers like Chrome and Firefox when opening a suspicious phishing website, but it includes information that explains the reasons why the website might be a scam. Such explanations are based on website features commonly used by AI-based systems to classify a website as phishing or not and aim to help users detecting phishing websites. To ensure a high understandability and effectiveness of the explanations, the C-HIP model was adopted to design such messages. In addition, the resulting messages have been iteratively refined performing a static analysis of their comprehension, sentiment, and readability.

Keywords: Polymorphic warning messages, Usable Security, Cybersecurity

1 Introduction

Remote work due to the COVID-19 pandemic is expected to be the new normal [1], suggesting a situation where people use even more their computers, smartphones, and tablets at home for several activities. These devices often lack professional antivirus protection programs, or firewalls [2]. In some cases, software in use may have simply reached the end of its life cycle, implying no relevant security updates as seen in windows 7 lately [3], obviously exposing users to attacks. Although the new normal would facilitate users' protection since they can comfortably work in their homes, this situation resulted in the upsurge of cybercrime and further exploitation of user vulnerabilities

increased by 600% in March 2020 [4, 5]. These vulnerabilities are often also due to certain human factors and poor designs of security warnings messages [5-7].

In most cases, the last barrier between victims and attackers are the warning messages. It is typical of phishing attacks: when browsers detect fraudulent phishing websites they show warning messages that ask users to open or not the target website. Therefore, users need the right information presented in a manner that they can understand easily, and at the time they need to make the decision [8]. There is a need for warnings to appear with specific features in a simple language, which are polymorphic [7] and that is not generalized in terms of look and feel [9, 10]. Understanding of the warning becomes an important factor since human decisions are involved. Current warning designs focus on describing the potentially dangerous outcome if the warning is not heeded. Existing literature on improving comprehension recommends using simple plain language avoiding technical jargon [5, 6], and to describe the specific risks clearly and being as brief [5].

This paper aims to improve the effectiveness of warning messages defending users from phishing attacks. In the last years, we worked on the design of a warning message that not only informs users of an ongoing attack but that also explains the reasons why the target website could be fraudulent [9][10]. The design of our solution is based on lessons drawn from warning literature on best practices [11-13]. This paper advances our previous research [10] focusing on the understandability of the explanations provided by the warning messages. For each type of explanation provided by the message, three variants have been designed following the C-HIP model. Then, an evaluation of resulting variants has been conducted performing a static analysis of the messages' comprehension, sentiment, and readability.

In the next section, an overview of related work of phishing attacks is provided. In Section 3 we present the design of warning messages and evaluation metrics. Finally, Section 4, discusses our findings with conclusions and future work.

2 Background and related work

2.1 Human factors in cybersecurity

There has been an upsurge in scams and malware attacks recently with phishing attacks, which increased by 600% in March 2020 [4]. Google also blocked 18 million malware and phishing emails related to the COVID19 virus daily in April 2020 [14]. According to [15], cybercrime will cost the world $6 trillion annually by 2021, up from $3 trillion in 2015, with 95% cases due to human errors attributed, in some cases, non-compliance to warning messages that alert users on cyberattacks. This has made research in warning messages grow considerably over the past decades [16-18]. During this time researchers have continued to investigate a wide variety of variables, e.g., user behavior, environmental stimuli, text evaluation. Again, the concepts of uncertainty and risks are difficult for people to evaluate when faced with issues that require critical decisions [19]. For designers of security systems, it is essential to understand how users would evaluate and take decisions regarding security [19, 20]. Authors in [21] showed that the

central problem of human interaction with IT security systems is that users should be able to make informed decisions without further help. They illustrated this by designing and implementing two applications that make visible the visualization of network events and the integration of action and configuration of available security mechanisms.

Felt et al in [22] evaluated whether Android users pay attention to, understand, and act on permission information during installation. Their study participants (Internet survey and laboratory) displayed low attention and comprehension rates. About 17% paid attention to permissions during installation, and only 3% of Internet survey respondents could correctly answer all three permission comprehension questions. These results suggest that current Android permission warnings do not help most users make correct security decisions.

In the study reported in [23], the authors observed a disparity between actual changes made by Windows 7 updates and the changes the participants thought were being made. A multi-method approach that collected (interview, survey, and computer log data) from 37 Windows 7 users and compared what the users thought was happening on their computers (interview and survey data), what users want to happen on their computer (interview and survey data), and what was going on (log data) was used. Results showed that 75% of participants had a misunderstanding about what was happening on their computer and that over 50% of the participants could not execute their intentions for computer management.

Bravo-Lillo et al [24] examined the behavior novice users exhibit when confronted with situations in which they should make security decisions. They demonstrate that these categories of users are not aware of the sensitivity of their data and mostly started to worry after deciding to allow access.

Fagan et al [25] investigated user motivations on why some users follow advice on security aspects and others do not. They conducted a survey study with 290 participants using a rational decision model as well as current thoughts on human motivation where they asked participants about their motivations regarding (not) updating, using a password manager, using two-factor authentication, and changing passwords frequently. The authors determined that following security advice was mainly a trade-off decision between convenience and security, where users actively considered features such as set-up time and weighed that against the potential security benefits. They concluded that the value of convenience may be used to help motivate the use of security tools and techniques.

Almuhimed et al in [26] investigated factors that may contribute to why people ignore warnings. Through an online survey-based experiment they did to gain more insight into the effects of reputation on warning adherence. Participants said that they trusted high-reputation websites more than the warnings; however, their responses suggest that a notable minority of people could be swayed by providing more information. Gainsbury et al in [27] conducted a field experiment to examine the impact of warning message content on gambling behavior, found that warning messages focused on self-appraisal (positively) framed messages were more frequently recalled than informative (negatively) framed messages, but that negatively framed messages were more influential.

In [28] the authors examined the impact of negative message framing on security technology adoption. Based on previous studies, they hypothesized that negatively framed messages would have a greater effect on the adoption of security technologies that detect system abuse than on technologies for prevention and that internet security managers should become more sensitive to how new security technologies are introduced and to the factors that help shape adoption intentions. Authors in [29] suggest that people may pay attention to warnings but are most likely to ignore those that do not map well onto a clear and understandable course of action. Experimental studies in [30, 31] indicate that a large percentage of users do not read computer warnings, rarely understand them, and are most likely not to heed them, even when there are obvious risks. Qualitative insight into warning assessment by users of different skill levels, age, and exposure is presented in [24] they conclude that all aspects of warning design need to be considered holistically to improve warnings. Stating that the process of reading a warning is central to warning message reception and understanding.

2.2 Design warning messages: the C-HIP Model

Warning messages play a fundamental role in defending users against cyberattacks since they often are the last barrier between the attacker and the victim. However, as we discussed above, warning messages shown by the browsers in case of phishing attacks often fail in helping users understand if the target website is fraudulent or not. Besides the guidelines proposed to design warning messages for phishing attacks [16], the design of these messages can also benefit from the use of models for the design of generic warnings. One of the most adopted models in the literature and that we used to design our warning messages is the Communication-Human Information Processing (C-HIP) model, which defines the critical route and sets the foundation for structuring warning messages (see **Fig. 1**) [17].

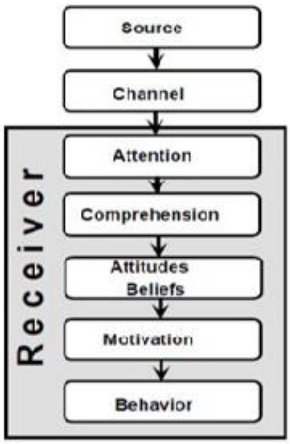

Fig. 1. C-HIP Model

The model summarizes the most important activity and entity involved in the communication of a warning. The model starts with a source delivering a warning through a channel to a receiver, who then takes it along with other stimuli (environmental or internal) that subject the message to a lot of distractions or distortions. It then identifies a set of steps between the delivery of a warning and the user's final behavior or response which is usually based on the resultant effect of the various processes such warnings had undergone. An essential part of a warning message is defined by [18] as there must be a signal word that should be noticeable (salient) e.g., Danger, Warning, Caution and Notice. This signal word helps increase the effectiveness of the warning. In 2011, Bravo-Lillo et al. [32] compiled a set of design guidelines and presented rules for descriptive text, which includes:

- Briefly describe the risk and consequences of not complying with advice;
- Illustrate clearly how to avoid the risk;
- Be transparent and avoid technical jargon were possible;
- Be brief as possible.

2.3 Evaluating warning messages comprehension

After a warning message captures a user's attention, the next step is message comprehension. Most warning designers assume that users understand the hazard been described and subsequently adhere to prescribed advice. This has proven not to be so as reported in [25, 33, 34]. Users exhibit different levels of comprehension and interpretations of given texts, which subsequently influences their actions or inaction. Due to technical complexity, novice users may not fully understand what URL mimicking warning means (a situation where an attacker mimics a genuine URL, thus re-directing unsuspecting users to a similar webpage to steal their sensitive credentials). Therefore, warning message text should be targeted at least skilled users stripping off complex technical terms as much as possible. To this aim, the design of warning messages for phishing attacks might benefit from the use of static evaluations of the readability and sentiment. In the following sections, we briefly report on some of the most adopted metrics for text readability and sentiment, which were also used to design the warning message proposed in this paper.

2.3.1 Sentiment Analysis

Sentiment Analysis involves determining the evaluative nature of a piece of text. These texts convey emotions which are key components in communication to effectively communicate messages and to understand reactions to messages [35, 36]. These messages could be classified as positive, negative, or neutral [37]. This analysis is common in customer reviews, newspaper headlines [38], novels and emails [36, 38-42], blogs and tweets [41, 42] and negative messages were found to appeal to certain behavioral anticipated responses. Surveys by [42, 43] give a summary were using Natural Language Processing (NLP) techniques (e.g., IBM Watson [44]), automated agents can gain the ability to process and analyze text at different levels of abstraction, exploiting the speed and computational power of modern systems. Within the computer security software

vendor community, the use of negatively framed messages to influence the adoption of their products is not novel. A growing trend among purveyors of information assurance and computer security technology is to employ negatively framed messages to provoke a favorable behavioral response among existing and potential clientele [45, 46].

2.3.2 Readability measures

Given a piece of text, readability metrics measure the degree to which a person can read, easily understand, and find interesting such text [47-49]. Reading sometimes may appear like a complex phenomenon dependent on several factors e.g., cognitive, behavioral, and social [50-52]. To measure readability, several formulas have been established in literature with each specific for different metrics. The most popular is the Flesch-Kincaid Grade Level, Flesch Reading Ease Formula, and SMOG formula [53-55]. A combination of word length, sentence length, and conversancy with word has been used to predict readability, with the background knowledge that longer words/sentences, which are usually used with a complex syntax, indicate greater reading difficulty and recall [56, 57]. Also, since shorter words tend to be more common than longer ones in English, longer words are considered less likely to be familiar to the reader [58]. While these assumptions do not account for individual readers' vocabulary and reading experience, simple metrics such as sentence and word length can provide a useful initial step in assessing readability. In [59] the authors investigate the application of readability measures to assess the difficulty of the descriptive text in warning messages. They agree that adapting such a measure to the needs of warning message design allows objective feedback on the quality of a warning's descriptive text. They concluded that an automated process will be able needful to assist software developers and designers in creating more readable and hence more understandable security warning messages. In the following sections, we provide a brief description of these widely adopted formulas and their evaluation results of our warning messages.

2.3.3 Flesch – Kincaid Grade Level and Flesch Reading Scores.

The Flesch–Kincaid readability tests are readability tests designed to indicate how difficult a provided text usually in English can be understood [60]. These involve primarily two groups of tests, the Flesch Reading Ease Test, and the Flesch–Kincaid Grade Level Test [54, 60]. Although they utilize the same metrics (word length and sentence length), they have different weighting ratios that are used to approximate the reading grade level and scores of a text. Flesch Reading Ease score is graded between 1 and 100, while the Flesch Kincaid Grade Level reflects the US education system needed to understand a text [49, 54]. They are both calculated with the same units, but they return two different readability scores. The authors of [53] explained that the higher the Reading score, the easier for a particular text to be read by the majority of people.

Reading Ease Score	Description	Predicted Reading Grade	Estimated % of US Reading Adults
0 – 30	Very difficult	College	4.5%
30 – 50	Difficult	College	33%
50 – 60	Fairly difficult	10^{th} – 12^{th}	54 %
***60 – 70**	**Standard**	**8^{th} – 9^{th}**	**83%**
70 – 80	Fairly Easy	7^{th}	88%
80 – 90	Easy	6^{th}	91%
90 – 100	Very easy	5^{th}	93%

Table 1. Description and predicted reading grade for Flesch Reading Ease Scores [53]

2.3.4 SMOG Index

SMOG is an acronym for Simple Measure of Gobbledygook, it is widely recognized to be a nod to Robert Gunning's Fog Index. Credited to G Harry McLaughlin [55], the SMOG Index estimates the years of education a person needs to be able to comprehend a passage, it was developed as an improvement of other readability measures [61]. It involves estimation of two statistics: First, the number of sentences of the selected article and the number of words with three or more syllables [62].

3 Toward explanations inside warning interfaces for phishing attacks

3.1 A polymorphic warning message to prevent phishing attacks

The goal of our research is to improve the effectiveness of warning messages against phishing attacks. To this aim, we already proposed an interface similar to the ones shown by browsers like Chrome and Firefox when opening a suspicious phishing website, but that also includes information clarifying the reasons why the website might be a scam [9, 10]. Such explanations are based on website features commonly used by AI-based systems to classify a website as a scam or not (e.g., by Google Safe Browsing [23]). We only considered those features that can be explained to and understood by non-technical users, i.e.:

- server location;
- website time life;
- presence in the Wayback Machine;
- rank in a search engine;
- fake HTTPS certificates;
- mimicked URLs;
- domain name.

To this aim, when a suspect website is detected, our system first computes the website features (e.g., HTTPS certificate = self-signed, time life = 2 days) and then it ranks all the feature values according to a metric we elaborated and that is based on the feature

entropy. In the end, the warning message selects and shows the most informative features. Details on the metric and on the algorithm for feature ranking are submitted for publication.

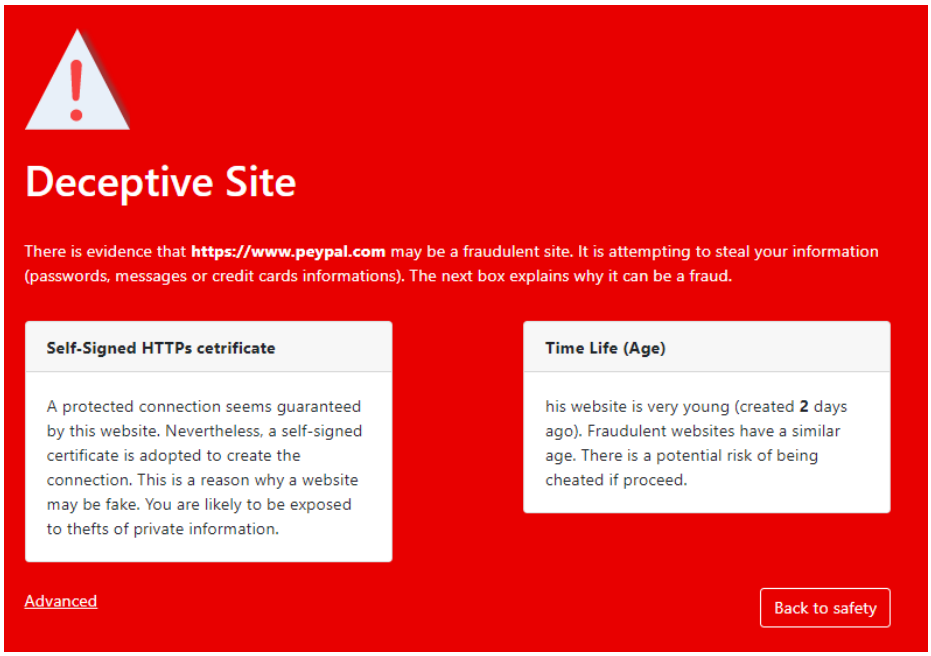

Fig. 2. Polymorphic warning message against phishing attacks

According to the example in **Fig. 2**, the proposed warning interface explains that the bank website going to be visited by the user is detected as phishing for two reasons: it uses a fake HTTPS certificate and it has been created just 2 days ago. Such information, properly structured and adequately shown to the users, help them to make informed decisions and avoid opening scam websites.

It is worth noticing the polymorphic behavior of the interface: the three panels show different information according to the suspect website, thus different reasons would be reported with different phishing websites. Thanks to this warning message, its polymorphic behavior and the explanations it provides, we address three important goals, i.e.:

1. *Prevent user habituation*: a polymorphic message decreases the clickthrough effect caused by the user habituation [22];
2. *Provide an explanation about the attack*: useful information about the causes of the phishing attacks support the users in deciding if the website is (or not) a phishing attack [23];

3. *Educate the users on cyberattacks and related risks*: a long-term training of the users on phishing attacks is performed since they understand the reasons for this attack [16, 24].

In the next section, we briefly describe how the C-HIP model has been used to design the feature explanations.

3.2 Warning text design

Based on best practices from warning literature [59], we designed our warning message texts aimed at warning users against attacks. In this paper, we report the design of two out of the seven indicators: Website time life (Age) and Hypertext transfer protocols (HTTPS). The design and the evaluation of the other 5 messages is an ongoing activity. The generation of the explanations is based on a generic pattern we purposely defined to instantiate each message, i.e.:

Feature value + illustrated example of feature [optional] + Hazard Identification + Effects of a successful attack

For example, a warning message text mimicking a URL indicator reads as follows:

"A protected connection seems guaranteed by this website. Nevertheless, a self-signed certificate is adopted to create the connection. This is a reason why a website may be fake. You are likely to be exposed to thefts of private information".

Feature	A protected connection seems guaranteed by this website
Illustrated example of feature [optional]	none
Hazard Identification	Nevertheless, a self-signed certificate is adopted to create the connection. This is a reason why a website may be fake
Effects of a successful attack	You are likely to be exposed to thefts of private information

Table 2. Schema elucidation

As the next steps, we produced three variants of messages (see **Table 3**) in line with warning guidelines indicated in [32, 52, 63], for each indicator which had the same objective but in different ways.

Features (Indicators)	**Message Variants**
Website time Life (Age)	1. This website was created recently (n days ago). This is typical of fraudulent websites. It likely aims to steal your private information. 2. 2.This website is very young (created n days ago). Fraudulent websites have a similar age. There is a potential risk of being cheated if you proceed. 3. 3.The target website was created n days ago. Young websites are famous for criminal activities. There is a potential risk if you proceed

Self-signed HTTPS certificate	4. This website seems to have a protected connection. However, its connection uses a self-issued certificate. This indicates it may be a fraud. You will most likely be exposed to thefts of private information. 5. A protected connection seems guaranteed by this website. Nevertheless, a self-signed certificate is adopted to create the connection. This is a reason why a website may be fake. You are likely to be exposed to thefts of private information. 6. This website seems to offer a safe connection. This is not safe since it is a self-validated certificate. Attackers self-validate their websites to cheat and defraud users. Your private information is at risk.

Table 3. Warning message variants for URL and HTTPS

We then subjected these messages to sentiment analysis and readability measures evaluations as discussed above. We also used online text Inspector tools listed in [63, 64] to measure the lexical diversity of text in the warning messages. **Table 4**. summarizes the results of the static analysis of the texts.

Warning message Variant	Flesch Reading Ease Score	Flesch-Kin. Grade Level	SMOG Index	Sentiment Analysis	Word count	Sentence count
Age (V1)	57.1	7.1	6.8	-0.90	22	3
Age (V2)	67.9	5.9	6	-0.95	26	3
Age (V3)	63	6.4	6	-0.39	23	3
HTTPs (V1)	61.4	6.8	7.8	-0.83	34	4
HTTPs (V2)	62.3	7.1	8.3	-0.59	40	4
HTTPs (V3)	62.6	6.7	7.2	-0.71	35	4

Table 4. Warning message statistics using online tools

All the warning texts had negative values for the sentiment analysis with ranges -0.39 to - 0.95. The readability scores, which include the word count (total number of words in a sentence), returned an average of 30 counts ranging between 22 to 40. Reach (a measure of the proportion of your target audience that can read your content easily) calibrated against the literate general public, so a reach of 100% indicates your content is readable by about 85% of the public, all returned an average of 100%. All readability measures metrics are in line with the recommendation proposed in [12, 58, 65].

4 Conclusion and future work

We set out to design warning messages for two indicators (Website time life (Age) and HTTPS), three variants for each, which provided explanations to users, in particular to those users that do not have expertise in IT or security and why they should not oblige

to attacker's request on phishing websites. During our investigations, we found several aspects of warning message texts that could improve user comprehension and adherence. First, the architectural model – CHIP Model, which we adopted in building our schema (Feature value + Hazard identification + effects of a successful attack), supported our explanation design. Then, an iterative process subjected the negatively framed text messages to sentiment analysis and finally evaluated them for readability compliance. The three variants messages for each selected indicator showed overwhelming compliance to our set out objectives - comprehension and adherence as evident in results from the online tools used. Some researchers still have reservations about the use of online tools for the evaluation of messages claiming that textual statistics alone cannot address certain emotions and complexities required in a message. We were able to address this concern by tailoring our warning messages to return negative values during the sentiment analysis. As described in the previous sections and in line with our objectives, it is common among computer security technology, to employ negatively framed messages to provoke a favorable behavioral response among existing and potential clientele [45, 46].

As the next steps, we will design more indictors which include: Server location, Domain name, Mimicked URLs, Rank in a search engine, and presence in the Wayback machine and subsequently evaluate our polymorphic warning messages with those found in popular browsers (chrome and firefox) involving real users. We agree that while readability formulas may not measure the context, prior knowledge, interest level, difficulty of concept, or coherence of text of users as illustrated in [48, 65], it however has the potential to provide designers and developers with an automatic tool that can estimate how readable and understandable a warning will be for their target audience thus suffixing for those developers that cannot afford specialist help.

References

[1] E. Brynjolfsson, J. J. Horton, A. Ozimek, D. Rock, G. Sharma, and H.-Y. TuYe, "COVID-19 and remote work: an early look at US data," National Bureau of Economic Research 0898-2937, 2020.

[2] J. Wiggen, "The impact of COVID-19 on cyber crime and state-sponsored cyber activities," 2020.

[3] E. Bott, "How many people still run windows 7 https://www.zdnet.com/article/as-support-ends-windows-7-users-head-for-the-exits/ (Access 28/10/2020) ".

[4] S. Gallagher and A. Brandt. (2020). *"Facing down the myriad threats tied to covid19," 2020, https://news.sophos.com/enus/2020/04/14/covidmalware (Accessed 28/10/2020).*

[5] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 581-590.

[6] B. Friedman, D. Hurley, D. C. Howe, E. Felten, and H. Nissenbaum, "Users' conceptions of web security: a comparative study," in *CHI'02 extended abstracts on Human factors in computing systems*, 2002, pp. 746-747.

[7] C. Bravo-Lillo, L. Cranor, S. Komanduri, S. Schechter, and M. Sleeper, "Harder to ignore? Revisiting pop-up fatigue and approaches to prevent it," in *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*, 2014, pp. 105-111.

[8] C. Jackson, D. R. Simon, D. S. Tan, and A. Barth, "An Evaluation of Extended Validation and Picture-in-Picture Phishing Attacks," in *Financial Cryptography and Data Security*, Berlin, Heidelberg, 2007, pp. 281-293.

[9] G. Desolda, F. Di Nocera, L. Ferro, R. Lanzilotti, P. Maggi, and A. Marrella, "Alerting Users About Phishing Attacks," in *HCI for Cybersecurity, Privacy and Trust*. vol. LNCS 11594, A. Moallem, Ed., ed Cham: Springer International Publishing, 2019, pp. 134-148.

[10] 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.

[11] M. Alsharnouby, F. Alaca, and S. Chiasson, "Why phishing still works: User strategies for combating phishing attacks," *International Journal of Human-Computer Studies*, vol. 82, pp. 69-82, 2015.

[12] M. Harbach, S. Fahl, P. Yakovleva, and M. Smith, "Sorry, I don't get it: An analysis of warning message texts," in *International Conference on Financial Cryptography and Data Security*, 2013, pp. 94-111.

[13] S. J. Greenwald, K. G. Olthoff, V. Raskin, and W. Ruch, "The user non-acceptance paradigm: INFOSEC's dirty little secret," in *Proceedings of the 2004 workshop on New security paradigms*, 2004, pp. 35-43.

[14] N. Kumaran and S. Lugani, "Protecting businesses against cyber threats during COVID-19 and beyond," *Google Cloud*, vol. 16, 2020.

[15] C. M. Williams, R. Chaturvedi, and K. Chakravarthy, "Cybersecurity Risks in a Pandemic," *Journal of Medical Internet Research*, vol. 22, p. e23692, 2020.

[16] M. S. Wogalter, *Handbook of warnings*: CRC Press, 2006.

[17] K. Laughery, D. DeJoy, and M. Wogalter, "Warnings and Risk Communication," *Taylor and Francis*, 1999.

[18] M. S. Wogalter, V. C. Conzola, and T. L. Smith-Jackson, "Research-based guidelines for warning design and evaluation," *Applied Ergonomics*, vol. 33, pp. 219-230, 2002/05/01 2002.

[19] R. West, "The psychology of security," *Communications of the ACM*, vol. 51, pp. 34-40, 2008.

[20] P. Kumaraguru, Y. Rhee, S. Sheng, S. Hasan, A. Acquisti, L. F. Cranor, *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, pp. 70-81.

[21] R. De Paula, X. Ding, P. Dourish, K. Nies, B. Pillet, D. Redmiles, *et al.*, "Two experiences designing for effective security," in *Proceedings of the 2005 symposium on Usable privacy and security*, 2005, pp. 25-34.

[22] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner, "Android permissions: User attention, comprehension, and behavior," in *Proceedings of the eighth symposium on usable privacy and security*, 2012, pp. 1-14.

[23] R. Wash, E. Rader, K. Vaniea, and M. Rizor, "Out of the loop: How automated software updates cause unintended security consequences," in *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*, 2014, pp. 89-104.

[24] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri, "Bridging the gap in computer security warnings: A mental model approach," *IEEE Security & Privacy*, vol. 9, pp. 18-26, 2011.

[25] M. Fagan and M. M. H. Khan, "Why do they do what they do?: A study of what motivates users to (not) follow computer security advice," in *Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016)*, 2016, pp. 59-75.

[26] H. Almuhimedi, A. P. Felt, R. W. Reeder, and S. Consolvo, "Your reputation precedes you: History, reputation, and the chrome malware warning," in *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*, 2014, pp. 113-128.

[27] S. M. Gainsbury, A. Russell, S. Gainsbury, D. Aro, D. Ball, C. Tobar, *et al.*, "Optimal content for warning messages to enhance consumer decision making and reduce problem gambling," 2015.

[28] J. D. Shropshire, M. Warkentin, and A. C. Johnston, "Impact of negative message framing on security adoption," *Journal of Computer Information Systems*, vol. 51, pp. 41-51, 2010.

[29] K. Witte, "Putting the fear back into fear appeals: The extended parallel process model," *Communications Monographs*, vol. 59, pp. 329-349, 1992.

[30] S. Egelman, L. F. Cranor, and J. Hong, "You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings," 2008.

[31] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor, "Crying Wolf: An Empirical Study of SSL Warning Effectiveness," in *USENIX security symposium*, 2009, pp. 399-416.

[32] 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.

[33] S. Furnell, "Why users cannot use security," *Computers & Security*, vol. 24, pp. 274-279, 2005.

[34] A. Herzberg, "Why Johnny can't surf (safely)? Attacks and defenses for web users," *Computers & Security*, vol. 28, pp. 63-71, 2009.

[35] M. Thelwall, "Heart and soul: Sentiment strength detection in the social web with sentistrength (summary book chapter)," *Cyberemotions: Collective emotions in cyberspace. Berlin, Germany: Springer*, pp. 119-134.

[36] S. M. Mohammad, "Challenges in sentiment analysis," in *A practical guide to sentiment analysis*, ed: Springer, 2017, pp. 61-83.

[37] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment Analysis of Short Informal Texts," *Journal of Artificial Intelligence Research*, vol. 50, pp. 723-762, 2014.

[38] J. R. Bellegarda, "Emotion analysis using latent affective folding and embedding," in *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, 2010, pp. 1-9.

[39] A. C. Boucouvalas, "Real time text-to-emotion engine for expressive internet communications," in *Proceedings of International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP-2002)*, 2002.

[40] V. Francisco and P. Gervás, "Automated mark up of affective information in english texts," in *International Conference on Text, Speech and Dialogue*, 2006, pp. 375-382.

[41] S. M. Mohammad, "Sentiment analysis: Detecting valence, emotions, and other affectual states from text," in *Emotion measurement*, ed: Elsevier, 2016, pp. 201-237.

[42] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data*, ed: Springer, 2012, pp. 415-463.

[43] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, pp. 1-135, 2008.

[44] G. Biondi, V. Franzoni, and V. Poggioni, "A deep learning semantic approach to emotion recognition using the IBM watson bluemix alchemy language," in *International Conference on Computational Science and Its Applications*, 2017, pp. 718-729.

[45] C. Whaley, "Security companies might be messing with IT managers' minds," *Computing Canada*, vol. 31, p. 17, 2005.

[46] A. C. Johnston, "FEAR APPEALS AND INFORMATION SECURITY BEHAVIORS: AN EMPIRICAL STUDY," *MIS Quarterly*, vol. 34, pp. 549-566, 2010.

[47] J. C. Richards, J. Platt, and H. Platt, "Longman dictionary of language teaching," *Applied Linguistics*, vol. 288, 1992.

[48] M. Zamanian and P. Heydari, "Readability of Texts: State of the Art," *Theory & Practice in Language Studies*, vol. 2, 2012.

[49] E. Dale and J. S. Chall, "The concept of readability," *Elementary English*, vol. 26, pp. 19-26, 1949.

[50] J. P. Gee, "Three paradigms in reading (really literacy) research and digital media," *Reading at a Crossroads?: Disjunctures and Continuities in Current Conceptions and Practices*, vol. 35, 2015.

[51] S. G. Paris and S. A. Stahl, *Children's reading comprehension and assessment*: Routledge, 2005.

[52] L. F. Cranor, "A framework for reasoning about the human in the loop," 2008.

[53] W. H. DuBay, "The Principles of Readability," *Online Submission*, 2004.

[54] R. Flesch, "A new readability yardstick," *Journal of applied psychology*, vol. 32, p. 221, 1948.

[55] G. H. Mc Laughlin, "SMOG grading-a new readability formula," *Journal of reading*, vol. 12, pp. 639-646, 1969.

[56] L. Feng, N. Elhadad, and M. Huenerfauth, "Cognitively motivated features for readability assessment," in *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, 2009, pp. 229-237.

[57] E. M. Redmiles, M. Morales, L. Maszkiewicz, R. Stevens, E. Liu, D. Kuchhal, *et al.*, "First Steps Toward Measuring the Readability of Security Advice," ed, 2018.

[58] M. Harbach, S. Fahl, T. Muders, and M. Smith, "Towards measuring warning readability," in *Proceedings of the 2012 ACM conference on Computer and communications security*, 2012, pp. 989-991.

[59] A. P. Felt, A. Ainslie, R. W. Reeder, S. Consolvo, S. Thyagaraja, A. Bettes, *et al.*, "Improving SSL Warnings: Comprehension and Adherence," presented at the ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea, 2015.

[60] R. Flesch, "Flesch-Kincaid readability test," *Retrieved October*, vol. 26, p. 3, 2007.

[61] M. A. Scranton, "SMOG grading: a readability formula by G. Harry McLaughlin," Kansas State University, 1970.

[62] S. Zhou, H. Jeong, and P. A. Green, "How consistent are the best-known readability equations in estimating the readability of design standards?," *IEEE Transactions on Professional Communication*, vol. 60, pp. 97-111, 2017.

[63] Webex. (December 28 2020). *https://www.webfx.com/tools/readable/check.php*.

[64] N. Hidayatillah and Y. Zainil, "THE READABILITY OF STUDENTS'TEXTBOOK USED IN SEMANTIC AND PRAGMATIC COURSE IN ENGLISH LANGUAGE EDUCATION PROGRAM OF UNP," *Journal of English Language Teaching*, vol. 9, pp. 144-159, 2020.

[65] P. Heydari and A. M. Riazi, "Readability of texts: Human evaluation versus computer index," *Mediterranean Journal of Social Sciences*, vol. 3, pp. 177-190, 2012.