Human Factors in Phishing Attacks: A Systematic Literature Review

GIUSEPPE DESOLDA, Università di Bari Aldo Moro
LAUREN S. FERRO, Sapienza Universitá di Roma
ANDREA MARRELLA, Sapienza Universitá di Roma
TIZIANA CATARCI, Sapienza Universitá di Roma
MARIA FRANCESCA COSTABILE, Università di Bari Aldo Moro

Phishing is the fraudulent attempt to obtain sensitive information by disguising oneself as a trustworthy entity in digital communication. It is a type of cyber attack often successful because users are not aware of their vulnerabilities or unable to understand the risks. This article presents a Systematic Literature Review (SLR) conducted to draw a "big picture" of the most important research works performed on human factors and phishing. The analysis of the retrieved publications, framed along the research questions addressed in the SLR, helps understanding how human factors should be considered to defend against phishing attacks. Future research directions are also highlighted.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; • **Security and privacy** → **Human and societal aspects of security and privacy**; **Phishing**.

Additional Key Words and Phrases: Phishing, Human Factors, Cybersecurity, Human-Computer Interaction

ACM Reference Format:
Giuseppe Desolda, Lauren S. Ferro, Andrea Marrella, Tiziana Catarci, and Maria Francesca Costabile. 2021. Human Factors in Phishing Attacks: A Systematic Literature Review. In *ACM Computing Surveys*. ACM, New York, NY, USA, Article 0, 34 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Technology is all around us, it is ubiquitous and, as a result, the modern world depends on it daily. However, while its usage has significantly improved our lives, it has also made us vulnerable and accessible to deceptive schemes and exploitation. As technology evolves, so do the procedures that we need to employ to protect and maintain the integrity of information. This impacts the rate at which loopholes are found or new approaches to exploit and access new systems or methods of protection. Contemporary research seeks new ways to anticipate and mitigate threats or, at the very least, identify and resolve them as fast as possible [30] [5] [65] [143]. To this end, two issues arise, which place humans at the center. The first is that humans can only perform a set number of tasks at one time [78]. Secondly, humans are often *the cause* of security breaches by either allowing them to occur or creating them, since they are not adequately supported by the interactive systems in understanding potential risks and threats [37][75][104].

It is predicted that global cybercrime costs will grow by 15 percent per year since 2021, reaching $10.5 trillion USD annually by 2025, up from $3 trillion USD in 2015 [127]. In this context, cyber attacks are emerging as problems caused

Manuscript submitted to ACM

not only by technological aspects but also by human factors neglected when designing interactive systems. The IBM Cyber Security Intelligence Index Report highlighted that close to 95% of security incidents are due to human errors [57]. One of the most spread and effective cyberattacks that leverages human factors is *phishing* [58]. In 2010 Google detected an average of 317 phishing websites per day, but in 2020 the number jumped to 5789, representing an increment of 1726% in a decade, and 25% more in reference to 2019 [15].

This article presents a systematic literature review (SLR) on human factors in phishing attacks. The SLR was conducted according to the Kitchenham methodology [67] and retrieved 52 key publications of interest, published in the period from 2001 to May 2019; they were analyzed along three dimensions, each one related to one of the three research questions that drove the SLR. This article provides several contributions to researchers interested in phishing attacks. First, it shows how human factors are exploited by attackers. More importantly, it helps understanding how human factors can be considered to defend against phishing attacks by reporting solutions proposed in the literature, which proved useful to limit the effectiveness of such attacks. It also highlights future research directions that deserve more attention, in order to help researchers and practitioners to propose new strategies and methods to better defend people and organizations against continuously increasing phishing attacks. Training users emerged as one of the most adopted solutions to limit these attacks; in particular, games were reported as valid methods for training. However, further research is needed for making user training more effective.

The article is organized as follows. Section 2 introduces some background concepts and clarifies the rationale of the SLR. Section 3 reports the details of the first two phases of the SLR, i.e., planning and conducting. Section 4 refers to the third phase of the SLR, reporting and analyzing the resulting publications. Section 5 discusses some very recent works, published from June 2019 to May 2021. Section 6 describes future challenges for researchers working on human factors in phishing. Section 7 concludes the article.

2 BACKGROUND

This section introduces some background on the concepts the SLR is based on, namely, cybersecurity, human factors and phishing. At the end, literature reviews and surveys already existing on phishing as well as on human factors in cybersecurity are briefly reported, in order to clarify the rationale of this research work.

2.1 Cybersecurity

A thorough literature review, as presented in this article, revealed that the term “cybersecurity” is broadly used to encapsulate any security topics related to information and communication technology. The Oxford English dictionary defines cybersecurity as "the state of being protected against the criminal or unauthorized use of electronic data, or the measures taken to achieve this" [93]. The definition provided by the Merriam-Webster dictionary also refers to the "measures taken to protect a computer or computer system (as on the Internet) against unauthorized access or attack" [82]. This entails the safeguarding of computer systems and the information they contain from penetration and malicious damage (e.g., hackers) or disruption (e.g., infected networks) [74]. In other perspectives, the concept of cybersecurity focuses on security concepts, security safeguards, policies, guidelines, risk management approaches, best practices, tools and technologies that can be used to protect cyber environment, organization and user assets [59]. In addition, cybersecurity involves reducing the risk of malicious attacks to software, computers and networks [9]. This includes the use of tools (e.g., anti-virus software, firewalls, good cyber-hygiene) to detect break-ins, stop viruses, block malicious access, enforce authentication, and enable encrypted communications.

Based on the above, in this article we consider the term cybersecurity as encompassing not only the act of protecting oneself and/or organizations against online attacks to their assets, but also what can be used to enforce and maintain the practice of safe online behavior.

Lastly, we acknowledge the interchangeability of the term with both "cybersecurity" and "cyber security"; and in some cases "cyber-security" being accepted and used within related content. However, for this paper, we have chosen to use its attached form *cybersecurity*.

2.2 Human Factors

A secure system relies on human users doing the right things according to cybersecurity policies. Therefore, it is important to understand how *human factors* may create weaknesses in a system that an attacker could exploit. We intend to consider how human factors can be addressed to improve approaches to cybersecurity. The International Ergonomics Association defines human factors as the "scientific discipline concerned with the understanding of the interaction among humans and elements of a system" [14]. In this field, theories, principles, methods and data are applied to improve human well-being and system performances. Being the discipline of human factors centered on users, it also considers the devices (e.g., PC, Mobiles, etc.), tasks (e.g., browsing the web), and environments (e.g., home or office) of the user.

Human factors have been studied in several areas, such as environmental design, health care, and also in cybersecurity [116]. Great attention to human factors has been devoted to the domain of aviation. *The Dirty Dozen* proposed by Dupont refers to twelve of the most common errors in maintenance activities due to specific human factors [44]; these errors are reported in the aviation domain as possible causes of accidents or incidents. However, there is neither a concise list of human factors within cybersecurity, nor a definitive description of relevant human factors. Thus, we have chosen to utilize the list outlined by Dupont as we consider it a valid basis for our investigation:

- **Lack of Communication**: people not communicating with each other within a working and/or online environment.
- **Complacency**: a feeling of self-confidence that can lead to a lack of awareness of potential dangers.
- **Lack of Knowledge**: not having specific knowledge and enough experience that can lead to poor decisions.
- **Distraction**: when a user's attention has been taken away from the task that they are required to do.
- **Lack of Teamwork**: not providing enough support towards a group of people, co-workers, etc, who rely on your support.
- **Fatigue**: it is a physiological reaction resulting from prolonged periods of work and stress.
- **Lack of Resources**: not having enough resources (e.g., time, tools, people, etc.) to complete a task.
- **Pressure**: pressure to meet a deadline interferes with our ability to complete tasks correctly.
- **Lack of Assertiveness**: not being able or allowed to express concerns or ideas.
- **Stress**: acute and chronic stress from working for long periods or other demanding issues such as family or financial problems.
- **Lack of Awareness**: not being aware of what happens in the surrounding (working or online) environment, often leading to an unconscious disconnection from what others are doing.
- **Norms**: workplace practices that develop over time, which can then influence other behaviors.

Even if in recent times publications on human factors within cybersecurity are gaining momentum, a wider picture to understand the current state of human factors within cybersecurity is still missing. This article is a contribution in this direction, since it analyzes human factors in relation to a specific type of cybersecurity attack, namely *phishing*.

2.3 Phishing

Cybersecurity attacks can be successful because users are not aware of their vulnerabilities and because of their *Lack of Knowledge* about consequences and risks. With social media users tending to share a great deal of their lives online, it is significantly easier for attackers to find ways to gather information about users and use it in ways to "convince" them that their identity and intentions are legitimate.

Phishing is one of the most effective cyber attacks. Various authors define phishing in slightly different ways. In this article, we adopt the definition of phishing proposed by Lastdrager in his literature survey on phishing attacks [72], where phishing is defined as: "a scalable act of deception whereby impersonation is used to obtain information from a target". A phishing attack usually consists of sending a message (e.g., an email), which appears as from a reputable organization (e.g., a bank), sounds urgent, claims to enclose important information, and invites the victims to open a website that is a clone of the original one (e.g., a clone of their own bank website). In the message, the victims are invited to provide personal information on the website, for example, they are required to login to the website for updating their profile information. In most cases, the victims are unlikely to check or question the website validity; thus, they open it providing the required information that, unfortunately, is stolen by the attackers who can use it, for example, to enter on the bank account on behalf of the victims to steal their money.

Phishing is often successful due to the vulnerabilities of users, namely to human factors related issues. A phishing attack requires preparation for it to be successful, which involves studying users, their behavior, their online posts, and even watching them online to gain valuable knowledge to use in a targeted approach. Also target websites are considered by the attackers to improve the effectiveness of their attacks, for example, by looking at the schedule where servers go down for (regular) maintenance, their content, etc. On one side, knowledge about victims allows attackers to send customized messages, for example, fans of a soccer team may receive an email regarding an offer to purchase their club's jerseys (in this case, the phishing website aims to steal their credit card information). On the other side, knowledge about target websites can improve the trustability on phishing messages; for instance, if a website has scheduled maintenance, its users may receive an email asking them to unlock their account after the maintenance by clicking on a link and sign-in on the website (in this case, the phishing website aims to steal users' credentials).

Phishing does not only occur via email. In several cases, users receive fraudulent phone calls or text messages that appear as though they have legitimately come from a company that the user has an account or service with. As a result, some users are unlikely to check or question their validity and unknowingly hand over their personal data to scammers, hackers, and other persons with malicious intentions. In addition, as a response to continuous updates of defensive solutions against phishing attacks, increasingly sophisticated and diversified attacks are proposed. A comprehensive overview of the variants of phishing attacks is reported by Chiew et al. [30], where a classification of the main components characterizing this attack is derived (see Figure 1). This classification firstly identifies the *medium*, which is used to start the attack, namely, Internet, SMS and voice. Each medium may use a *vector*, i.e., the vehicle for launching the attack. Examples of vectors for the Internet are e-mail, eFax, instant messaging (e.g., social network messages), websites. The last layer of this classification is called *technical approaches* and reports all the technological solutions available to deploy a phishing attack, for example, JavaScript obfuscation, man-in-the-middle, SQL injection. Each vector can exploit one or more of these technical approaches to perform the attack. From this classification, it

emerges that a phishing attack is very complex and it can be performed in several ways. This has led to the spreading of new terms indicating variations of phishing attacks. Some of the most popular variants are:

- **Spear-Phishing**: it is the fraudulent practice of *sending emails* that appear to be from a known or trusted sender and *are targeted* to individuals to reveal confidential information.
- **Vishing**: it is the fraudulent practice of making *phone calls* or leaving *voice messages* appearing to be from reputable companies to obtain personal information
- **SMiShing**: also known as SMS phishing, it consists of sending SMS or instant messages on victim's smartphone; such messages appear as sent by a trusted source (e.g., our bank) and invite the victim to click a link as in the case of more traditional phishing, or even to download an attachment. In this last case, the attachment installs malware like a rootkit or a backdoor to guarantee the scammers to access to everything (contacts, email messages, application data, etc.).
- **Pharming**: it is a scamming practice where malicious code is installed on a personal computer or server, which in-turns misdirects users to fraudulent sites without their knowledge or consent.

A phishing attack is successful for several reasons. For example, the attacker can gain an advantage by performing reconnaissance about users and/or the company they work for. This information allows the attacker to communicate with a user by using familiar terms or colloquial phrases. As a result, the attackers can improve the "legitimacy" as being someone the user can trust. Therefore, users are likely to feel more comfortable with who they are interacting with and subsequently lower their guard. The COVID-19 pandemic has provided the opportunity for increasing attacks based on the fear of users and their working situation (i.e. working from home) [47][84]. According to data gathered

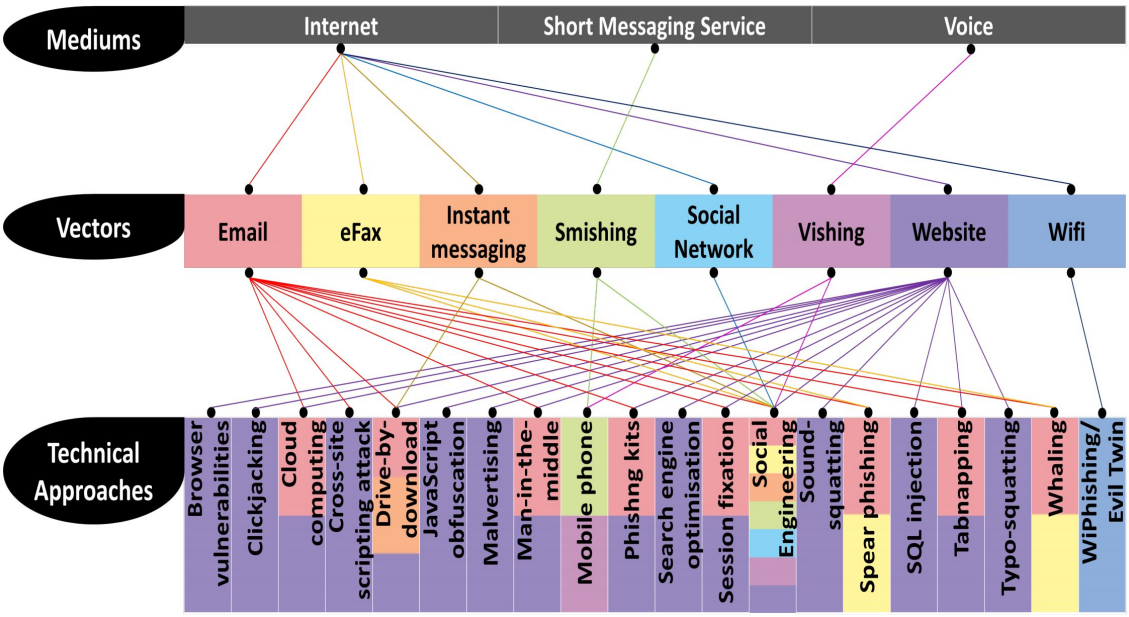

Fig. 1. A framework reporting the main components of a phishing attack and the relationships among them [30]

and analyzed by Atlas VPN[1], the number of phishing websites spiked by 250% amid COVID-19 quarantine [21], with 18 million scam emails being blocked by Google daily [125], and thousand of malicious coronavirus-related websites being created daily [21]. The attack increase is echoed by several security experts and companies [123]. However, while users' knowledge and awareness of phishing and security-related issues should be improved, other human factors issues should be addressed. For example, high-pressure workplaces (resulting in high levels of *Stress*) or situations with a strong cultural influence (e.g., *Norms*) should be considered in the analysis, and approaches to mitigate the risks have to be devised.

2.4 Literature reviews and surveys on phishing

The complexity and importance of phishing attacks have been demonstrated by increasingly growing research activity that is framed in different surveys and literature reviews. Although phishing attacks exploit social and psychological aspects of victims, most of the research is focused on technical aspects (see for example [30]). In the survey reported in [5], the authors evaluated and compared different techniques, focusing on machine learning solutions, to detect phishing emails. In [65], it has been proposed a survey on phishing mitigation techniques like detection, offensive defense, correction, and prevention. To the best of our knowledge, that survey is the only one that marginally covers human factors in phishing, as part of the defensive mechanisms. However, similar to the other surveys, the emphasis is on the technical and procedural aspects.

A few literature reviews deal with the topic of human factors in cybersecurity, marginally covering phishing attacks. For example, in [143] papers on human factors and their related issues in cybersecurity and privacy are reviewed, focusing the attention on big data. In particular, human factors are analyzed by considering desktop behaviors, mobile behaviors, and online behaviors, and security and privacy issues in daily human practices are identified and used to propose both users' behavioral patterns and solutions to detect abnormal, vulnerable and malicious actions. A literature review on information security culture is presented in [54]; a framework summarizing human factors that contribute to the security culture of an organization is also proposed.

Human factors play a central role in phishing attacks, which have rapidly and dramatically increased in the last years their number as well as their effectiveness. However, to date, there is no publication that surveys the literature on human factors in phishing attacks; this SLR aims to remedy this lack.

3 PLANNING AND CONDUCTING THE SLR

To identify and classify the publications on human factors within the context of phishing, we conducted a Systematic Literature Review (SLR) via a scientific, thorough, and reproducible approach. According to Kitchenham [67], the SLR requires 3 phases: *planning*, *conducting* and *reporting*. This section describes the planning and the conducting phases, while the reporting phase is detailed in Section 4.

3.1 Planning the SLR

Planning included the following activities:

(1) Formulation of research questions to scope the search;
(2) Definition of the search strings;
(3) Selection of data sources;

[1]https://atlasvpn.com

(4) Definition of inclusion criteria.

3.1.1 Formulation of the research questions. The main goal of our SLR is to assess the current state of research about human factors in phishing attacks and to understand how human factors could be exploited to mitigate the impact of *phishing attacks*. With this objective in mind, we formulated the following research questions:

RQ1: Which specific topics relating to human factors in phishing attacks are discussed within the literature?
RQ2: What solutions exist that address human factors to reduce the success of phishing attacks?
RQ3: What are the main vulnerable human factors in phishing attacks?

RQ1 focuses on identifying which specific topics relating to human factors on phishing attacks are concretely tackled by the research literature. RQ2 investigates the solutions that, addressing human factors, are or have been used to reduce the success of phishing attacks in the event where information has been given to the scammer, which has resulted in some form of loss and/or damage (e.g., financial, data, etc.). RQ3 aims to ascertain which human factors are more susceptible to phishing attacks; this question is key to access the current landscape of how human factors are considered within the context of phishing.

3.1.2 Definition of the search strings. We defined 28 search strings by deriving terms from: (i) an initial assessment of some of the most cited papers dealing with phishing within the field of cybersecurity, and *(ii)* our knowledge of the subject matter. We first determined that the term *phishing* is very generic and it would allow us to retrieve the majority of publications in this area. As important terms we also selected *malware*, *hacking*, and *password*, since they are often used in relation to phishing attacks. The rationale for choosing them is:

- *malware*: it involves any software intentionally designed to cause damage and steal personal information from a computer, server, client, or computer network. A wide variety of malware exists (e.g., virus, trojan, rootkit, etc.) and some types of phishing involve malware, e.g., SMiShing invites victims to download on their devices attachments that install malware to gain access to all the device data.
- *hacking*: it is one of the most common and generic terms used in cybersecurity to indicate methods, techniques and operations aimed at exploiting weaknesses in a computer system or network to gain unauthorized access to sensitive information.
- *password*: a user's password (and account information) is often the first piece of information stolen by the attacker with a successful phishing scam.

It is worth observing that variations in spelling (e.g., behaviour/behavior; modeling/modelling) were also included within the searches to ensure that relevant papers were not excluded or overlooked.

Since the above four terms are very generic and commonly used in different research fields, to scope our search in the area of human-computer interaction we decided to combine each of them with the terms "human factors", "human computer interaction", "HCI", "behavio(u)r mode(l)ling" and "user model(l)ing". It is worth noticing that variations like "human-computer interaction" (with the hyphen) do not need to be included in the query, since the search engines do not take the hyphen into account and thus there is no impact on the results. On the contrary, acronyms such as "HCI" have to be used with their longer version "human computer interaction", since in many cases they return varying results. We intentionally excluded the terms "cybersecurity/cyber security/cyber-security" because these terms are too general for a focused search.

According to these criteria, the final search strings used for our search were the following:

(1) "human factors" *and* "x"
(2) "HCI" *and* "x"
(3) "human computer interaction" *and* "x"
(4) "behavior modeling" *and* "x"
(5) "behaviour modelling" *and* "x"
(6) "user modeling" *and* "x"
(7) "user modelling" *and* "x"

where "x" is substituted by either "phishing" or "malware" or "hacking", or "password", for a total of 28 strings. For example, the first 4 strings are: *"human factors" and "phishing'*; *"human factors" and "malware'*; *"human factors" and "hacking'*; *"human factors" and "password"*.

3.1.3 Selection of data sources. Two types of data sources have been considered as a ground of this SLR: (1) scientific digital libraries, and (2) a selection of conference proceedings and journals.

The examined libraries were ACM Digital Library[2], IEEE Xplore Digital Library[3], Springer Link[4], Elsevier Science Direct[5], Scopus[6], and Google Scholar[7]. During the definition of the search strings, we discovered that the search engines of the most popular scientific libraries performed differently when specifying the search string. The same search needed to be performed in various ways depending on the library (i.e., by using different syntax). Each library contained various options for searching content. For example, they permitted to search for keywords within an article's title. In some cases, other options allowed for the same keywords to be searched within abstracts, full text, and/or a combination (e.g., title, abstract, and keywords).

All searches were conducted on the entire databases, without being confined to a particular area over another. This was due to the multidisciplinary nature of research that human factors on cybersecurity encompasses, spanning various fields such as psychology, engineering, security, and information technology.

All searches utilized relative syntax. All terms that consisted of more than one word, for example *human factors*, were included in quotation marks" " (e.g. "human factors"). In addition, to ensure that terms were searched together, the Boolean operator "AND" was included for querying the digital libraries, with the exception of the ACM Digital library that required the "+" symbol instead of AND.

The other data sources we searched included the proceedings of nine specific conferences and four important journals relevant for the topics of the SLR, namely:

- ACM Conference on Human Factors in Computing Systems (CHI), ACM User Interface Software and Technology (UIST), ACM Intelligent User Interfaces (IUI), ACM Designing Interactive Systems (DIS), IFIP TC13 Conference on Human Computer Interaction (INTERACT), Advances in Human Factors in Cybersecurity (AHFE),
- Human Factors and Ergonomics Society Annual Meeting (HFES), Workshop on Usable Security Privacy (USEC), European Workshop on Usable Security (EuroSEC).

[2]https://dl.acm.org/
[3]https://ieeexplore.ieee.org/
[4]https://link.springer.com/
[5]https://www.sciencedirect.com
[6]https://www.scopus.com/
[7]https://scholar.google.com/

- IEEE Transactions on Human-Machine Systems, ACM Transactions on Computer-Human Interaction (TOCHI), International Journal of Human-Computer Studies (IJHCS), Journal of the Human Factors and Ergonomics Society (HFES).

Despite the fact the IEEE Security and Privacy Magazine is a technology-oriented journal, we analyzed the articles of the special issue on *Secure or Usable?*, because it is highly cited in the literature [38].

3.1.4 Definition of inclusion criteria. The last activity, as suggested, e.g., by [49, 90, 101, 126], consists of defining inclusion criteria to ensure an unbiased selection of relevant publications. Therefore, a publication is retained if it satisfies all of the following inclusion criteria:

- IN1: the publication deals with phishing related issues that take into account human factors (i.e., publications in which human factors are only mentioned without being an important topic of the research were excluded);
- IN2: the publication has been peer-reviewed;
- IN3: in case a study is reported in more publications (presenting partial results), the publication reporting the most complete study is retained;
- IN4: the publication is written in English.

3.2 Conducting the SLR

After planning, the SLR is conducted. By taking into account what was suggested by Kitchenham et al. [67], we performed two main activities, called *literature review execution* and *data synthesis*; they are described in Section 3.2.1 and in Section 3.2.2, respectively.

3.2.1 Literature review execution. This activity was performed during April-May 2019 and followed a three-phase search process:

(1) **Phase 1 - Digital library search**: we searched each of the digital libraries indicated in Section 3.1.3 using the search strings described in Section 3.1.2;
(2) **Phase 2 - Conference and Journal search:** we searched journals and conferences mentioned in Section 3.1.3 by using the search strings described in Section 3.1.2;
(3) **Phase 3 - Backward snowballing search:** we checked references and citations of the publications resulting from the previous two phases, in order to identify possible relevant publications. This method is suggested in [134].

The search across the digital libraries (Phase 1) yielded a total of 1857 potentially relevant publications. The number of publications resulting from the search of each digital library can be seen at https://cutt.ly/gnGcV69. After duplicate removal, a dataset of 1210 publications was obtained. In order to make feasible and accurate the analysis of the publications, we decided to consider only the publications from 2014, thus reducing the dataset to 698 publications. It is worth remarking that this does not mean that papers published before 2014 are not considered in this SLR, because Phase 3 allowed us to retrieve further publications, without any time constraint, by analyzing the references of publications selected in Phase 1 and Phase 2. The 698 publications were then analyzed by reviewing their abstract, introduction, and conclusion, taking into account the inclusion criteria. In some cases, the whole paper was reviewed to determine if it was appropriate for the SLR. At the end, Phase 1 resulted in a total of 23 publications. The details of this first phase, and all the results, can be found at https://cutt.ly/OnGvkOk.

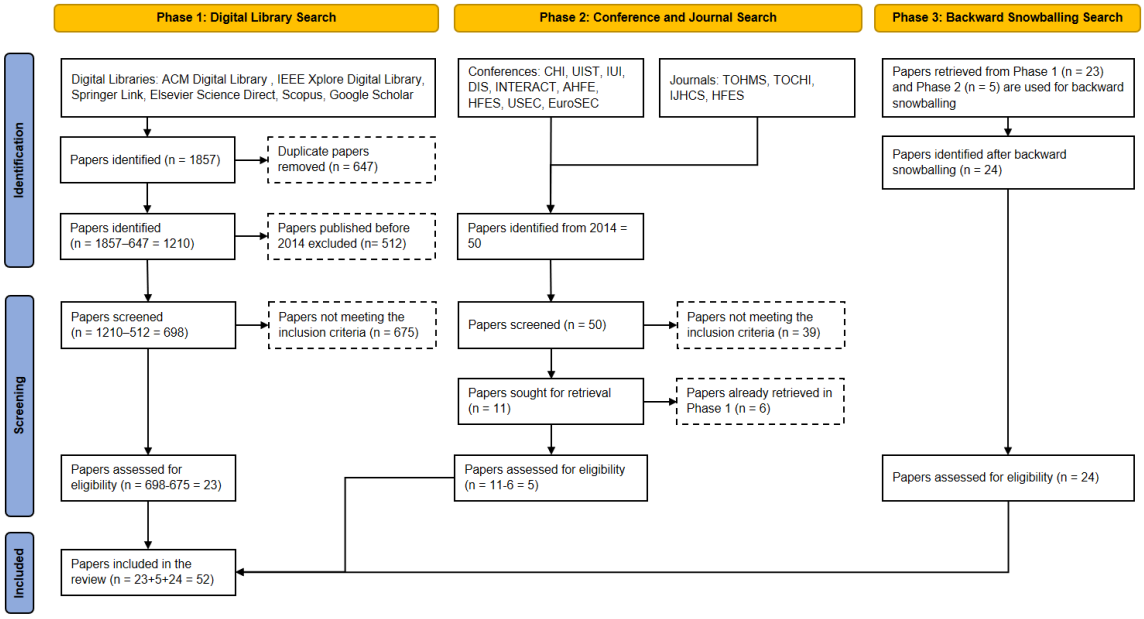

Fig. 2. Flow diagram summarizing the selection of the publications along the 3 search phases

Phase 2 consisted of searching on conference proceedings and journals indicated in Section 3.1.3 by using the same search strings of Phase 1. Again, the search was restricted to publications from 2014. Publications in the special issue mentioned in Section 3.1.3 were also analyzed. As result, a further 50 potentially relevant publications were found. The specific numbers for each journal and each conference resulting from the search can be seen at https://cutt.ly/gnGcV69. Once we reviewed these publications and applied the inclusion criteria, a total of 11 publications resulted relevant. We added them to the set resulting from Phase 1 and, after removing 6 duplicates, the total number of 28 publications was obtained. The details of this search phase, and all the results, can be found at https://cutt.ly/gnGv01M

The third and last search (Phase 3) used the *"Backward snowballing method"* [134], which consists of using the references in the publications already obtained and their citations, in order to identify additional papers. As remarked before, this phase mitigates possible limitations of Phase 1 and Phase 2, since it permits to retrieve relevant papers published before 2014. No further iterations of the backward snowballing method were executed. At the end of this phase, 24 additional publications were identified, bringing to 52 the total number of publications retrieved by the SLR. The details of this third search, and all the results, can be found at https://cutt.ly/SnGbkJt. Figure 2 summarizes the three search phases.

3.2.2 Data synthesis. The 52 publications resulting from the three search phases are listed in an Excel spreadsheet available at https://cutt.ly/gnGcV69, in alphabetical order of the title. They are also shown in Table 1, listed in alphabetical order of the first author's last name and indicating the reference number in the references at the end of this article.

Some important considerations emerge from the search process and its results. The first being that the use American or British spelling is an important factor, as in some cases the number of results varied significantly based on the spelling of the search terms. One must search using both to ensure that no publications are missed. The second is

Table 1. The list of the 52 publications selected by the SLR in the period from 2001 to May 2019

Authors	Reference	Authors	Reference
Abroshan et al. (2018)	[1]	Marforio et al. (2016)	[79]
Alissa et al. (2018)	[4]	Mayer et al. (2017)	[80]
Alohali et al. (2017)	[6]	McElwee et al. (2018)	[81]
Alsharnouby et al. (2015)	[7]	Metalidou et al. (2014)	[83]
Arachchilage et al. (2016)	[11]	Ndibwile et al. (2017)	[85]
Ardi and Heidemann (2016)	[12]	Noureddine et al. (2017)	[87]
Asfoor et al. (2018)	[13]	Nurse (2018)	[89]
Avery et al. (2017)	[16]	Oliveira et al. (2017)	[91]
Benias and Markopoulos (2018)	[17]	Onarlioglu et al. (2012)	[92]
Canfield et al. (2016)	[23]	Pham et al. (2017)	[97]
Canova et al. (2015)	[24]	Safa et al. (2015)	[105]
Chen et al. (2018)	[28]	Sasse et al. (2001)	[108]
Choong and Theofanos (2015)	[31]	Sawyer and Hancock (2018)	[109]
Corradini and Nardelli (2018)	[36]	Sheng et al. (2010)	[111]
Downs et al. (2007)	[41]	Sheng et al. (2007)	[112]
Downs et al. (2006)	[42]	Stainbrook and Caporusso (2018)	[117]
Flores and Ekstedt (2012)	[50]	Stanton et al. (2005)	[118]
Gangire et al. (2019)	[52]	Steves et al. (2019)	[119]
Jansson and von Solms (2013)	[60]	Wang et al. (2012)	[128]
Jensen et al. (2017)	[62]	Wash and Cooper (2018)	[129]
Karakasiliotis et al. (2006)	[64]	Wen et al. (2019)	[130]
Kirlappos and Sasse (2015)	[66]	Williams and Li (2017)	[131]
Kumaraguru et al. (2009)	[68]	Williams et al. (2018)	[132]
Kumaraguru et al. (2010)	[69]	Williams et al. (2017)	[133]
Lévesque et al. (2018)	[73]	Xiong et al. (2018)	[137]
Lim et al. (2016)	[76]	Zhao et al. (2016)	[145]

related to the use of HCI as an acronym of human-computer interaction; they returned varying results, suggesting that acronyms for popular terms should be also searched when their longer version is used. As a further consideration, the not very high number of retrieved papers (52) indicates that this topic is still novel. However, the analysis of the temporal distribution, shown in Figure 3, indicates that the interest in human factor in phishing attacks is growing, with a sharp increase from 2015. The number of publications in 2019 is not meaningful since the literature review was executed during April-May 2019.

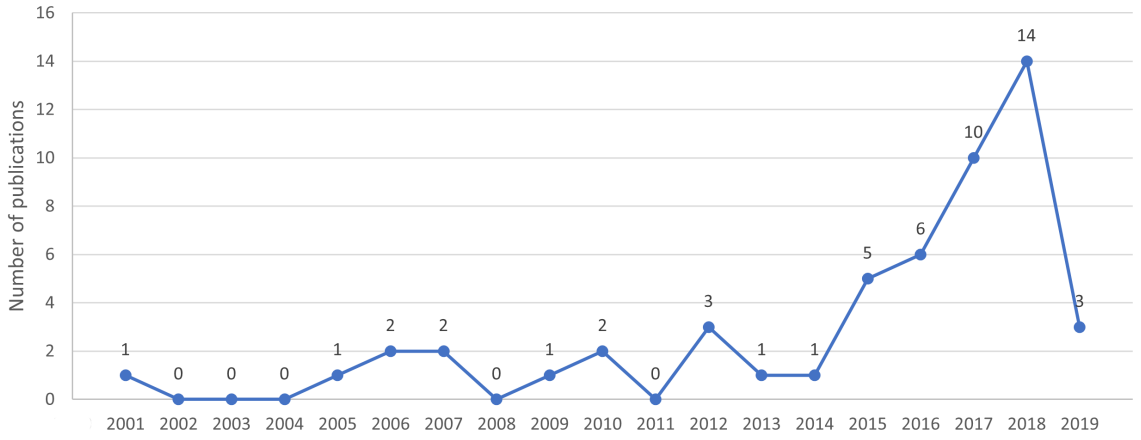

Fig. 3. Distribution over time of the publications selected by the SLR up to May 2019

4 REPORTING AND ANALYZING THE SLR RESULTS

This section reports the analysis of the SLR results framed along the three research questions presented in Section 3.1.1.

4.1 Which specific topics relating to human factors in phishing attacks are discussed within the literature? (RQ1)

In order to be as objective as possible in identifying the topics, we analyzed the retrieved publications based on the Latent Dirichlet Allocation (LDA) technique [20], widely recognized as one of the most effective techniques for latent topics distribution within a corpus [61]. LDA is an unsupervised generative probabilistic method for modeling a corpus of documents as a set of topics. It was elaborated by Blei, Ng and Jordan in 2003 [20], and today is one of the most popular methods in topic modeling. The term latent refers to the fact that topics are not known a priori, they are hidden in the data, i.e., the 52 publications in our case. The idea is that a set of documents refers to various topics and every topic is characterized by various terms. LDA has been widely used in different domains like medical sciences (e.g., [144][63][95][136]), geography (e.g. [39][46][124][141][114]), and political science (e.g. [27][35][56]).

Table 2. The 5 topics most discussed in the 52 selected papers, as emerged from the LDA analysis, and their related terms with weights

Topic No.	Topic Label	Weighted terms
1	Policies	0.025*password + 0.015*human + 0.010*behavior + 0.010*policy + 0.010*indicator + 0.009*model + 0.009*system + 0.009*employee + 0.007*design + 0.007*work”
2	Organization Susceptibility	0.017*behavior + 0.017*organization + 0.014*research + 0.009*end + 0.008*knowledge + 0.008*result + 0.008*individual + 0.007*base + 0.006*training + 0.005*survey
3	User Awareness	0.011*factor + 0.009*people + 0.008*awareness + 0.008*employee + 0.007*system + 0.007*technique + 0.007*human + 0.007*behavior + 0.006*influence + 0.006*risk
4	Modelling	0.011*human + 0.007*internet + 0.007*model + 0.007*threat + 0.007*website + 0.007*behaviour + 0.007*training + 0.006*well + 0.006*base + 0.005*system
5	Defensive Mechanisms	0.011*training + 0.010*website + 0.010*system + 0.009*result + 0.009*game + 0.009*participant + 0.008*identify + 0.008*design + 0.008*people + 0.007*factor

The application of LDA to identify the topics in the 52 papers is described in the Appendix. The resulting 5 topics are listed in Table 2. The first column reports a number that identifies the topic. The topic label in the second column has been determined by manually analyzing the papers referring to each one of the five topics; indeed, even if the LDA automatically identifies topics within a corpus of documents associating each paper to one or more topics, the topic label must be defined by the analysts. The label has been defined as the word(s) that, according to the authors of this article, better summarize(s) the topic in the associated papers. The third column reports the terms related to each topic, as they result from the LDA analysis. For each term, the weight is indicated: it is a number ranging from 0 to 1 that provides the probability of the presence of that term in the papers for which the topic is relevant. Table 3 reports the 52 papers and their topics. It is worth remarking that the topics identified for and associated with each paper are not exclusive since a paper can be associated with one or more topics. For example, paper with reference [50] is associated with three topics while paper [52] is associated with only one. In the rest of this section, the five topics are discussed with reference to some representative papers.

Topic 1 - Policies. Policies for privacy and security are often suggested or imposed by organizations, public administrations and companies to regulate the use of interactive systems by their employees or users in general (e.g, citizens, students). One of the most common examples regards policies for passwords, typically provided to generate and manage safe credentials. However, despite passwords being one of the most adopted and strong authentication mechanisms, it poses several challenges from the point of view of the users due to the complex policy requirements (e.g., length, character usage, expiration, reuse). In addition, besides satisfying such requirements, users have to manage multiple passwords, resulting in a high cognitive overload. This problem has been addressed for example by Abroshan et al. [1], who performed a set of studies in real organizations to investigate the behavior of employees when dealing with their organization password policies; they found that employees generated weak passwords, reused the same or similar ones and, in the worst cases, wrote down their password, e.g., on a post-it. Thus, even with strong security policies, an organization remains exposed to cyberattacks due to these human errors. This misalignment between policies and humans has been researched also by Alissa et al. [4]: the authors proposed a psychometric tool to measure humans concerning cybersecurity policies. This questionnaire systematizes a set of measures (e.g., password, email, identity), identifies for each of them a list of policies (for a password, for example, the requirements to generate strong passwords like length and the inclusion of special characters) and evaluate the user through a set of questions (e.g., "Do you prefer to use the default password?"). This questionnaire can be administered to users (e.g, company employees, citizens, university students) to investigate their behavior when dealing with cybersecurity policies provided by an organization. The questionnaire results should provide useful indications to prevent or improve wrong behaviors (e.g., if it results that people adopt default passwords, the systems should impose users to set up a different password). Password policies can be also complemented with an automatic check of the password robustness. For example, Marforio et al. [79] created a predictive model to classify the user-generated password in terms of convenience and security. This model can be used in conjunction with policies to assist users in following the right strategies for password generation.

Another interesting aspect that should be considered when an organization defines policies for privacy and security is the 'prevalence effect', which has been largely investigated in psychology as the phenomenon where a human is more likely to miss a target with a low frequency than a target with a high frequency [135]. Sawyer and Hancock analyzed this effect in a study involving 300 participants exposed to phishing emails at rates of 1%, 5% and 20% [109]. The results demonstrated that the prevalence effect exists in phishing cyber attacks, since participants that received the 1% of phishing emails result in lower accuracy in detecting the attack. This underlines the difficulty for users to identify

Table 3. Associations between papers and their topics

Reference	Topics and Weights	Reference	Topics and Weights
[1]	[(3, 0.999)]	[79]	[(1, 0.830), (5, 0.169)]
[4]	[(1, 0.999)]	[80]	[(5, 0.999)]
[6]	[(3, 0.999)]	[81]	[(2, 0.999)]
[7]	[(5, 0.646), (3, 0.353)]	[83]	[(3, 0.999)]
[11]	[(5, 0.694), (1, 0.305)]	[85]	[(4, 0.999)]
[12]	[(5, 0.999)]	[87]	[(1, 0.999)]
[13]	[(3, 0.999)]	[89]	[(4, 0.999)]
[16]	[(5, 0.999)]	[91]	[(5, 0.999)]
[17]	[(1, 0.999)]	[92]	[(4, 0.999)]
[23]	[(5, 0.770), (1, 0.229)]	[97]	[(4, 0.999)]
[24]	[(4, 0.525), (5, 0.474)]	[105]	[(4, 0.999)]
[28]	[(5, 0.999)]	[108]	[(1, 0.999)]
[31]	[(1, 0.999)]	[109]	[(4, 0.910), (1, 0.089)]
[36]	[(3, 0.5724748), (4, 0.426)]	[112]	[(5, 0.999)]
[42]	[(3, 0.999)]	[111]	[(5, 0.999)]
[41]	[(3, 0.542), (5, 0.456)]	[117]	[(5, 0.573), (1, 0.426)]
[50]	[(2, 0.516), (1, 0.342), (4, 0.141)]	[118]	[(2, 0.999)]
[52]	[(4, 0.999)]	[119]	[(4, 0.999)]
[60]	[(2, 0.746), (3, 0.253)]	[128]	[(2, 0.999)]
[62]	[(2, 0.999)]	[129]	[(4, 0.662), (5, 0.337)]
[64]	[(3, 0.999)]	[130]	[(5, 0.507), (2, 0.492)]
[66]	[(3, 0.890), (2, 0.109)]	[131]	[(3, 0.999)]
[68]	[(5, 0.9999)]	[132]	[(3, 0.999)]
[69]	[(5, 0.999)]	[133]	[(4, 0.999)]
[73]	[(5, 0.999)]	[137]	[(5, 0.999)]
[76]	[(3, 0.999)]	[145]	[(5, 0.999)]

phishing emails sent at a low rate, thus organization policies should warn users to pay attention also in case of suspect emails with a low rate.

Topic 2 - Organization Susceptibility. Several factors affect the susceptibility of organizations to phishing attacks. For example, the spear-phishing technique aims to fraud specific organization employees by sending them emails that appear to come from a known or trusted sender. Lévesque et al. [73] investigated the behavior of individuals that interact with spear-phishing emails. Their study demonstrates that visceral triggers (e.g., the urgency of response), phishing deception indicators (e.g., grammar and spelling errors), and phishing knowledge strongly influence the employees when exposed to phishing. In addition, McElwee et al. [81] investigated how organization security managers can limit the employees' susceptibility to phishing attacks. The ground of this analysis was a four-year study performed in an organization that simulated phishing exercises to train employees on phishing attacks. This study revealed that providing repeated and targeted exercises is indeed a good training to limit employees' susceptibility to phishing attacks. It also shows that incentives like bonuses and rewards given for successfully spotting training on phishing attacks might not be effective to reduce employees' susceptibility.

In another study performed within an organization, Flores and Ekstedt researched how organizational and human factors are related in the context of security [50]. Domain experts were interviewed and their answers were analyzed following an inductive approach that led to the identification of a model, which is characterized by six constructs related

to (1) information security leadership, (2) organizational structures, (3) information security processes, (4) security knowledge transfer, (5) shared organizational security, and (6) knowledge. This model can be adopted by organizations to shape their internal structure and processes, to improve their security. Another contribution to defend organizations from phishing attacks comes from [60], which demonstrates the effectiveness of simulated phishing attacks combined with training as a solution to limit the effectiveness of phishing attacks.

Topic 3 - User Awareness. One of the most critical aspects considered in this field is the awareness of users to phishing attacks. For example, Williams et al. highlighted that different human aspects (e.g., self-deception, self-control, self-awareness, trust, motivation, expertise) and contextual factors (e.g, emotions, culture, organization) impact on users' awareness [131]. This analysis has been distilled in a holistic framework that summarizes all the potential relationships among human aspects, contextual factors and existing types of phishing attacks.

Williams et al. studied how organizations attempt to raise awareness of spear-phishing emails among their staff [132]. In order to analyze message-related factors that cheat users, nine spear-phishing emails were sent to 62000 employees over 6 weeks. The results revealed that including the authority cues in a phishing email increases the possibility that a user clicks a fraudulent link in the email. The authors also performed six focus groups in a different organization to explore additional human and contextual factors impacting user awareness of phishing. The most important factors that emerged are: the degree of exposure to external emails, the use of centralized inboxes, information overload within the work environment, and the role of social and technical support in enhancing the perception of self-efficacy. Similarly, Asfoor et al. studied bank customers to understand which human factors mostly influence the awareness of phishing attacks [13]. The resulting human factors were security concerns, security attentiveness, competency, IT knowledge, years of PC usage, and gender.

Topic 4 - Modelling. The study of models plays an important role in preventing phishing attacks. To design interactive systems characterized by security and/or privacy features two kinds of users must be taken into account, i.e., the attacker and the victim [120][22]. Williams and Li defined mental models of attackers and victims [133]. According to the resulting models, attacker features are: goal (e.g., obtain as many credit card details as possible), strategies (deception, diversion, and exploitation of lack of user knowledge), design of the phishing website (e.g., the website is visually similar to the genuine one), and limitations (fake websites might not fool all the victims). On the other hand, victim features are: perceptive and attentive processes (expert users paid attention to high-security user interface objects like the SSL padlock icon in the address bar, while novice users focused their attention to attractive, but less important, logos), knowledge and learning (e.g., the correct interpretation of the website content depends on what users know and have learned about their security values), memory (users that regularly applying their security knowledge are more likely to be able to recall critical security knowledge), external factors (time or social pressure), and risk-based decision making (e.g., the users evaluate if the overall risk of providing their information is outweighed by the benefits of doing this). These two models have been implemented into a prototype used to simulate the user modeled with respect to phishing websites.

Other works focus on different elements to improve user models. For example, Noureddine et al. try to address victims' uncertainties to build a more effective model [87]. From the attackers' perspective, works such as Nurse [89], Abroshan et al. [1], or Benias and Markopoulos [17] analyzed the motivations behind cybercriminals and how they exploit human factors and psychological aspects to succeed in their attacks. Attacks require careful planning and a convincing approach. If attackers simply email users without a conscious strategy about the attack (e.g., deceiving a user or manipulating their experience to redirect them to a fraudulent website), they are more likely to arouse suspicion. It is evident that, to be successful, the attackers spend a lot of time and effort.

Topic 5 - Defensive Mechanisms. This topic emerged as the most important in the entire corpus, characterizing 20 out of 52 papers. This underlines the importance of defensive mechanisms based on human aspects besides the ones based on technological features. One of the most explored solutions consists of training and educating users to prevent and detect phishing frauds. For example, *PhishGuru* is an online game that educates users on this type of attack while they are using email [112]. It teaches users to avoid falling for phishing attacks by delivering a training message when the user clicks on the URL in a simulated phishing email. The authors found that participants were less likely to click on a link in a (simulated) phishing email. In addition, they concluded that participants aged 18-25 years were consistently more vulnerable to phishing attacks than older participants, suggesting that age may play a role in the attention devoted to phishing emails. Thus, they propose to train young people in a way that specifically targets high school and college students. Besides PhishGuru, Sheng et al. proposed *Anti-Phishing Phil*, a system that trains users to identify phishing URLs through a gamified approach [111]. The authors found that the participants who played Anti-Phishing Phil could better identify fraudulent websites than other participants, thanks to the training messages and the ways they are presented.

Another game-based training on phishing attacks has been reported by Wen et al. [130]. The game, *"What.Hack"*, aims to train users on phishing by involving them in a role-playing game that simulates phishing from which the players have to defend. The user study demonstrated that a training resource exploiting a role-playing game is more effective and engaging than traditional forms of training.

A different approach for training users consists of integrating training messages inside warning messages for phishing attacks [137]. The user study demonstrated that embedded training messages have positive short- and long-term effects on user training and that this approach is effective to assist users in identifying phishing websites.

In an alternative approach, Wash and Cooper [129] compared traditional facts-and-advice training against training that uses a simple story to convey the same lessons. They found that facts-and-advice training works better than not training users, but only when presented by a security expert. On the other hand, stories did not work as well as facts-and-advice, but work better when told by a peer rather than a security expert. To this end, their results suggested that the source of training materials in conjunction with the type of materials can impact the security outcomes.

The need for more custom defensive mechanisms was studied by Oliveira et al. [91]. The authors analyzed phishing susceptibility according to user age. To this aim, a user study was conducted involving 158 participants in their homes without informing them about the receipt of emails to ensure natural behaviour. The study results showed that women are more vulnerable, thus a good defense strategy should take gender into account; it also remarked the need for training and educational tools for older users. Finally, it is worth mentioning a phishing toolkit that automatically constructs phishing websites of unlimited levels; these websites provide a good means to validate the effectiveness of defensive mechanisms against phishing attacks [145].

4.2 What solutions exist that address human factors to reduce the success of phishing attacks? (RQ2)

The retrieved 52 papers propose different solutions that, based on human factors, try to reduce the success of phishing scams. As reported in Table 4, we have organized them into four main categories, which are discussed in this section. Notice that publications [133] and [137] fall into two categories.

1. User Interface. This category includes papers that propose to modify the user interface (UI) to elicit certain behaviors or to inform users of their dangerous actions. Examples include the change of the icon visibility, the addition of interface messages or certain UI elements, or making users more attentive to the consequences of their actions via on-screen messages. For example, Williams and Li [133] studied the cognitive processes involved in assessing the

Table 4. Solutions based on human factors to reduce the success of phishing scam: four categories

Category	Included papers	References
User Interface	Papers that focus on modifying elements of the user interface to improve user security.	[7] [13] [23] [28] [42] [41] [73] [91] [117] [133] [137]
Attitude, Behavior, Psychological aspects	Papers that focus on users' attitudes, behaviors, characteristics or other psychological aspects to improve user security.	[1] [4] [6] [11] [17] [31] [52] [62] [64] [66] [80] [81] [87] [92] [97] [105] [109] [118] [128] [131]
Knowledge, Education, and Training	Papers that focus on educating users or providing training material (games of other applications) to improve their knowledge about security-related issues.	[12] [24] [36] [60] [68] [69] [76] [79] [85] [108] [112] [111] [129] [130] [132] [137]
Framework, Models, and Taxonomies	Papers that focus on frameworks / models / taxonomies that support designers in order to improve user security.	[16] [50] [83] [89] [119] [133] [145]

validity of web pages depending on the presence of the HTTPS padlock icon in the browser navigation bar, which usually indicates a secure website. Their study showed that often users fail to understand the role of the HTTPS padlock, thus becoming victims of phishing websites. The authors suggest extending interfaces by incorporating additional security indicators.

Xiong et al. propose two training-embedded warning interfaces for phishing attacks [137]. Such interfaces were evaluated to understand if they help users to identify phishing websites, also comparing them with the warning interface of Google Chrome that does not provide any train. The experimental results indicate that embedded training should be included in warning messages to prevent phishing attacks and to enable security training at scale. Zhao et al. [145] studied how phishing websites are designed to appear legitimate. They implemented a toolkit that permits to easily generate web pages that provide examples of phishing attacks, addressing both traditional and Web Single Sign-On phishing. An evaluation of the toolkit has shown that it is effective for researchers interested in studying user behavior when exposed to phishing attacks.

2. Attitude, Behavior, Psychological aspects. This category includes papers that focused on changing behavior, attitudes, or other psychological aspects (e.g. cognition) of users, both in a personal and professional context. Choong and Theofanos investigated strengths and weaknesses of security policies of a company by looking at how its employees perceived these policies [31]. They found that, even if such policies should contribute to increase the overall security, they often neglect human factors and thus they may not yield the desired results. For example, policies often require the creation of passwords that must be very long and complex, resulting difficult to remember. This problem is amplified also by policies that often require changing passwords. Users compensate these difficulties by behaving in wrong ways, for example by taking notes of their passwords on a post-it attached to their PC monitor.

In their study, Lévesque et al. examined the interactions among users, antivirus software, and malware as they occur on systems [73]. The results indicate that users who declared a high level of computer expertise were more susceptible to phishing attacks. Instead, they found that age and gender do not significantly correlate with successful malware attacks, which appears to be partially in contrast with the findings in [88].

Avery et al. [16] analyzed physiological aspects that push users to be deceived during attacks like phishing, scamming, watering hole, clickbait and repackaging. The identified philological components in a phishing attack are: instant gratification (e.g., ease opportunity to save or earn money), victim identification (e.g., desire to assist those in need),

reputation (e.g., the attack relies on the reputation of a company to deceive recipients), and undesirable consequences (e.g., victims are warned that their account will be destroyed if they do not do what asked).

Corradini and Nardelli [36] claimed that employees are the first line of defense from cyber attacks. They suggest that analyzing the users' perception of risks is the first step to tailor educational programs aimed at changing the attitude of users towards cyber attacks. The already mentioned questionnaire developed by Alissa et al. [4] is intended to measures human behavior towards cybersecurity policies and identify wrong behaviors; questionnaire results should be used to correct behaviors that can make users more vulnerable.

3. Knowledge, Education and Training. This category includes papers that focus on informing users about various cybersecurity-related topics to improve their cybersecurity knowledge. In general, these strategies focus on notifying the user as events happen, [137] or via training programs [36] [13], or simulations [130] to improve the user knowledge about cybersecurity issues. The aim is to minimize the likelihood that a cyber attack occurs and, if it does, users are aware of how to handle them (e.g. alert security, disconnect their devices from a network, etc.). More details on defensive mechanisms have been already reported in Section 4.1 - Topic 5.

Marforio et al. [79], Ndibwile et al. [85], Ardi et al. [12], and Canova et al. [24] developed applications that provide various methods such as creating fake login accounts to access suspicious sites, browser plug-ins to help with detection of suspicious sites, and the deployment of personalized security indicators to decrease phishing attacks on mobile applications, to educate, train, or assist users in a cyber/phishing attack. Indeed, the interaction with mobile devices poses different challenges since user interfaces are considerably different than the ones of PC, and user's focus is limited to a few visual elements. Maforio et al. [79] report the findings from a user study on the effectiveness of personalized security indicators for mobile applications. They propose to customize a mobile app (e.g., the one for e-banking) with a personal picture. This personal element is intended to help users recognize attacks against the customized app: if a malicious component infects the mobile device and overrides a mobile app (e.g., the one for e-banking cited before) to steal personal information, the user is facilitated to recognize the attack since the custom image is missing in the fake version of the app. The preliminary results revealed that personalized indicators could help users detecting applications that perform phishing attacks.

Other authors, e.g., Kumaraguru et al. [69], Ndibwile et al.[85], Wen at al. [130], and Sheng et al. [112], propose interactive game-based solutions to train users about various security-related issues. These articles suggest approaches more effective than traditional training methods that are based, e.g., on the use of videos providing demonstration about possible attacks that users watch passively. Instead, games and other methods allow users to interact in real-time with the training material. In this way, immediate feedback on a user action is provided, albeit good or bad, and consequently information is given on how users may improve their behavior. Furthermore, and perhaps more importantly, such user interactions can be recorded in an attempt to provide a statistical/analytical overview to a security analyst, thus giving insights about aspects of security practices that need improvements. For example, *UnPhishMe* is a mobile application that takes advantage of particular weaknesses of phishing sites such as the input information field, which required authentication [85]. As a result, UnPhishMe allows users to create fake login account credentials that mimicked the user login procedure. The application then determined if the login page changed to another web page after an authentication attempt by monitoring any changes in the hash code. Ultimately, the authors determined that the application is a 96% effective way to assist users in identifying phishing attacks.

Another approach to improve user knowledge and awareness of phishing attacks is based on the use of browser plugins. Ardi and Heidemann [12] developed *AuntieTuna*, a plugin for Google Chrome that presents alerts to users when they open a phishing web page. This extension checks if each visited page is a potential phishing website based

on snapshots of known good websites that a user adds to a whitelist. For example, a user adds a snapshot of Bank-X. For each webpage visited by the user the extension checks if the webpage looks like Bank-X and, if so, it is marked as phish and blocked. The authors demonstrated the ability of the plugin to reduce the likelihood that a user would enter their personal information into a fraudulent site and compromise their (or others) data.

Lastly, Lim et al. [76] propose a security training system that prepares users for email phishing and phone SMSishing attacks. The proposed training consisted of sending e-mail or SMS with an URL of a virtual phishing site, in the same way as the real phishing attacks, and alerts users when the webpage is open explaining that these kinds of messages and URLs might be dangerous. The evaluation of this system revealed that both the click rate of virtual phishing e-mails and threatening links decreased over time from 16% to 12%. They concluded that training against security threats in phishing e-mail for individual users would be possible through the proposed security training system.

4. Frameworks, Models and Taxonomies. This category includes papers that provide indications about how to approach human factors related issues of cybersecurity problems. For instance, behavioral or psychological elements are used to build frameworks, models or taxonomies (see, e.g., [81][52][16][87]). Gangire et al. defined a conceptual model of user behavior for advancing information security-compliant behavior in organizations [52]. This model identifies some factors influencing information security behavior in organizations, namely, perceived competence, perceived relatedness, and perceived autonomy. This model can be adopted, for example, by companies to analyze the employees' behavior while dealing with their security policies and, if necessary, to correct their wrong behavior with the final goal of reducing the effectiveness of phishing attacks against their organization employees.

Metalidou et al. [83] developed a preliminary framework that identifies correlation of human factors with the lack of information security awareness. The first dimension is the "Lack of Motivation", meaning that employees need to be motivated in order to follow secure behaviors and practices, as well as the management needs to identify what motivates their employees. The second one is "Lack of awareness", which refers to a lack of general knowledge about cyber attacks, for example, users do not know how important is to specify a strong password to prevent phishing attacks. The third dimension is "Users' Risky Belief", and an example is when users believe that the installation of anti-virus software is not crucial for their information. The fourth dimension is "Behavior", which indicates the users' risky behavior or missing prevention behavior. An example is the generation of weak passwords or the annotation of passwords on physical or digital sheets that can be easily intercepted by other persons. The fifth dimension is "Inadequate Use of Technology", which intends that even the best technology cannot succeed in safety issues without human cooperation. An example of inappropriate use of technology is the unauthorized reconfiguration of systems.

Nurse has developed an introductory taxonomy that summarizes cybercrimes, showing that they may continue to occur if not addressed appropriately [89]. This taxonomy focuses on five different types of cyber-criminal attacks that exploit human factors, including phishing. Nurse's contends that there should be an increase in identifying approaches to prevent, detect, and deter the behavior of attackers. In this way, synthesized insights from other fields can address issues related to cybercrime and attackers.

Steves et al. have developed the *Phish Scale*, an approach to grade phishing emails used to train users [119]. It is based on two dimensions: i) the number of cues contained in the email message, and ii) the premise alignment for the target audience. The considered cues are: errors contained in the message like spelling and grammar irregularities; technical indicator, as the type of the attachment; visual presentation indicator, like the logo imitation; language and content, for example, the sense of urgency; common tactic, like limited time offer. The premise alignment indicates how the email is relevant for the victim, rated as high, medium, and low. This supports phishing trainers to determine the difficulty of their phishing exercises and to explain the percentage of victims of a specific exercise.

It is evident that there is a need for more concise frameworks, models, and taxonomies that also share a common vocabulary. However, what the proposed frameworks, models, and taxonomies have all demonstrated is a more methodological approach to understand the human factors within the context of cybersecurity.

4.3 What are the main vulnerable human factors in phishing attacks? (RQ3)

The manual analysis of the papers revealed that sometimes human factors are not discussed explicitly as they are defined (e.g., in The Dirty Dozen). A human factor may not be referred to as such but rather as a cause of a problem; for example, *Lack of Knowledge* is not explicitly mentioned but it is often said that users do not know enough about phishing scams. Thus, we read the papers and identified the key areas that were addressed and why they were areas of concern from a human factor perspective. For instance, if one paper discusses issues around the availability of training material, we interpreted this as *Lack of Resources*. Similarly, if there were issues relating to deceptive strategies by attackers being successful we would consider this *Lack of Awareness* by the users. From the careful analysis of the 52 publications, the five human factors discussed in the rest of this section were those most referred to as vulnerable.

The first is *Lack of Knowledge*. Many papers focused on educating the user in some way or at least provide guidelines to improve the user knowledge about a cybersecurity-related issue or to improve their daily behavior/attitude towards them. In several cases, improving the user knowledge centered on providing relevant training to users. Other approaches provided users with information at the moment that it was relevant, e.g., indicating that a URL might not be legitimate.

The second human factor often addressed is *Lack of Resources*. In many cases, users were not educated on cybersecurity-related issues and how to avoid them both in a personal and business context. To address this issue, various approaches, frameworks, applications and tools to create more "cyber aware" users were developed. These solutions were quite different, each addressing either a specific issue or a general consensus of what users know (or do not know) about cybersecurity practice. Therefore, while there is an apparent *Lack of Resources* when it comes to resources to educate users about such issues, the approaches are very different, making it hard to establish a baseline for a direction to target future work.

The third human factor is *Lack of Awareness*, since it is often remarked that users do not pay enough attention to what they do or they do not notice changes in their interactions (e.g., a different website URL). In many cases, Lack of Awareness is related to *Lack of Knowledge* or even to *Lack of Resources* and *Complacency*; indeed, users could not have any support by tools or people to better inform them and, in some cases, even security experts fall victims of cyber attacks or cyber threats, possibly because of overconfidence that reduces their awareness during online interactions.

The fourth is *Norms*, since users are influenced by practices developed over time. In some publications (e.g., [36] [62]), issues surrounding workplace and environmental culture were highlighted as having an impact in several ways. Sasse et al. [108] discussed *social issues* and *informal work procedures* as causes for concern; for example, cases where people do not share their password among colleagues or lock their screen when they leave their desk can be seen as a sign of mistrust towards their colleagues rather than good cybersecurity practice. The tendencies for such practices are evident in the questionnaire developed by Alsharnouby et al.[7], where questions such as *"Passwords should not be shared with anyone"* or *"Securing workstations (screen lock or logout) prior to leaving area to prevent unauthorized access"* are also featured. Indeed, the adoption and practice of norm related behavior can also relate to other key human factors such as *Lack of Assertiveness*, where users do not want to feel ostracized from their colleagues and therefore adopt practices, which might be dangerous, because they want to maintain a positive social environment within the workplace. Consequently, a user's attitude towards the company and their colleagues may also be an influencing factor. Pham et al. discussed the impact of social influence on individual's behavior and beliefs on motivation and its normative

and informative effects on employees [97]. In a similar context, Mayer et al. studied 14 behavioral factors towards improving user internet security behavior [80]. Their analysis revealed that norms had a weak effect on users' behavior; suggesting that more research is needed to understand how norms might influence a user in other ways.

The fifth human factor often addressed is *Complacency*, since users are influenced by either what they already know or what is done within the workplace, which leads them to under-evaluate potential dangers. This was particularly evident in publications that discuss workplace cultures and behavior. In many ways, *Complacency* appears to be a result of users over-estimating their cybersecurity practices or assuming that they do not have anything worthwhile to make themselves a target. This could occur for many reasons such as *Lack of Knowledge*, workplace *Norms* [97] [80] [105], novice mistakes due to *Lack of Awareness* [118].

The above discussion highlights significant areas of concern related to vulnerable human factors in phishing; it emerges that the research community has to still devote a lot of effort in order to limit the success of phishing attacks. Section 6 will address some relevant future challenges.

5 SOME RECENT TRENDS

The reported SLR was performed during April-May 2019 and retrieved publications from 2001. In order to get a more complete and updated view, we recently looked at further works published from June 2019 to May 2021. Specifically, we analyzed publications that appeared in this two-year period in the proceedings of the conferences and in the journals that were considered in Phase 2 of our SLR. We found 14 publications, a number that confirms the trend highlighted in Figure 3, i.e., the interest in human factor in phishing attacks is growing, with a sharp increase from 2015. These recent publications fall in one of the first three categories presented in Table 4 (see Section 4.2). Some of them analyze specific elements of the phishing emails (e.g., suggested links, email content, etc.) in order to understand how they may fool users (similarly to what has been discussed in previous papers, e.g., [119]), some others report studies that compare behaviors or other characteristics of the users (e.g., age), while two of them analyze the prevalence effect, which has been also investigated in previous publications that focused on policies for privacy and security (see Section 4.1). Only 4 publications address some novel aspects: 3 of them focus on the analysis of persuasion principles that could be exploited in the phishing email content, the fourth one is about the use of AI algorithms to analyze users' behavior to possibly predict users' ability in recognizing phishing emails. These 14 recent publications are briefly discussed in this section.

Phishing attacks are becoming increasingly sophisticated, often attempting to trick users by manipulating the message content. One of the most deceptive elements is the suggested link (URL), which tries to make users believe that clicking on it will open the genuine site. URL confusion stems from a misalignment between user URL-parsing strategies and URL complexity (e.g., multiple sub-domains). Solutions to the problem can be either to educate users or redesign URL identity to reduce technical complexity, making URLs more user-friendly as suggested by Reynolds et al. in [103]. Users should accurately compare a URL with an expected destination in order to understand whether it could be a fraudulent URL. One approach to understand users' URL parsing strategies is presented by Albakry, Vaniea, and Wolter, who studied users' perception of 23 URLs with various structures to see if users could predict where the URL would lead and how safe they feel clicking on it [2]. They found that those who use technology regularly did not perform significantly better than those that do not. This is very concerning because it shows that regular use of technology does not provide adequate security experience or knowledge, and therefore we cannot assume that such regular users are any safer than others. This further strengthens the call for both more targeted training as well as tools to facilitate users to detect suspicious URLs. In [103], it is remarked that more education could help users be wary of

URL obfuscations. Thus, training confirms as an important defensive mechanism (see Section 4.1) and it is proposed again as one of the most significant solutions to reduce the success of phishing attacks (see Section 4.2).

In an attempt to understand the concerns surrounding URLs, Althobaiti, Meng, and Vaniea also contend that the average non-technical users do not have the knowledge to safely judge the validity of a URL in comparison to security experts [8]. Therefore, users are often less likely to question or even notice discrepancies in URL names, such as the use of identical-looking UTF8-encoded characters. However, attackers take advantage of other common human detection weaknesses such as misspellings (e.g., www.twitter.com vs. www.twittter.com) or typosquatting where attackers substitute characters like "vv" for "w" or capital "i" for lowercase "L" that look identical with most of the default fonts. One other tactic is redirecting or shorting URLs. However, redirected URLs are almost impossible to identify from the URL itself and even computers have difficulty identifying them without opening the URL's destination. To alleviate this problem, the authors propose a tool that assists users in judging the safety of URLs; specifically, the tool provides useful indications with reference to the specific URL the user is handling. The authors experimented different versions of the report providing these indications; they found that the longer version of the report allows users to accurately judge URLs (93% accurate) and that the summary version of the report still provides benefits (83% accurate).

The attackers devote even more attention to the email content, as it happens in spear-phishing attacks. In a study of 2020, Sharmal and Bashir discuss that scammers use languages specifically targeted to their victims, for instance to illicit fear of losing important data and/or to make the email appear trustworthy [110]. They identified several attack patterns as well as information examples that phishing emails often exploit, like traces of user's online activity (e.g., posts on social networks). The study by Pfeffel, Ulsamer, and Müller [96] tries to understand which elements of a phishing email mainly attract user attention; users mostly look at the subject and body of an email, whereas a fraudulent email address is more likely to trigger an alert. Thus, it might be worth studying ways to draw the users' focus to the email sender address, especially if it is not part of the usual users' addresses. It is also suggested that users should devote more time to process possible phishing scams, in order to analyze more carefully other lexical components. This reinforces the fact that people must be educated in recognizing phishing emails through proper training programs, and that they should be motivated in spending more time when reading and analyzing emails. A novel approach about training is presented in [138], where training information is provided in warning messages. This study provides two interesting results that are worth investigating in future research. The first one is that warning messages with embedded training do not overload the users and the message is still effective (even if it is longer because it provides further information). The second relevant result is about the long-term effects of the embedded training. Indeed, users were invited one week later to evaluate the legitimacy of further webpages without using warnings. It emerged that users previously exposed to the training-embedded warning messages differentiated legitimate and fraudulent webpages better than other users exposed to traditional warning messages. Thus training can be integrated inside warning messages without lowering the user performance in recognizing phishing websites.

From the analysis of the SLR, age emerged as having an impact on phishing susceptibility, even if contrasting results were found (see [73][88][112]). Two studies published in the last two years still provide contrasting results [77][106]. Lin et al. found that 43% of users (n=158) were susceptible to the simulated phishing emails, with older women showing the highest susceptibility [77]. In contrast, in [106] the authors found that older adults (age of 65+) and younger adults (age of 18+ and 65-) do not differ in classifying phishing emails. Older adults were only slower than younger, appearing to be more cautious in classifying emails. Anyway, accuracy was always lower than desirable (younger adults 66%, older adults 72%), suggesting that there is still much to do in educating people to identify phishing emails.

Two recent publications address the prevalence effect [107][113], which was discussed in articles retrieved in the SLR (see Section 4.1). In [107] the authors investigated if email load and phishing prevalence influence the identification of phishing emails. About email load, they found that the more emails users have in their inbox, the more difficult they perceive email classification, but this perception does not influence their ability to classify phishing emails. However, the authors argue that email load might become critical in case of multitasking. About the prevalence effect, it mainly emerged that low prevalence, i.e., low number of phishing emails in users' mailbox, results in the lower ability of users to detect phishing emails. These results are in line with previous findings, such as those in [109], which demonstrate decreased phishing sensitivity with fewer phishing emails. In [113], the prevalence effect has been investigated from another perspective, i.e., to understand whether the frequency of experiencing phishing emails during a training activity influences the identification of phishing attacks. Results revealed that participants who received more frequent phishing during the training had a higher hit rate during post-training activity, but also a higher false alarm rate at identifying phishing emails.

The novelty of 3 of the 14 publications of the last two years is that they study persuasion principles used in phishing emails to make the attack more successful [94][122][48]. These publications refer to the six persuasion principles presented by Cialdini in [34], which are not discussed here for the sake of brevity. The study in [94] includes a survey that exposed 985 participants to phishing emails implementing the six persuasion principles. The authors found that phishing emails based on *consistency* and *reciprocity* principles were the most successful, while those based on *scarcity* and *social proof* principles were the least successful. A similar study has been carried out focusing on company employees [122]. The study exposed 56000 participants of an international financial organization to five types of phishing emails that exploited the three persuasion principles that Cialdini, in his general study, indicates as the most effective, namely *social proof*, *authority* and *scarcity*. These three principles were used both in isolation and in combination. The results demonstrated that *social proof* is the most effective attack vector, followed by *authority* and *scarcity*. This study did not consider the other three principles, as in [94], and addressed a different type of population. It is worth noticing that *social proof* resulted in the most effective principle to persuade company employees in trusting the phishing emails they got; indeed, *social proof* refers to the fact that people want to be seen doing what other people, often peers, are doing. The studies in [122][94] indicate the need for further research along this interesting perspective determined by persuasion principles. A first attempt along this new direction comes from the article in [48], which proposes a list of Principles of Persuasion in Social Engineering that are specific for the cybersecurity context, since the Cialdini's principles are general-purpose. The study conducted a thematic analysis on a set of 194 phishing emails ranging from 2008 to 2017. The most prominent persuasion principles for phishing attacks that emerged are: *authority*, *reciprocation*, *integrity*, *strong affect*. While the first two principles are already defined by Cialdini [34], the last two emerged as new in this context. In particular, the *integrity* principle says that "people tend to believe that others usually express their true feelings and needs when they make a statement". The *strong affect*, instead, refers to the use of a "heightened emotional state to distract the person from performing a logical evaluation of the situation" (e.g., fear, excitement, or panic).

Practical implications of identifying new persuasion principles that might be effective in phishing attacks can range from the study of user behavior exposed to persuasion strategies in phishing emails to the development of tools for the automatic recognition of phishing emails. Such tools might exploit Artificial Intelligence (AI), e.g., algorithms based on sentiment analysis of the email content. AI is actually exploited in [142] to analyze interaction-based behavior and predict users' ability in recognizing phishing emails; the results show that slow mouse movements indicate high awareness of phishing emails and could be used to determine the likelihood of users falling victim of phishing attacks.

The current emphasis on AI will generate very likely more works on studying users' behaviors, in order to propose new solutions that might limit the success of phishing attacks.

6 FUTURE CHALLENGES

Our analysis demonstrates that the study of human factors in phishing attacks is still a quite novel research direction, since there was a significant growth of research papers since 2015. Despite the valuable research work, phishing remains very effective and the most frequent attack [58], primarily because attackers leverage on human factors to cheat users. There is a need to improve the efforts of the research community, and this section proposes some future directions that can help towards limiting the success of phishing attacks.

6.1 Specifying human factors in cybersecurity

Even if we referred to Duponts "Dirty Dozen" [44], identifying human factors in many of the publications required significant effort and interpretation, because human factors are not necessarily described as so or they are described in varying ways when referring to the same thing (e.g., *Lack of Awareness* and unaware users). Moreover, there may also be other human factors relevant to cybersecurity and phishing, which are not adequately referenced in the literature. The recent studies in [48, 106, 122], which investigate persuasion principles used in phishing emails, stimulate new ideas to investigate which human factors make people more susceptible to persuasion strategies used by attackers and to identify defensive solutions. Further efforts should be devoted towards agreeing on a common ground of relevant human factors within cybersecurity, so that researchers studying the implications of human factors can refer to it.

6.2 Anticipating phishing variants

Phishing attacks are evolving very quickly following and exploiting technological advances. However, the modus operandi typically consists of looking for defensive solutions when a new attack appears, thus exposing users to the new threats before solving them. Attacks, where possible, should be predicted and anticipated without chasing them, and thus reversing the classic defensive approach. This principle is adopted in war context and it is known as "The best defense is a good offense", it proposes to be proactive instead of passive, leading to a strategic advantage. This principle should be always adopted in cybersecurity, and in the context of phishing attacks it can help limit their effectiveness. For example, while until a couple of years ago phishing was an attack almost entirely launched by email, today it is also spreading to social channels and by instant messaging on apps like WhatsApp and Telegram. Knowing the nature and processes of phishing attacks and the new technology trends, researchers should define solutions to prevent new attacks. For instance, deepfake is a new frontier of artificial intelligence that permits the creation of synthetic media like audio or video, where a person in a video or his/her voice in an audio track is replaced with someone else. Deepfake is gaining momentum in fake news, hoaxes, and financial fraud [29]. With deepfake technology going viral in the next few years, it wouldn't be surprising if fake videos, images and audio will be sent to victims to request credentials and personal data, such as a video of a family member asking for the bank's credentials to perform a transaction. Similarly, the spread use of mobile devices must push the identification of ad-hoc defensive solutions. For example, company employees are used to perform part of their work by using smartphones or tablets, thus company policies should take into account this. But also the design of general-purpose defensive solutions, like warning messages or antivirus, should better consider the nature of mobile devices (e.g., reduced screen size, mobility, low user attention) to curb these attacks.

6.3 Need for design indications

This SLR revealed a lack of mature and widely-adopted indications for designing solutions that could limit phishing attacks. While in HCI there are well-established indications to improve system usability and UX, e.g. principles, standards, guidelines, design rules, heuristics [98], in the field of usable security, and in particular for phishing attacks, the maturity reached does not guarantee robust solutions in the real world. For example, warning messages implemented in modern browsers to alert users on phishing attacks might be significantly improved, as demonstrated in [45][102][55]; however, browser developers seem to neglect these proposals [40]. These messages often contain jargon that is too technical; consequently, non-technical users do not understand the risks related to open a phishing web page. In addition, such messages display the same warning interfaces, accustoming users to the same message and causing them, over time, to skip the message, thus accessing malicious sites.

Another area of phishing attacks that would benefit from more effective indications is password generation policies. Some works have shown the importance of having users generate robust and mnemonic passwords [139][70]. However, in real contexts where several passwords for daily work and personal activities have to be managed, users often tend to behave incorrectly, for example, they always enter the same passwords or write them on small sheets of paper exposed to the public (e.g. post-its attached to the monitor) [1]. Guidelines on password management exist that should help users to manage simple and robust passwords, but users neglect them, exposing themselves to possible threats to their security or that of the organizations they work for. In addition to a good understanding of why such behaviors, more effective guidelines should be defined to help users manage passwords in the real world.

6.4 More awareness on humans as the weakest link

This SLR focuses on human factors in phishing attacks since it has been widely demonstrated that the human is, in general, the weakest link in the cybersecurity chain, and this is especially true for phishing attacks [57]. This SLR highlighted important contributions on the study of human factors: on the behavior of an individual that interacts with phishing emails (e.g., [73]), on the relationship between organizational and human factors (e.g, [50]), on user awareness on phishing attacks (e.g., [131]), and on how human factors influence this awareness (e.g, [13]). However, there is still a lot to understand about human factors, how they are related to phishing attacks and how their study can help in defining more appropriate solutions. Multidisciplinary teams should focus more and more on the study of human factors, with objectives such as: to improve the definition of the mental models of attackers and victims [133]; to understand the victims' behaviors in real and daily contexts when interacting with potential fraudulent emails [73]; to investigate how culture, skills, gender, age, and other contextual factors might expose users to phishing attacks; to understand why policies to generate passwords fail in the real world [4].

6.5 Training, training and more training

This SLR has shown that one of the most adopted solutions to fight phishing is prevention through user training. Games like PhishGuru [68], What.Hack [130] or training systems like Anti-Phishing Phil [112] [111] have been discussed in this article. However, it remains to be studied to what extent these tools are adopted, since there is not much evidence about their acceptance and adoption. Furthermore, since training is a valid and promising solution against phishing attacks, research should concentrate on the most vulnerable types of users, for example children. In [71], an interesting approach to train children is presented. Vulnerable users are also "digital immigrants" [99], who often do not have a good IT education but use computers to access, for example, public administration or bank services, or just websites for

entertainment. More targeted studies could contribute to the definition of mechanisms and tools for ad-hoc training, possibly using metaphors, jargon, historical references, or analogies closer to specific age groups, cultures, skills. Because training requires time, it is often avoided. Less time-demanding approaches should be designed, for example integrating into warning messages for phishing attacks fragments of knowledge on how to recognize phishing sites besides alerting the users [10]. The results of the recent study in [138] indicate that providing training information in warning messages does not overload users and the messages are effective. The need for more effective and tailored training programs has been also underlined in a recent literature review on cybersecurity social engineering training programs, which found that, despite the existing training programs, hackers are still successful against trained users [3].

7 CONCLUSIONS

This article has provided a deep overview of phishing attacks and their relationship with human factors as causes as well as defense solutions to them. Despite phishing attacks are widely recognized among the most spread and effective attacks and human factors play a central role in them, the SLR revealed a lack of systematized knowledge on these topics. Existing surveys and SLR dealt with phishing attacks from more technical and technological perspectives. Instead, the SLR presented in this article, through the analysis of the reviewed papers, provides answers to different research questions, leading to: 1) identify the hottest topics investigated within the literature (policies, organization susceptibility, user awareness, use of various models, defensive mechanisms), 2) classify effective solutions that may reduce the success of phishing attacks (proper design of user interfaces, user training, systematization of human factors in frameworks or similar); 3) highlight the most vulnerable human factors exploited by attackers during phishing scams (Lack of Knowledge, Lack of Resources, Lack of Awareness, Norms, Complacency).

All these analyses supported the identification of future challenges that should drive the research in the next years, in order to get a more systematic view of human factors and phishing attacks within: the need for a common ground that specifies human factors in cybersecurity, the importance of anticipating phishing variants to identify in advance ways to defend against them, the need for design indications to develop systems robust against phishing attacks, a deeper understanding and awareness that humans are the weakest link, the importance of user training as one of the most promising solutions to defend against phishing.

This SLR paves the way to deeply understand phishing attacks beyond technical aspects, contributing to fight against them. An important aspect that this SLR highlighted is that a systematization of human factors within cybersecurity is still missing; this deserves special attention, in order to build a common basis in the study of causes, consequences and remedies of cyber attacks, from the users' point of view. It is worth remarking that today cybersecurity is different from the past, not only because of the growing number and variety of cyber attacks [58] but also because nowadays it involves national and supranational (e.g., EU) critical infrastructures. Therefore, to protect the cyberspace, which today represents one of the most important fields where modern wars are fought, human factors play an important role that researchers and practitioners have to consider when developing systems, algorithms, organizational processes and everything that involve computer systems and human security.

Acknowledgements. The work of Maria Francesca Costabile and Giuseppe Desolda has been supported by the Italian Ministry of University and Research (MUR) under grant PRIN 2017 "EMPATHY" and by the PON projects LIFT, TALIsMAn and SIMPLe. The work of Tiziana Catarci, Lauren S. Ferro and Andrea Marrella has been supported by the "Dipartimento di Eccellenza" grant, the H2020 projects DATACLOUD, DESTINI and FIRST, the Italian project RoMA - Resilience of Metropolitan Areas, and the Sapienza grant BPbots.

REFERENCES

[1] Hossein Abroshan, Jan Devos, Geert Poels, and Eric Laermans. 2018. Phishing Attacks Root Causes. In *Risks and Security of Internet and Systems*. Springer, 187–202. https://doi.org/10.1007/978-3-319-76687-4_13

[2] Sara Albakry, Kami Vaniea, and Maria K Wolters. 2020. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[3] Hussain Aldawood and Geoffrey Skinner. 2019. Reviewing Cyber Security Social Engineering Training and Awareness Programs–Pitfalls and Ongoing Issues. *Future Internet* 11, 3 (2019), 73. https://doi.org/10.3390/fi11030073

[4] Khalid Adnan Alissa, Hanan Abdullah Alshehri, Shahad Abdulaziz Dahdouh, Basstaa Mohammad Alsubaie, Afnan Mohammed Alghamdi, Abdulrahman Alharby, and Norah Ahmed Almubairik. 2018. An Instrument to Measure Human Behavior Toward Cyber Security Policies. In *21st Saudi Computer Society National Computer Conf. (NCC)*. IEEE, 1–6. https://doi.org/10.1109/ncg.2018.8592978

[5] A. Almomani, B. B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani. 2013. A Survey of Phishing Email Filtering Techniques. *IEEE Communications Surveys Tutorials* 15, 4 (2013), 2070–2090. https://doi.org/10.1109/SURV.2013.030713.00020

[6] Manal Alohali, Nathan Clarke, Steven Furnell, and Saad Albakri. 2017. Information security behavior: Recognizing the influencers. In *2017 Computing Conf.* IEEE, 844–853. https://doi.org/10.1109/sai.2017.8252194

[7] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. 2015. Why phishing still works: User strategies for combating phishing attacks. *Int. J. of Human-Computer Studies* 82 (2015), 69–82. https://doi.org/10.1016/j.ijhcs.2015.05.005

[8] Kholoud Althobaiti, Nicole Meng, and Kami Vaniea. 2021. I Don't Need an Expert! Making URL Phishing Features Human Comprehensible. In *The ACM CHI Conference on Human Factors in Computing Systems 2021*. Association for Computing Machinery (ACM).

[9] Edward G Amoroso. 2007. *Cyber Security*. Silicon Press.

[10] Joseph Aneke, Carmelo Ardito, and Giuseppe Desolda. 2020. Designing an Intelligent User Interface for Preventing Phishing Attacks. In *Beyond Interactions*. Springer, 97–106. https://doi.org/10.1007/978-3-030-46540-7_10

[11] N.A.G. Arachchilage, S. Love, and K. Beznosov. 2016. Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior* 60 (2016), 185–197. https://doi.org/10.1016/j.chb.2016.02.065

[12] Calvin Ardi and John Heidemann. 2016. AuntieTuna: Personalized Content-based Phishing Detection. In *Workshop on Usable Security (USEC'16)*. Internet Society. https://doi.org/10.14722/usec.2016.23012

[13] Ayman Asfoor, Fiza Abdul Rahim, and Salman Yussof. 2018. Factors Influencing Information Security Awareness of Phishing Attacks from Bank Customers' Perspective: A Preliminary Investigation. In *Adv. in Int. Syst. and Comp.* Springer, 641–654. https://doi.org/10.1007/978-3-319-99007-1_60

[14] IEA International Ergonomics Association. 2021. Definition, Domains of Specialization, Systemic Approach. https://iea.cc/definition-and-domains-of-ergonomics/. Accessed: 2021-26-04.

[15] AtlasVPN. 2021. A record 2 million phishing sites reported in 2020, highest in a decade. https://atlasvpn.com/blog/a-record-2-million-phishing-sites-reported-in-2020-highest-in-a-decade, note = Accessed: 2021-06-10.

[16] J. Avery, M. Almeshekah, and E. Spafford. 2017. Offensive deception in computing. In *12th Int. Conf. on Cyber Warfare and Security, ICCWS'17*. Academic Conf. Int. Limited, 23–31.

[17] Nikos Benias and Angelos P. Markopoulos. 2018. Hacking the human: Exploiting primordial instincts. In *2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conf.* IEEE, 1–6. https://doi.org/10.23919/seeda-cecnsm.2018.8544934

[18] Jonathan M. Bischof and Edoardo M. Airoldi. 2012. Summarizing Topical Content with Word Frequency and Exclusivity. In *29th Int. Conf. on Int. Conf. on Machine Learning (ICML'12)*. Omnipress, Madison, WI, USA, 9–16.

[19] David M. Blei and John D. Lafferty. 2009. Visualizing Topics with Multi-Word Expressions. arXiv:0907.1013 [stat.ML]

[20] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022.

[21] John C. 2020. Google Registers a 350% Increase in Phishing Websites Amid Quarantine. https://atlasvpn.com/blog/google-registers-a-350-increase-in-phishing-websites-amid-quarantine Accessed: 2021-06-10.

[22] L Jean Camp. 2009. Mental models of privacy and security. *IEEE Tech. and Soc. Magazine* 28, 3 (2009), 37–46. https://doi.org/10.1109/MTS.2009.934142

[23] Casey Inez Canfield, Baruch Fischhoff, and Alex Davis. 2016. Quantifying Phishing Susceptibility for Detection and Behavior Decisions. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58, 8 (2016), 1158–1172. https://doi.org/10.1177/0018720816665025

[24] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Benjamin Reinheimer. 2015. NoPhish App Evaluation: Lab and Retention Study. In *Workshop on Usable Security (USEC'15)*. Internet Society. https://doi.org/10.14722/usec.2015.23009

[25] Allison June-Barlow Chaney and David M Blei. 2012. Visualizing topic models. In *Sixth Int. AAAI Conf. on Weblogs and Social Media*. AAAI press. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewPaper/4645

[26] Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L. Boyd-graber, and David M. Blei. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. In *Advances in Neural Information Processing Systems 22*. Curran Associates, Inc., 288–296. http://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.pdf

[27] Bi Chen, Leilei Zhu, Daniel Kifer, and Dongwon Lee. 2010. What Is an Opinion About? Exploring Political Standpoints Using Opinion Scoring Model. In *Twenty-Fourth AAAI Conf. on Artificial Intelligence*. AAAI press. http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1863

[28] Jing Chen, Scott Mishler, Bin Hu, Ninghui Li, and Robert W. Proctor. 2018. The description–experience gap in the effect of warning reliability on user trust and performance in a phishing-detection context. *Int. J. of Human-Computer Studies* 119 (2018), 35–47. https://doi.org/10.1016/j.ijhcs.2018.05.010

[29] Bobby Chesney and Danielle Citron. 2019. Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review* 107 (2019). https://doi.org/10.2139/ssrn.3213954

[30] Kang Leng Chiew, Kelvin Sheng Chek Yong, and Choon Lin Tan. 2018. A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications* 106 (2018), 1 – 20. https://doi.org/10.1016/j.eswa.2018.03.050

[31] Yee-Yin Choong and Mary Theofanos. 2015. What 4,500+ People Can Tell You – Employees' Attitudes Toward Organizational Password Policy Do Matter. In *Int. Conf. on Human Aspects of Information Security, Privacy, and Trust*. Springer, 299–310. https://doi.org/10.1007/978-3-319-20376-8_27

[32] Jason Chuang, Yuening Hu, Ashley Jin, John D Wilkerson, Daniel A McFarland, Christopher D Manning, and Jeffrey Heer. 2013. Document exploration with topic modeling: Designing interactive visualizations to support effective analysis workflows. In *NIPS Workshop on Topic Models: Computation, Application, and Evaluation*. https://mimno.infosci.cornell.edu/nips2013ws/nips2013tm_submission_17.pdf

[33] Jason Chuang, Christopher D. Manning, and Jeffrey Heer. 2012. Termite: Visualization Techniques for Assessing Textual Topic Models. In *Int. Conf. on Advanced Visual Interfaces (AVI '12)*. Association for Computing Machinery, 74–77. https://doi.org/10.1145/2254556.2254572

[34] Robert B Cialdini. 2009. *Influence: Science and practice*. Vol. 4. Pearson education Boston, MA.

[35] Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on Twitter: It's not easy!. In *Seventh Int. AAAI Conf. on Weblogs and Social Media*. AAAI press. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6128/6347

[36] Isabella Corradini and Enrico Nardelli. 2018. Building Organizational Risk Culture in Cyber Security: The Role of Human Factors. In *Adv. in Int. Syst. and Comp*. Springer, 193–202. https://doi.org/10.1007/978-3-319-94782-2_19

[37] Lorrie Faith Cranor. 2008. A Framework for Reasoning about the Human in the Loop. In *1st Conf. on Usability, Psychology, and Security (UPSEC'08)*. USENIX Association. https://dl.acm.org/doi/10.5555/1387649.1387650

[38] L. F. Cranor and S. Garfinkel. 2004. Editors' Introduction: Secure or Usable? *IEEE Security Privacy* 2, 5 (2004). https://doi.org/10.1109/MSP.2004.69

[39] Marco Cristani, Alessandro Perina, Umberto Castellani, and Vittorio Murino. 2008. Geo-located image analysis using latent representations. In *IEEE Conf. on Computer Vision and Pattern Recognition*. 1–8. https://doi.org/10.1109/CVPR.2008.4587390

[40] Giuseppe Desolda, Francesco Di Nocera, Lauren Ferro, Rosa Lanzilotti, Piero Maggi, and Andrea Marrella. 2019. Alerting Users About Phishing Attacks. In *21st Int. Conf. on Human-Computer Interaction*. Springer, 134–148. https://doi.org/10.1007/978-3-030-22351-9_9

[41] Julie S. Downs, Mandy Holbrook, and Lorrie Faith Cranor. 2007. Behavioral response to phishing risk. In *Proc. of the anti-phishing working groups 2nd annual eCrime researchers summit*. ACM, 37–44. https://doi.org/10.1145/1299015.1299019

[42] Julie S. Downs, Mandy B. Holbrook, and Lorrie Faith Cranor. 2006. Decision strategies and susceptibility to phishing. In *2nd Symp. on Usable privacy and security - SOUPS '06*. ACM, 79–90. https://doi.org/10.1145/1143120.1143131

[43] Susan T Dumais. 2004. Latent semantic analysis. *Annual review of information science and technology* 38, 1 (2004), 188–230.

[44] Gordon Dupont. 1997. The Dirty Dozen Errors in Maintenance. In *11th Meeting on Human Factors in Aviation Maintenance and Inspection*.

[45] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *26th Int. Conf. on Human Factors in Computing Systems (CHI '08)*. ACM. https://doi.org/10.1145/1357054.1357219

[46] Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing. 2010. A Latent Variable Model for Geographic Lexical Variation. In *2010 Conf. on Empirical Methods in Natural Language Processing (EMNLP '10)*. Association for Computational Linguistics, 1277–1287. https://dl.acm.org/doi/10.5555/1870658.1870782

[47] Jessica Ellis. 2020. *COVID-19 Phishing Update: Campaigns Exploiting Hope for a Cure*. https://info.phishlabs.com/blog/covid-phishing-update-campaigns-addressing-a-cure

[48] Ana Ferreira and Soraia Teles. 2019. Persuasion: How phishing emails can influence users and bypass security measures. *International Journal of Human-Computer Studies* 125 (2019), 19–31.

[49] Arlene Fink. 2019. *Conducting research literature reviews: From the internet to paper*. Sage publications.

[50] Waldo Rocha Flores and Mathias Ekstedt. 2012. A model for investigating organizational impact on information security behavior. In *Pre-ICIS Workshop on Information Security and Privacy (SIGSEC)*.

[51] B. Fuglede and F. Topsoe. 2004. Jensen-Shannon divergence and Hilbert space embedding. In *Int. Symp. on Information Theory (ISIT '04)*. https://doi.org/10.1109/ISIT.2004.1365067

[52] Yotamu Gangire, Adele Da Veiga, and Marlien Herselman. 2019. A conceptual model of information security compliant behaviour based on the self-determination theory. In *2019 Conf. on Information Communications Technology and Society (ICTAS'19)*. IEEE, 1–6. https://doi.org/10.1109/ictas.2019.8703629

[53] Matthew J Gardner, Joshua Lutes, Jeff Lund, Josh Hansen, Dan Walker, Eric Ringger, and Kevin Seppi. 2010. The topic browser: An interactive tool for browsing topic models. In *NIPS workshop on challenges of data visualization*, Vol. 2. Whistler Canada.

[54] Henry W Glaspie and Waldemar Karwowski. 2017. Human Factors in Information Security Culture: A Literature Review. In *Int. Conf. on Applied Human Factors and Ergonomics*. Springer, 269–280. https://doi.org/10.1007/978-3-319-60585-2_25

[55] Sanjay Goel, Kevin Williams, and Ersin Dincelli. 2017. Got Phished? Internet Security and Human Vulnerability. *Journal of the Association for Information Systems* 18, 1 (2017). https://doi.org/10.17705/1jais.00447

[56] Derek Greene and James P Cross. 2015. Unveiling the Political Agenda of the European Parliament Plenary: A Topical Analysis. In *ACM Web Science Conf. (WebSci '15)*. 1–10. https://doi.org/10.1145/2786451.2786464

[57] IBM. 2014. IBM X-Force Threat Intelligence Index (2014). https://i.crn.com/sites/default/files/ckfinderimages/userfiles/images/crn/custom/IBMSecurityServices2014.PDF. Accessed: 2021-06-10.

[58] IBM. 2020. IBM X-Force Threat Intelligence Index (2020). https://www.ibm.com/security/digital-assets/xforce-threat-intelligence-index-map. Accessed: 2021-06-10.

[59] ITU. 2009. *Overview of Cybersecurity. Recommendation ITU-T X.1205.* http://www.itu.int/rec/T-REC-X.1205-200804-I/en

[60] K. Jansson and R. von Solms. 2013. Phishing for phishing awareness. *Behaviour & Information Technology* 32, 6 (2013), 584–593. https://doi.org/10.1080/0144929x.2011.632650

[61] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. *Multimedia Tools Appl.* 78, 11 (2019), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

[62] Matthew Jensen, Alexandra Durcikova, and Ryan Wright. 2017. Combating Phishing Attacks: A Knowledge Management Approach. In *50th Hawaii Int. Conf. on System Sciences (2017)*. ScholarSpace. https://doi.org/10.24251/hicss.2017.520

[63] Zaixing Jiang, Xuezhong Zhou, Xiaoping Zhang, and Shibo Chen. 2012. Using link topic model to analyze traditional Chinese Medicine Clinical symptom-herb regularities. In *14th Int. Conf. on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 15–18. https://doi.org/10.1109/HealthCom.2012.6380057

[64] A. Karakasiliotis, S. M. Furnell, and M. Papadaki. 2006. Assessing end-user awareness of social engineering and phishing. In *7th Australian Information Warfare and Security Conf.* Security Research Institute (SRI). https://doi.org/10.4225/75/57A80E47AA0CB

[65] M. Khonji, Y. Iraqi, and A. Jones. 2013. Phishing Detection: A Literature Survey. *IEEE Communications Surveys Tutorials* 15, 4 (2013), 2091–2121. https://doi.org/10.1109/SURV.2013.032213.00009

[66] Iacovos Kirlappos and Martina Angela Sasse. 2015. Fixing Security Together: Leveraging trust relationships to improve security in organizations. In *Workshop on Usable Security (USEC'15)*. Internet Society. https://doi.org/10.14722/usec.2015.23013

[67] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University* 33, 2004 (2004), 1–26.

[68] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. 2009. School of phish: a real-world evaluation of anti-phishing training. In *5th Symp. on Usable Privacy and Security (SOUPS '09)*. ACM. https://doi.org/10.1145/1572532.1572536

[69] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. 2010. Teaching Johnny not to fall for phish. *ACM Trans. on Internet Technology* 10, 2 (2010), 1–31. https://doi.org/10.1145/1754393.1754396

[70] Cynthia Kuo, Sasha Romanosky, and Lorrie Faith Cranor. 2006. Human selection of mnemonic phrase-based passwords. In *2nd Symp. on Usable privacy and security - SOUPS '06*. ACM, 67–78. https://doi.org/10.1145/1143120.1143129

[71] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. 2017. How effective is anti-phishing training for children?. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*. 229–239.

[72] Elmer EH Lastdrager. 2014. Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science* 3, 1 (2014). https://doi.org/10.1186/s40163-014-0009-y

[73] Fanny Lalonde Lévesque, Sonia Chiasson, Anil Somayaji, and José M. Fernandez. 2018. Technological and human factors of malware attacks: A computer security clinical trial approach. *ACM Trans. on Privacy and Security* 21, 4 (2018), 1–30. https://doi.org/10.1145/3210311

[74] James A Lewis. 2006. Cybersecurity and Critical Infrastructure Protection. *Center for Strategic and Int. Studies* (2006).

[75] Divakaran Liginlal, Inkook Sim, and Lara Khansa. 2009. How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management. *Computers & Security* 28, 3-4 (2009), 215–228. https://doi.org/10.1016/j.cose.2008.11.003

[76] Ilkwon Lim, Young-Gil Park, and Jae-Kwang Lee. 2016. Design of Security Training System for Individual Users. *Wireless Personal Communications* 90, 3 (2016), 1105–1120. https://doi.org/10.1007/s11277-016-3380-z

[77] Tian Lin, Daniel E Capecci, Donovan M Ellis, Harold A Rocha, Sandeep Dommaraju, Daniela S Oliveira, and Natalie C Ebner. 2019. Susceptibility to spear-phishing emails: Effects of internet user demographics and email content. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 5 (2019), 1–28.

[78] Cindy Lustig and Warren Meck. 2009. The Overflowing Brain: Information Overload and the Limits of Working Memory. *The New England Journal of Medicine* 360, 14 (2009).

[79] Claudio Marforio, Ramya Jayaram Masti, Claudio Soriente, Kari Kostiainen, and Srdjan Čapkun. 2016. Evaluation of Personalized Security Indicators as an Anti-Phishing Mechanism for Smartphone Applications. In *2016 CHI Conf. on Human Factors in Computing Systems (CHI '16)*. ACM, 540–551. https://doi.org/10.1145/2858036.2858085

[80] Peter Mayer, Alexandra Kunz, and Melanie Volkamer. 2017. Reliable Behavioural Factors in the Information Security Context. In *12th Int. Conf. on Availability, Reliability and Security*. ACM. https://doi.org/10.1145/3098954.3098986

[81] Steven McElwee, George Murphy, and Paul Shelton. 2018. Influencing Outcomes and Behaviors in Simulated Phishing Exercises. In *SoutheastCon 2018*. IEEE, 1–6. https://doi.org/10.1109/secon.2018.8479109

[82] Merriam-Webster. 2018. Definition of Cybersecurity in English by Merriam-Webster. https://www.merriam-webster.com/dictionary/cybersecurity Accessed: 2021-06-10.

[83] Efthymia Metalidou, Catherine Marinagi, Panagiotis Trivellas, Niclas Eberhagen, Christos Skourlas, and Georgios Giannakopoulos. 2014. The Human Factor of Information Security: Unintentional Damage Perspective. *Procedia - Social and Behavioral Sciences* 147 (2014), 424–428. https://doi.org/10.1016/j.sbspro.2014.07.133

[84] Elizabeth Montalbano. 2020. *Top Email Protections Fail in Latest COVID-19 Phishing Campaign.* https://threatpost.com/top-email-protections-fail-covid-19-phishing/154329/

[85] Jema David Ndibwile, Youki Kadobayashi, and Doudou Fall. 2017. UnPhishMe: Phishing Attack Detection by Deceptive Login Simulation through an Android Mobile App. In *12th Asia Joint Conf. on Information Security (AsiaJCIS)*. IEEE, 38–47. https://doi.org/10.1109/asiajcis.2017.19

[86] David Newman, Youn Noh, Edmund Talley, Sarvnaz Karimi, and Timothy Baldwin. 2010. Evaluating Topic Models for Digital Libraries. In *10th Annual Joint Conf. on Digital Libraries (JCDL '10)*. ACM, 215–224. https://doi.org/10.1145/1816123.1816156

[87] Mohammad A. Noureddine, Andrew Marturano, Ken Keefe, Masooda Bashir, and William H. Sanders. 2017. Accounting for the Human User in Predictive Security Models. In *22nd Pacific Rim Int. Symp. on Dependable Computing (PRDC '17)*. IEEE, 329–338. https://doi.org/10.1109/prdc.2017.58

[88] Jude Jacob Nsiempba, Fanny Lalonde Lévesque, Nathalie de Marcellis-Warin, and José M. Fernandez. 2018. An Empirical Analysis of Risk Aversion in Malware Infections. In *Risks and Security of Internet and Systems (CRiSIS'17)*. Springer, 260–267. https://doi.org/10.1007/978-3-319-76687-4_18

[89] Jason RC Nurse. 2018. Cybercrime and You: How Criminals Attack and the Human Factors That They Seek to Exploit. *The Oxford Handbook of Cyberpsychology* (2018). https://doi.org/10.1093/oxfordhb/9780198812746.013.35

[90] Chitu Okoli and Kira Schabram. 2010. A guide to conducting a systematic literature review of information systems research. (2010).

[91] Daniela Oliveira, Natalie Ebner, Harold Rocha, Huizi Yang, Donovan Ellis, Sandeep Dommaraju, Melis Muradoglu, Devon Weir, Adam Soliman, and Tian Lin. 2017. Dissecting Spear Phishing Emails for Older vs Young Adults: On the Interplay of Weapons of Influence and Life Domains in Predicting Susceptibility to Phishing. In *2017 CHI Conf. on Human Factors in Computing Systems (CHI '17)*. ACM, 6412–6424. https://doi.org/10.1145/3025453.3025831

[92] Kaan Onarlioglu, Utku Ozan Yilmaz, Engin Kirda, and Davide Balzarotti. 2012. Insights into User Behavior in Dealing with Internet Attacks. In *NDSS Symp.* https://www.ndss-Symp..org/wp-content/uploads/2017/09/P08_1.pdf

[93] Oxford-Press. 2018. Definition of Cybersecurity in English by Oxford Dictionaries. https://en.oxforddictionaries.com/definition/Cybersecurity Accessed: 2021-06-10.

[94] Kathryn Parsons, Marcus Butavicius, Paul Delfabbro, and Meredith Lillie. 2019. Predicting susceptibility to social influence in phishing emails. *International Journal of Human-Computer Studies* 128 (2019), 17–26.

[95] Michael J Paul and Mark Dredze. 2011. You Are What You Tweet: Analyzing Twitter for Public Health. In *Fifth Int. AAAI Conf. on Weblogs and Social Media*. AAAI press. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2880/3264

[96] Kevin Pfeffel, Philipp Ulsamer, and Nicholas H Müller. 2019. Where the user does look when reading phishing mails–an eye-tracking study. In *International Conference on Human-Computer Interaction*. Springer, 277–287.

[97] Hiep Cong Pham, Duy Dang Pham, Linda Brennan, and Joan Richardson. 2017. Information Security and People: A Conundrum for Compliance. *Australasian Journal of Information Systems* 21 (2017). https://doi.org/10.3127/ajis.v21i0.1321

[98] Jennifer Preece, Yvonne Rogers, and Helen Sharp. 2019. *Interaction Design: Beyond Human-Computer Interaction, 5th Edition*. Wiley.

[99] Marc Prensky. 2001. Digital Natives, Digital Immigrants. *On the Horizon* 9, 5 (2001).

[100] Daniel Ramage, Evan Rosen, Jason Chuang, Christopher D Manning, and Daniel A McFarland. 2009. Topic modeling for the social sciences. In *NIPS workshop on applications for topic models: text and beyond*, Vol. 5. 27.

[101] Justus Randolph. 2009. A guide to writing the dissertation literature review. *Practical Assessment, Research, and Evaluation* 14, 1 (2009).

[102] Robert W Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. 2018. An experience sampling study of user reactions to browser warnings in the field. In *2018 CHI Conf. on Human Factors in Computing Systems (CHI '18)*. 1–13. https://doi.org/10.1145/3173574.3174086

[103] Joshua Reynolds, Deepak Kumar, Zane Ma, Rohan Subramanian, Meishan Wu, Martin Shelton, Joshua Mason, Emily Stark, and Michael Bailey. 2020. Measuring identity confusion with uniform resource locators. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[104] Rimvydas Rukšėnas, Paul Curzon, and Ann Blandford. 2008. Modelling and analysing cognitive causes of security breaches. *Innovations in Systems and Software Engineering* 4, 2 (2008), 143–160. https://doi.org/10.1007/s11334-008-0050-7

[105] Nader Sohrabi Safa, Mehdi Sookhak, Rossouw Von Solms, Steven Furnell, Norjihan Abdul Ghani, and Tutut Herawan. 2015. Information security conscious care behaviour formation in organizations. *Computers & Security* 53 (2015), 65–78. https://doi.org/10.1016/j.cose.2015.05.012

[106] Dawn M Sarno, Joanna E Lewis, Corey J Bohil, and Mark B Neider. 2020. Which phish is on the hook? Phishing vulnerability for older versus younger adults. *Human factors* 62, 5 (2020), 704–717.

[107] Dawn M Sarno and Mark B Neider. 2021. So Many Phish, So Little Time: Exploring Email Task Factors and Phishing Susceptibility. *Human Factors* (2021), 0018720821999174.

[108] M A Sasse, S Brostoff, and D Weirich. 2001. Transforming the 'Weakest Link'–a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal* 19, 3 (2001), 122–131. https://doi.org/10.1023/a:1011902718709

[109] Ben D. Sawyer and Peter A. Hancock. 2018. Hacking the Human: The Prevalence Paradox in Cybersecurity. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 60, 5 (2018), 597–609. https://doi.org/10.1177/0018720818780472

[110] Tanusree Sharma and Masooda Bashir. 2020. An analysis of phishing emails and how the human vulnerabilities are exploited. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 49–55.

[111] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. 2010. Who Falls for Phish?: A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions. In *2010 CHI Conf. on Human Factors in Computing Systems (CHI '10)*. ACM, 373–382. https://doi.org/10.1145/1753326.1753383

[112] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. 2007. Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. In *3rd Symp. on Usable privacy and security - SOUPS '07*. ACM, 88–99. https://doi.org/10.1145/1280680.1280692

[113] Kuldeep Singh, Palvi Aggarwal, Prashanth Rajivan, and Cleotilde Gonzalez. 2019. Training to Detect Phishing Emails: Effects of the Frequency of Experienced Phishing Emails. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications Sage CA: Los Angeles, CA, 453–457.

[114] Sergej Sizov. 2010. Geofolk: latent spatial semantics in web 2.0 social media. In *Third ACM Int. Conf. on Web Search and Data Mining (WSDM '10)*. 281–290. https://doi.org/10.1145/1718487.1718522

[115] Justin Snyder, Rebecca Knowles, Mark Dredze, Matthew Gormley, and Travis Wolfe. 2013. Topic Models and Metadata for Visualizing Text Corpora. In *Proc. of the 2013 NAACL HLT Demonstration Session*. Association for Computational Linguistics, 5–9. https://www.aclweb.org/anthology/N13-3002

[116] Human Factors Ergonomics Society. 2021. Human Factors and Ergonomics Society - Technical Groups. https://www.hfes.org/Connect/Technical-Groups. Accessed: 2021-06-10.

[117] Michael Stainbrook and Nicholas Caporusso. 2018. Convenience or Strength? Aiding Optimal Strategies in Password Generation. In *Advances in Intelligent Systems and Computing*. Springer, 23–32. https://doi.org/10.1007/978-3-319-94782-2_3

[118] Jeffrey M. Stanton, Kathryn R. Stam, Paul Mastrangelo, and Jeffrey Jolton. 2005. Analysis of end user security behaviors. *Computers & Security* 24, 2 (2005), 124–133. https://doi.org/10.1016/j.cose.2004.07.001

[119] Michelle P. Steves, Kristen K. Greene, and Mary F. Theofanos. 2019. A Phish Scale: Rating Human Phishing Message Detection Difficulty. In *Workshop on Usable Security (USEC'19)*. Internet Society. https://doi.org/10.14722/usec.2019.23028

[120] Timothy Summers, Kalle J Lyytinen, Tony Lingham, and Eugene A Pierce. 2013. How Hackers Think: A Study of Cybersecurity Experts and Their Mental Models. In *Third Annual Int. Conf. on Engaged Management Scholarship*. https://dx.doi.org/10.2139/ssrn.2326634

[121] Matt Taddy. 2012. On Estimation and Selection for Topic Models. In *Fifteenth Int. Conf. on Artificial Intelligence and Statistics*, Vol. 22. PMLR, 1184–1193. http://proceedings.mlr.press/v22/taddy12.html

[122] Ronnie Taib, Kun Yu, Shlomo Berkovsky, Mark Wiggins, and Piers Bayl-Smith. 2019. Social engineering and organisational dependencies in phishing attacks. In *IFIP Conference on Human-Computer Interaction*. Springer, 564–584.

[123] Dean Takahashi. 2020. *Unit 42: Phishing attacks are thriving during the pandemic.* https://venturebeat.com/2020/04/14/unit-42-phishing-attacks-are-thriving-during-the-pandemic/

[124] Hong Tang, Li Shen, Yinfeng Qi, Yunhao Chen, Yang Shu, Jing Li, and David A Clausi. 2012. A multiscale latent Dirichlet allocation model for object-oriented clustering of VHR panchromatic satellite images. *IEEE Trans. on Geoscience and Remote Sensing* 51, 3 (2012), 1680–1692. https://doi.org/10.1109/TGRS.2012.2205579

[125] Joe Tidy. 2020. *Google blocking 18m coronavirus scam emails every day.* https://www.bbc.com/news/technology-52319093

[126] Richard J Torraco. 2005. Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review* 4, 3 (2005), 356–367. https://doi.org/10.1177/1534484305278283

[127] Cybersecurity Ventures. 2020. Cybercrime To Cost The World $10.5 Trillion Annually By 2025. https://cybersecurityventures.com/cybercrime-damage-costs-10-trillion-by-2025. Accessed: 2021-06-10.

[128] Jingguo Wang, Tejaswini Herath, Rui Chen, Arun Vishwanath, and H. Raghav Rao. 2012. Research Article Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email. *IEEE Trans. on Prof. Comm.* 55, 4 (2012), 345–362. https://doi.org/10.1109/tpc.2012.2208392

[129] Rick Wash and Molly M. Cooper. 2018. Who Provides Phishing Training?: Facts, Stories, and People Like Me. In *2018 CHI Conf. on Human Factors in Computing Systems (CHI '18)*. ACM. https://doi.org/10.1145/3173574.3174066

[130] Zikai Alex Wen, Zhiqiu Lin, Rowena Chen, and Erik Andersen. 2019. What. Hack: Engaging Anti-Phishing Training Through a Role-playing Phishing Simulation Game. In *2019 CHI Conf. on Human Factors in Computing Systems (CHI '19)*. ACM. https://doi.org/10.1145/3290605.3300338

[131] Emma J. Williams, Amy Beardmore, and Adam N. Joinson. 2017. Individual differences in susceptibility to online influence: A theoretical review. *Computers in Human Behavior* 72 (2017), 412–421. https://doi.org/10.1016/j.chb.2017.03.002

[132] Emma J. Williams, Joanne Hinds, and Adam N. Joinson. 2018. Exploring susceptibility to phishing in the workplace. *Int. J. of Human-Computer Studies* 120 (2018), 1–13. https://doi.org/10.1016/j.ijhcs.2018.06.004

[133] Nick Williams and Shujun Li. 2017. Simulating Human Detection of Phishing Websites: An Investigation into the Applicability of the ACT-R Cognitive Behaviour Architecture Model. In *2017 3rd Int. Conf. on Cybernetics (CYBCONF)*. IEEE, 1–8. https://doi.org/10.1109/cybconf.2017.7985810

[134] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*. 1–10.

[135] Jeremy M Wolfe, Todd S Horowitz, and Naomi M Kenner. 2005. Rare items often missed in visual searches. *Nature* 435, 7041 (2005), 439–440.

[136] Yonghui Wu, Mei Liu, W Jim Zheng, Zhongming Zhao, and Hua Xu. 2012. Ranking gene-drug relationships in biomedical literature using latent dirichlet allocation. In *Biocomputing 2012*. World Scientific, 422–433.

[137] Aiping Xiong, Robert W. Proctor, Weining Yang, and Ninghui Li. 2018. Embedding Training Within Warnings Improves Skills of Identifying Phishing Webpages. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 61, 4 (2018), 577–595. https://doi.org/10.1177/0018720818810942

[138] Aiping Xiong, Robert W Proctor, Weining Yang, and Ninghui Li. 2019. Embedding training within warnings improves skills of identifying phishing webpages. *Human factors* 61, 4 (2019), 577–595.

[139] Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. 2004. Password memorability and security: Empirical results. *IEEE Security & privacy* 2, 5 (2004), 25–31. https://doi.org/10.1109/MSP.2004.81

[140] Shuangyan Yi, Zhihui Lai, Zhenyu He, Yiu ming Cheung, and Yang Liu. 2017. Joint sparse principal component analysis. *Pattern Recognition* 61 (2017), 524 – 536. https://doi.org/10.1016/j.patcog.2016.08.025

[141] Zhijun Yin, Liangliang Cao, Jiawei Han, Chengxiang Zhai, and Thomas Huang. 2011. Geographical topic discovery and comparison. In *20th Int. Conf. on World Wide Web (WWW'11)*. 247–256. https://doi.org/10.1145/1963405.1963443

[142] Kun Yu, Ronnie Taib, Marcus A Butavicius, Kathryn Parsons, and Fang Chen. 2019. Mouse behavior as an index of phishing awareness. In *IFIP Conference on Human-Computer Interaction*. Springer, 539–548.

[143] Xichen Zhang and Ali A Ghorbani. 2020. Human Factors in Cybersecurity: Issues and Challenges in Big Data. In *Security, Privacy, and Forensics Issues in Big Data*. IGI Global, 66–96. https://doi.org/10.4018/978-1-5225-9742-1.ch003

[144] Yin Zhang, Min Chen, Dijiang Huang, Di Wu, and Yong Li. 2017. iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. *Future Generation Computer Systems* 66 (2017), 30–35. https://doi.org/10.1016/j.future.2015.12.001

[145] Rui Zhao, Samantha John, Stacy Karas, Cara Bussell, Jennifer Roberts, Daniel Six, Brandon Gavett, and Chuan Yue. 2016. The highly insidious extreme phishing attacks. In *25th Int. Conf. on Computer Communication and Networks (ICCCN)*. IEEE. https://doi.org/10.1109/icccn.2016.7568582

APPENDIX

The Latent Dirichlet Allocation (LDA) technique is acknowledged as a very effective technique for identify topics within a corpus of documents [20]. This Appendix describes in more details how the LDA technique, as reported in Section 4.1, has been applied to identify the topics in the 52 papers selected by the SLR. It is worth remarking that this approach is similar to the Latent Semantic Analysis (LSA), but the main advantage of LDA is that the topic distribution is assumed to have a sparse Dirichlet prior [43]. This permits to represent a document as a small set of topics and each topic is codified through a small set of terms.

Before executing the LDA, we performed a pre-processing phase to prepare the corpus. For each paper, we only considered title, abstract, introduction and conclusion because they are the most informative and significant parts to use for summarizing its content. These four parts of each paper were extracted manually and translated from PDF to text by removing all the new line characters. A plain text file was created, composed of 52 rows, each for each paper. This text file was manually fixed to resolve problems introduced by the automatic translation from PDF to text (e.g., removal of hyphen symbols between two words that in the original PDF were a single word split due to carriage return).

Next, the 52 rows were cleaned by converting each of them into a list of lowercase tokens, and ignoring tokens that were too short or too long. In addition, we also removed email addresses, newline chars, quotes, English stop-words and a custom list of stop-words. The list of custom stop-words was defined iteratively by building LDA models and identifying those terms present in each topic with high weights, for example phishing, attack, phisher. The cleaning phase ended with the creation of bigrams and trigrams (in general, an n-gram consists in a n-words frequently occurring together in a document, e.g., human computer interaction is a trigram), and by performing a lemmatization of the resulting tokens to reduce inflected (or derived) words to their word stem, base or root form.

The cleaned corpus was processed using the *gensim* library to train the LDA model, which is the most adopted Python library for topic modelling, document indexing and similarity retrieval with large corpus. It is worth noticing that an important step we preliminary performed to build the LDA models was the estimation of the best number of topics. To this end, we exploited the *topic coherence*, which measures the relative distance between words within a topic, and it is widely adopted to choose the best number of topics when building LDA models. Since the LDA technique runs better with a large corpus and the one we analyzed includes only 52 documents, to build the LDA models we exploited the *passes* parameter of the gensim library, which permits to deal with small corpus. After experimenting with different values, we set the *passes* parameter equal to 100, meaning that each model related to a given number of topics has been built 100 times and the final LDA model is an 'average' of the 100 models. In order to identify the most suitable

number of topics, we analyzed the topic coherence scores of models with different number of topics. More specifically, we initially built 50 LDA models, i.e., the number of topics ranged from 1 to 50. We started using 50 since it would be a very high number of topics for 52 papers. We found out that LDA models have coherence scores increasingly lower after 5 topics, which indicates that the obtained models are less meaningful. This is evident in Figure 4, which shows the coherence values of the 9 LDA models; after 5, there is a visible drop of the coherence score. The models with a number of topics greater than 9 provide a coherence score even lower.

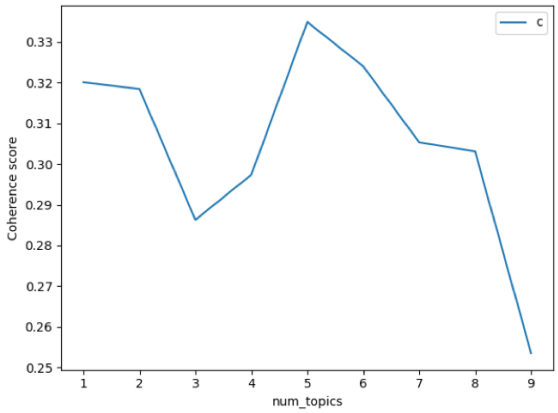

Fig. 4. Graph of the coherence values of the 9 LDA models built on the retrieved papers

This analysis has been triangulated with a manual inspection of the 9 LDA models, which confirmed that models with more than 5 topics are characterized by topics having terms with very low weights, i.e., these topics are meaningless. As a result, considering the coherence scores and the manual analysis of LDA models, we opted for an LDA model with 5 topics. Topics have to be interpreted by the analysts and it is widely recognized this interpretation could be very hard [26][100]. To this aim, the analysts usually consider different approaches based on statistical methods [19][86][121][18] or exploit proper visualizations [53][25][115][33][32]. One of the most recent and effective solutions is *LDAVis*. It is a web-based, interactive visualization that helps analysts easily inspect topic-term relationships of an LDA model. The interactive version of our LDA model built with LDAVis is available at this link https://bit.ly/38rwMlc. In this Appendix, we only report a screenshot of a section of LDAVis (see Figure 5), which shows the 5 topics as circles in a 2D plane (in each circle, the topic number indicated in Table 2 is shown). Since the LDA topics are calculated in a multidimensional scale, in order to project the inter-topic distances onto a 2D space LDAVis uses a multidimensional scaling method, i.e., the Joint Sparse Principal Component Analysis [140]. According to this reduction, x and y axes are the resulting two principal components of the starting inter-topics distances indicated as PC1 and PC2. The final topics are then depicted as circles whose center is determined by the Jensen–Shannon divergence method, which measures the similarity of probability distributions [51]: since the LDA method models a corpus of documents as a set of topics represented by terms and their probability, the Jensen-Shannon method measures the similarity between the probability distributions of each topic. The resulting values provide the centers of the circles onto the 2D plane and thus the circle distances indicate the similarity or difference of the represented topics. Finally, the topic's importance, which depends on the number of papers associated with the topic, is represented by the areas of the circles.

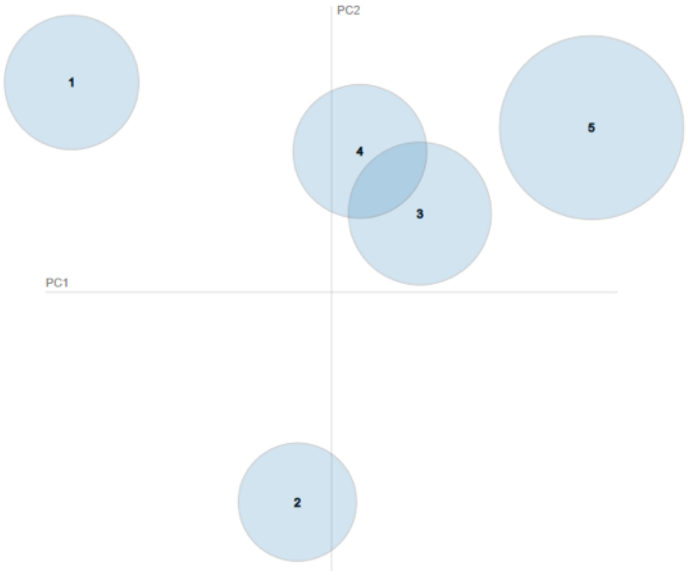

Fig. 5. Inter-topic Distance Map

Manuscript Accepted – Please Update CSUR-2020-0468.R2

Computing Surveys <onbehalfof@manuscriptcentral.com> 7 giugno 2021 16:05
Rispondi a: albert.zomaya@sydney.edu.au
A: marrella@diag.uniroma1.it
Cc: giuseppe.desolda@uniba.it, lsferro@diag.uniroma1.it, marrella@diag.uniroma1.it, catarci@diag.uniroma1.it, maria.costabile@uniba.it

07-Jun-2021

Dear Prof. Andrea Marrella:

Congratulations on the acceptance of your manuscript, "Human Factors in Phishing Attacks: A Systematic Literature Review". Your paper has been returned to your Author Center. You will find it under “Manuscripts Accepted for First Look.” Within two weeks, please upload your final files according to the instructions below:

Please note that CSUR has a limit of 35 pages. If the FINAL VERSION of your paper exceeds 35 pages, you will need to do the following:

- Either shorten the paper, or remove sections to be included as online only supplemental materials.

OR

- Seek out permission from the Associate Editor or Editor-in-Chief to go over the 35 page limit.

1. The final pdf of your paper (compiled from your final source files) – to upload as Main Document.
2. Zipped source files* (editable Word or LaTex files along with bibliography) formatted according to the style file found at: https://www.acm.org/publications/authors/submissions/, and any high resolution figures (if applicable) in .jpg, .png, .tif, or .eps format. *Upload these files as the “Other” designation.
3. If you have any supplementary online-only material (text and/or multimedia) for publication in the Digital Library, please provide a brief description of your material. A short “readme.txt” file will appear in the DL along with your supplementary material describing its content and whatever requirements there are for using it.

*IMPORTANT: Your zipped files should not be unpacked! 1. Delete previous files. 2. Upload your pdf as “Main Document (PDF)” as the first file. 3. Upload your zipped files, selecting the “Other” designation. 3a) If a supplemental file exceeds 100MB, load it into a separate zip. 4) Click on PDF to view your paper, then click "Save and Continue."

Please also include:

1. Computing Classification Systems Terms: http://www.acm.org/about/class/2012
2. Additional keywords and phrases

NB: If you have material owned by a third party, you must secure permission for its use before publication can proceed. If this is the case, please carefully read the guidelines at http://www.acm.org/publications/third-party-material and:

1. Obtain written permissions from the copyright holders.
2. Add the appropriate attributions to the figure captions in your paper (giving credit to the copyright holder).
3. Include the source of the third-party material in the references.

All author rights forms are now filled electronically through the ACM e-Rights Transfer Application. For further information about your rights options, please visit authors.acm.org. For questions concerning third-party material, please contact Stacey Schick, Assistant Editor, at schick@hq.acm.org.

Prior to publication, you will receive author proofs from the production manager.

Once your manuscript is published, we recommend that you use the ACM Author-Izer service, which allows you to post a link to your published article on your home page or institutional repository. Visitors to your personal bibliography can then download the definitive version of the article free-of-charge from the ACM DL. These downloads will be recorded as part of your DL usage statistics. Details may be found at: http://www.acm.org/publications/acm-author-izer-service.

Kindly submit your updates as soon as possible to avoid unnecessary delays in production.

Sincerely,
Computing Surveys Editorial Office