applied sciences

MDPI

Review

A Study on the Psychology of Social Engineering-Based Cyberattacks and Existing Countermeasures

Murtaza Ahmed Siddiqi [1], Wooguil Pak [1,*] and Moquddam A. Siddiqi [2]

[1] Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea; murtazasiddiqi@ynu.ac.kr
[2] Department of Sociology, University of Karachi, Karachi 75270, Pakistan; moquddam85@gmail.com
* Correspondence: wooguilpak@yu.ac.kr

Abstract: As cybersecurity strategies become more robust and challenging, cybercriminals are mutating cyberattacks to be more evasive. Recent studies have highlighted the use of social engineering by criminals to exploit the human factor in an organization's security architecture. Social engineering attacks exploit specific human attributes and psychology to bypass technical security measures for malicious acts. Social engineering is becoming a pervasive approach used for compromising individuals and organizations (is relatively more convenient to compromise a human compared to discovering a vulnerability in the security system). Social engineering-based cyberattacks are extremely difficult to counter as they do not follow specific patterns or approaches for conducting an attack, making them highly effective, efficient, easy, and obscure approaches for compromising any organization. To counter such attacks, a better understanding of the attack tactics is highly essential. Hence, this paper provides an in-depth analysis of the approaches used to conduct social engineering-based cyberattacks. This study discusses human vulnerabilities employed by criminals in recent security breaches. Further, the paper highlights the existing approaches, including machine learning-based methods, to counter social engineering-based cyberattacks.

Keywords: cybersecurity; cyberattack; information security; machine learning; psychology; phishing; social engineering

Citation: Siddiqi, M.A.; Pak, W.; Siddiqi, M.A. A Study on the Psychology of Social Engineering-Based Cyberattacks and Existing Countermeasures. *Appl. Sci.* **2022**, *12*, 6042. https://doi.org/10.3390/app12126042

Academic Editors: Howon Kim and Thi-Thu-Huong Le

Received: 24 May 2022
Accepted: 11 June 2022
Published: 14 June 2022

1. Introduction

At present, the digital world is growing at an exponential rate. People are embracing the internet as a new way of communication, business, knowledge, and entertainment. However, this shift in paradigm is raising serious concerns for the online security and privacy of users. Despite rigid security measures, a high number of internet attacks are conducted by exploiting application design vulnerabilities, scamming, or advanced technical methods [1,2]. Among these attack methods, scamming has been around even before the existence of computers and the internet. In terms of cybersecurity, scamming or phishing are categorized as social engineering (SE) attacks. Attacks that involve exploiting human vulnerabilities are classified as SE attacks [3]. SE attacks are conducted by exploiting human vulnerabilities, such as deception, persuasion, manipulation, or influence [4].

The mentioned vulnerabilities are exploited to acquire confidential information, unauthorized access, knowledge of cybersecurity measures, etc. The main concern for security experts in countering SE-based attacks is the absence of a pattern or methodology. In some cases, even the target is unaware that they are being manipulated or influenced by the perpetrator. Due to such hazy characteristics of SE-based attacks, existing countermeasures struggle to stop such attacks. Recently, there were several large-scale security breaches, compromising the vulnerable human factor [5–7]. Despite the advancements in technology, humans play an integral part in functional organization. This human element in the organizational security chain is inevitably targeted and exploited by hackers. Figure 1 shows some of the most common tools and a basic flow of cyberattacks based on SE [8,9].

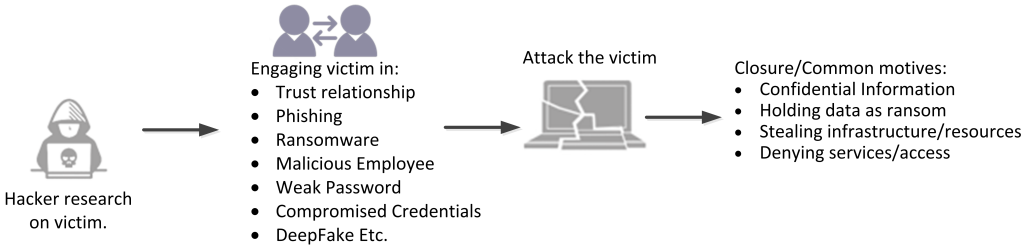

Figure 1. A basic flow of social engineering-based cyberattacks.

In the last two years, the most prominent mediums used to conduct SE attacks were social media and smishing attacks. The majority of cyberattacks based on SE rely on actual interactions between the attacker and the victim. In some cases, SE attacks can involve a simple phone call, with an individual impersonating an employee to garner information, such as a password or a PIN code. In 2020, Americans lost approximately USD 29.8 billion in phone scams [10]. Table 1 provides a broad overview of attacks based on SE methods. The attacks highlighted in Table 1 are some of the most prominent breaches in the last few years. These breaches occurred due to a combination of human errors and SE attacks. Based on Table 1, it can be seen that human error can also play a key role in conducting SE attacks. To conduct SE attacks, hackers can even influence a victim into making an error. Common methods used to influence targets are discussed later in the paper. Other than the financial costs, data breaches can also lead to the loss of customers due to security concerns. These attacks are highly impactful and easy to conduct. Based on the studies conducted for this work, as well as our understanding, the countermeasures for SE-based attacks lack the understanding of human behavior. Few of the published works highlight the interconnection between human vulnerabilities and SE attacks. This study further enhances the existing knowledge of human behavioral vulnerabilities and how they are exploited by hackers. After highlighting the importance and impact of cyberattacks based on SE, the rest of the paper is organized as follows. Section 2 provides a brief overview of well-known cyberattacks based on SE and how they are conducted. Section 3 provides a perspective on common human emotions or errors that can be exploited to conduct SE attacks. The section also presents the current standings and concerns of machine learning (ML) and existing countermeasures against SE-based cyberattacks. Section 4 highlights the concerns of existing solutions, including a summary of the topics discussed in this paper. Section 5 concludes the paper.

Table 1. Some of the most prominent social engineering-based cyberattacks.

Reference	Company	Date	Details/Damage	Breach Method/Tools
[5]	Saudi Aramco	2021	The hackers claimed that they had almost 1 terabyte worth of Aramco data and demanded USD 50 million as ransom.	Phishing email.
[11]	Microsoft	2021	Several MS Office users fell for the phishing email scam. Each victim was scammed for USD 100 to 199.	Business email compromise (BEC) attack, Phishing email
[6,7]	Marriott	2020–2018	On both occasions, hackers acquired access to millions of guest records. These records included guest names, addresses, contact numbers, and encrypted credit card information.	Phishing email, compromised credentials of two Marriott employees, remote access Trojan (RAT), Mimikatz post-exploitation tool.
[12]	Twitter	2020	Hackers compromised 130 Twitter accounts. Each account had at least 1 million followers. Hackers used 45 highly-influential accounts to promote a Bitcoin scam.	Impersonation-based SE attack, spear-phishing attacks.
[13]	Shark Tank	2020	Shark Tank host lost USD 400,000 after falling for a scam email.	Phishing email.
[14]	Toyota	2019	The Toyota Boshoku Corporation lost USD 37 million after falling victim to a BEC attack.	Phishing email (i.e., BEC)

Table 1. *Cont.*

Reference	Company	Date	Details/Damage	Breach Method/Tools
[15]	Energy firm (U.K based)	2019	The Chief Executive Officer (CEO) was deceived and scammed for USD 243,000 by the hackers.	Deepfake phishing impersonation
[16,17]	Google and Facebook	2015–2013	The phishing emails cost both Google and Facebook over USD 100 million.	Spear

2. Social Engineering Attacks

Social engineering can be defined as a process used to exploit human psychology, rather than a sophisticated hacking method [18]. With growing reliance on technology, SE is becoming a key tool for cyberattacks. Conversely, due to the increasing number of cyberattacks, the technical methods used to counter cyberattacks are also improving [19]. These continuous improvements of security methods are making technical attacks difficult to execute. On the other hand, SE is proving to be a highly successful approach used to conduct cyberattacks. Cyberattacks based on SE can facilitate attackers in several ways, i.e., infiltrating organizational networks, bypassing firewalls, infecting systems with malware, opening back-doors in the organization network, etc. Cyberattacks that use SE can exploit or influence human behavior in many ways. For instance, human error can be used to initiate an attack. Attackers can enforce an individual to err by influencing the decision-making process. Details on how decision-making is influenced are discussed later in the paper. Similarly, several other human behaviors can be exploited to conduct cyberattacks based on SE [20]. In this section, some of the most common cyberattacks based on SE are discussed.

2.1. Phishing Attack

Phishing attacks are among the most successful attack methods in SE-based attacks. Each day, millions of scam emails are sent by hackers, some are detected and blocked by different technical solutions [21]. However, some scam emails do manage to evade these systems. Phishing attacks typically start with a scam email that lures a victim into a trap. For instance, a phishing email may appear to come from an authentic source.

The method of luring a victim can vary depending on who the victim is, i.e., the email may ask the victim to click the link for your travel expense receipt, click the link to win a prize, etc. Falling victim to such phishing emails is based on human behavior attributes [22,23]. The process of a common phishing attack can be seen in Figure 2. A phishing attack can be conducted in several different ways. Two of the most common methods can be seen in Figure 2. As these attacks evolve, the goal of any type of phishing attack is to steal personal credentials or information. Some phishing attack types can be seen in Table 2 [24,25].

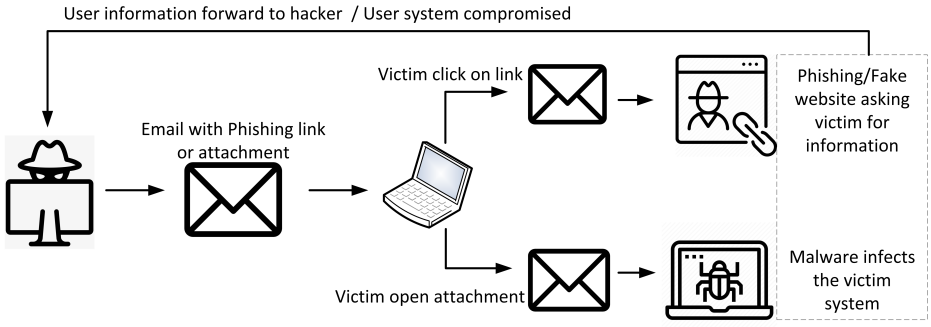

Figure 2. The common flow of phishing attacks.

Table 2. Common types of phishing attacks.

Phishing Attack	Description
Spear	Tailored attack to target a specific individual. For instance, an employee is targeted to gain access to an organization's network.
Whaling	The intended target is usually a high-profile individual. The attack requires considerable time to find the opportunity or means to compromise the individual's credentials.
Vishing	Vishing or voice phishing is an attack based on SE. Vishing is a fraudulent call intended to acquire classified information or credentials of the target individual.
Smishing	Smishing is a text message format of a vishing attack. In smishing, the only difference is that it is based on a text message rather than a call.
Impersonation/business email compromise (BEC)	BEC is an attack that requires planning and information. In a BES attack, the attacker impersonates a company executive, outsourced resource, or a supplier to acquire classified information, access to an organizational network, etc.
Clone	Clone phishing is an email-based phishing attack. In these attacks, the malicious actor knows most of the business applications being used by a person or an organization. Based on this knowledge, the attacker will clone a similar email disguised as an everyday email from the application to extract critical information or even credentials from the target.
Social media phishing	In social media phishing, the attacker works by observing the target individual's social media and other frequently visited sites to collect detailed information. Then, that attacker plans the attack based on acquired information. The attacker can use the gathered information to trick the victim in many ways.
Distributed spam distraction (DSD)	The DSD attack is executed in two steps. In the first step, the victim is spammed with phishing emails mirroring an authentic or credible source, i.e., a new letter, magazine, software company, etc. These fake emails contain a link that leads the victim to a web page that is a copy of an authentic and credible company's website. The second step depends on how the attacker plans to conduct the SE attack, i.e., the fake page could ask the victim for the login information (to garner further or confidential information) to confirm the identity and proceed further.

2.2. Dumpster Diving

The dumpster diving attack is a low-tech method used to obtain information on the target [26] . The process involves going through trash and looking for torn documents, receipts, and other forms of paper that could contain information on the target. An individual may throw away a piece of paper that may contain information about a password, pay slip, bill, credit card information, or something containing critical information. Such information can help hackers with several methods to conduct a SE attack on an individual. Dumpster diving is also among the most common methods of identity theft [27]. Figure 3 elaborates on some of the content that can be found in an organization's dumpster. As seen in Figure 3, the dumpster could contain any of the mentioned information. Normally these documents or papers are torn and thrown into the trash. A malicious actor can go through the dumpster and retrieve information to assist in conducting a SE-based attack.

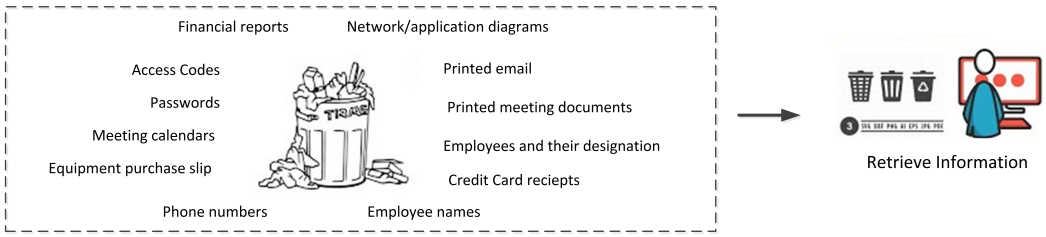

Figure 3. Content that can be found via dumpster-diving in an organization's trash.

2.3. Scareware

Scareware can be defined as a type of SE attack that is based on human emotions, i.e., anxiety, shock, manipulation, etc. [28] . The attack uses human emotions to manipulate the user into installing malicious software. The steps involved in a scareware attack can be seen in Figure 4 [29]. It can be seen (in Figure 4) that hackers use pop-up-based alerts on different sites to engage the target. Once the target clicks the pop-up, he/she is targeted using misinformation. The misinformation is intended to influence the target to perform an act out of panic. The intended act may ask the target for sensitive information or to

buy a product to solve the fake issue. In a scareware attack, the hacker only needs to persuade the victim to click on a link. For this persuasion, the attacker can use numerous techniques to influence the victim into installing the scareware malware. The graphical interface of scareware is in many ways an integral part of deceiving victims. The visual representation of the scareware (e.g., a pop-up or a scan report) meritoriously presents a credible and trustworthy application. Most forms of scareware malware adopt color schemes, font styles, and logos that are similar to known brands of antivirus or software products, e.g., Microsoft, Norton antivirus, etc. [29].

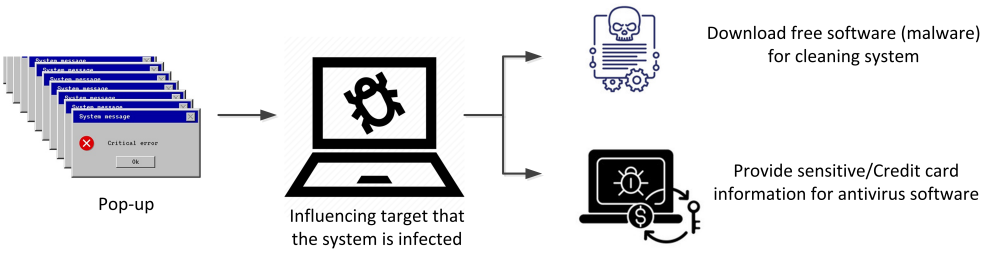

Figure 4. Steps of a scareware attack.

2.4. Water Hole

A water hole attack or water holing is a SE attack inspired by the hunting method of predators in the jungle. In the real world, the predators in the jungle wait near a water hole to attack their prey [30]. Figure 5 provides an overview of how a water hole attack is commonly conducted [31]. In step one, the attacker identifies the organization to target, then uses surveys and other means to identify the browsing habits of the employees. Based on this information, the hackers identify the frequently visited website by the employees. In step two, the attacker compromises the legitimate website. Compromising a secure site could be near impossible, so hackers must identify the website that can be compromised (one that is frequently visited by the employees). In step three, the hackers direct the employee from the original website to a malicious website. This malicious website attempts to identify the vulnerabilities in the victim's system. To identify the vulnerabilities, hackers use different methods, i.e., operating system fingerprints, analyzing the user–agent string, etc. In step four, the hacker exploits the vulnerability identified by the earlier scan. Once the system is compromised, the hacker can further advance the attack by infecting other systems on the network, achieving the desired goal [32].

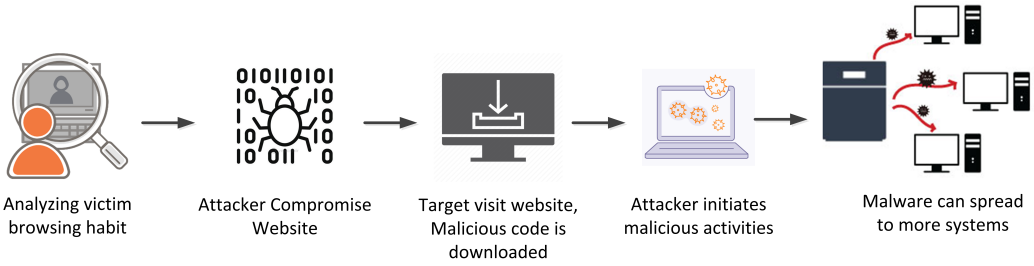

Figure 5. Steps of a water hole attack.

2.5. Reverse Social Engineering

Reverse engineering attacks are among the simplest yet most effective attack methods in SE attacks. Reverse engineering attacks are carried out in two steps. First, the malicious actor creates a problem for the target. The core idea behind the creation of the problem is to initiate an interaction with the target. Second, the malicious actor approaches the victim with the solution to that problem [33]. For example, the foe identifies a potential target in an organization and intentionally creates a problem related to the information technology (IT) department. Later, the foe poses as a person from the IT department, or a benefactor, and offers assistance. Such acts are used to gain the trust of the intended target. With time, the

malicious actor gains more trust and exploits the trust factor to gain sensitive information or manipulate the victim [34]. Reverse engineering attacks are very common attacks at an organizational level.

2.6. Deepfake

Deepfake is a recent and highly convincing technique used to conduct SE attacks. Cybercriminals use deepfakes to forge images, audio, and video to achieve a particular goal. In cybersecurity, the deepfake is a growing threat [35]. One of the most well-known algorithms for generating deepfake content is generative adversarial networks (GANs) [36]. GANs are a combination of two artificial neural networks (ANNs).

These ANNs are called detectors and synthesizers. These ANNs are trained using large datasets of real images, audio, and video clips. Then, the synthesizer ANN generates deepfake content and the detector ANN attempts to distinguish the authenticity of the content. The cycle of generating deepfake content continues until the detector ANN is no longer able to identify the generated fake content as fake. Due to this rigorous process of generation and validation, the generated forged content by GAN is very difficult to identify as fake. Figure 6 illustrates an overview of the process of creating deepfake material. In Figure 6, it can be seen that two different faces i.e., Face A and Face B are used to train a network. Later the network is used to generate Face A with expressions or audio from Face B. The newly generated image with the original Face-A interpretation by Face-B can be used to confuse or influence a victim. In Table 1, deepfake was employed for the SE attack conducted on a UK-based energy firm. In the attack, the deepfake voice was used to scam the CEO of the company [15]. Other than scamming, deepfake has also been used in several other criminal activities, i.e., blackmailing, damaging reputation, fake news, misinformation, mass panic, etc.

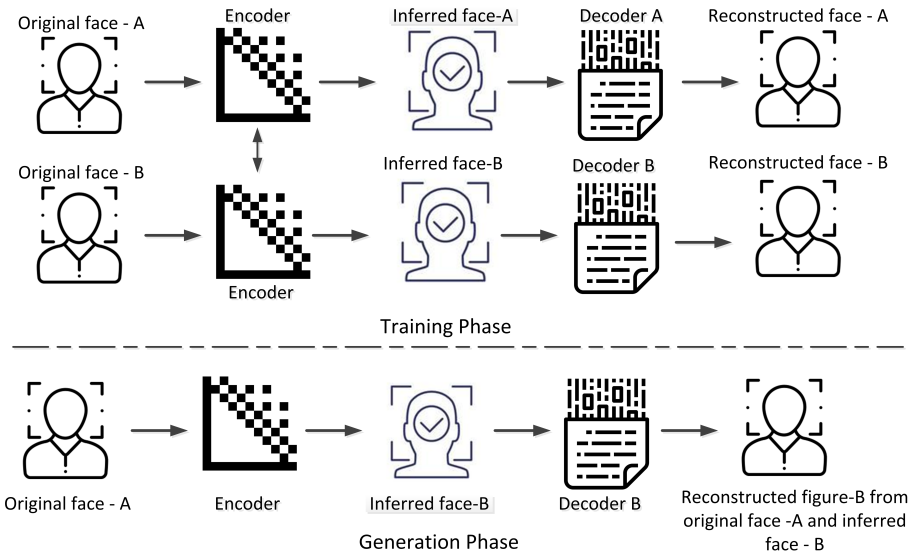

Figure 6. Simplified illustration of training and generating deepfake images or videos.

3. Influence Methodologies

To initiate a SE attack, the attacker needs some form of influence on the target. This section discusses the different methods to influence and compromise a victim. The prominent human behavioral aspects that are used to initiate a SE attack are social influence, persuasion, attitude and behavior, trust, fraud, decision making, emotions, language, reasoning, etc. [3]. The mentioned behavioral aspects are also used in cyberattacks based on SE. Based on the victim attributes, the attackers exploit the most suitable human vulnerability. The attackers can also use multiple vulnerabilities at different stages of an attack.

3.1. Social Influence

Social influence includes both intentional and unintentional efforts to alter another person's behavior or attitude. Usually, social influence works through peripheral processing. In such an approach, the victim may be unaware of the influence attempt by the attacker [37]. In general, social influences are categorized into three types utilitarian, value-expressive, and informational [38].

Informational influence refers to the influence of an individual over a particular group. For example, while shopping, a member of a group suggests what to buy based on earlier 'experience' information. The utilitarian influence refers to the influence on an individual or a group based on a reward or something similar. Value expression influence is influence due to interests or beliefs. For instance, on social media, people join groups based on their hobbies and interests. To further elaborate on how these influence methods work, Table 3 provides a more detailed look into social influence interlinked with SE attacks.

Table 3. Social influence methods related to cyberattacks based on social engineering.

Types of Social Influence	Description
Group influence [38]	Conformity is a variation in behavior to agree with others. A group can influence such behavior. On social media, online groups may be used to influence victims into falling for a SE attack. For example, a social media group with hundreds of subscribers can present a subscriber with a malicious link and influence the victim by informing that every member must go to the link to be a part of this new group for future events and information.
Informative influence/normative influences [39]	In SE attacks, the foe often uses particular information and setups with the help of informational/normative influences. For example, informing the victim about some free software and convincing the victim to install it by providing information on software importance and ease of use. Such an approach can be used to puzzle or manipulate the victim into performing certain actions or revealing information that benefits the foe.
Social exchange theory/reciprocity norm [34,40]	Such methods of influence are used in reverse SE attacks. The social exchange theory highlights that people make decisions on the value (intentionally or unintentionally) of a relationship. For example, while working in a corporate environment, coworkers exchange favors based on their relations with each other.
Moral influence/social responsibility [41]	SE attacks use moral influence or social responsibility in two ways. One way is that the foe exploits the victim's helpful nature to extract information or to gain favor to facilitate the attack. The second way is to exert pressure of social responsibility norms or moral duty on the victim during the SE attack. This pressure of moral duty influences the victim's behavior. Specifically, if the victim is not keen to offer any help. An example could be an online group to help animals. The malicious actor can identify the individuals who are highly motivated and willing to help. The attacker can target that victim for financial gain or can fabricate a story to target the moral values of the victim to extract information.
Self-disclosure/rapport relation-building [42,43]	Research shows that during the process of building social relations, self-disclosure causes a willingness to reveal more to people who show connections to us. Adversaries use this SE method on victims who feel the need to connect with someone special.

3.2. Persuasion

Social engineering relies on persuasion methods to manipulate victims into performing actions or revealing confidential information. Persuasion is a well-known method that is used in several other domains, such as sales, marketing, insurance, etc. [44]. As per Robert Cialdini [45], there are six main principles of persuasion—reciprocation, commitment and consistency, social proof, authority, liking, and scarcity. Reciprocation defines the behavior in which an individual replies to kindness with kindness. For example, if a coworker buys you lunch; you will feel obliged to buy him/her lunch the next time. A sense of commitment and consistency can be defined as a desire to be consistent with behavior. This behavior can include moral values, music, favorite food, etc. Social proof can be referred to as peer pressure. It refers to the intentional or unintentional acts of doing what everyone else is doing, e.g., if a group is looking out of a window, anyone who sees them will also look out of the window. The principle of liking is simply the act of agreeing with people whom we like. Similarly, the principle of authority refers to the act or tendency of individuals to follow authority. The principle of scarcity refers to the approach of persuasion using time-based constraints. For example, limited-time sales are used to persuade potential buyers to buy products before time runs out and the product price is increased. These six

principles play a key role in SE attacks that rely on persuasion. Further, Table 4 highlights some of the persuasion methods used in conducting cyberattacks based on SE.

Table 4. Persuasion types used in social engineering attacks.

Types of Persuasions	Description
Similarity [46]	The similarity of interests invites likeness and dissimilarity leads to dislike. The criminal tends to use this as an effective approach to gain the trust of the victim. For example, on social media platforms, a foe may join groups that are similar to the groups joined by a potential target. Such similarities can help build a relationship of trust between a hacker and a victim.
Distraction/ manipulation [25,47,48]	Research shows that moderate distraction does facilitate persuasion. Distraction is used as an effective tool in manipulation attacks. An example of a distraction-based SE attack is the DSD attack highlighted in Table 2.
Curiosity [49]	The majority of individuals are curious by nature. In a SE attack, human curiosity can be exploited in many ways. For example, the attacker can send a phished email or an infected file as an attachment with a curious subject line, i.e., you are fired, annual performance report, employee layoff list, etc.
Persuasion using authority/ credibility [50,51]	Most people tend to comply in front of an authoritarian figure. On the internet, hackers use symbols and logos that reflect authenticity and authority. For example, an official logo of taxation, law enforcement, etc., to show authority and credibility can be an effective approach to initiate a SE attack.

3.3. Attitude and Behavior

The theory of planned behavior (TPB) describes a psychological model to predict behavior. The TPB model can be seen in Figure 7 [52]. The 'attitude toward the behavior' defines the motivation factors, i.e., the effort a person is willing to put in, to show a certain behavior. For example, an individual on social media might not be willing to share personal information but might participate in an online activity where privacy might be in significant danger. The 'subjective norm' defines an individual's behavior influenced by his/her social circle. A person may align his/her behavior to be associated with a group. The 'perceived behavior control' indicates the level of control an individual has over a certain behavior. The three branches of TPB can be used to encourage a victim into revealing information based on the behavioral stimulus [53].

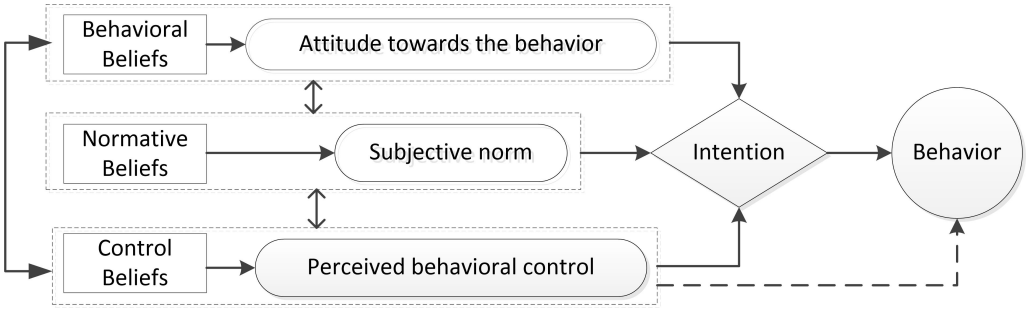

Figure 7. Theory of planned behavior.

The methods to influence the attitudes and behaviors of victims can further be divided into sub-domains, as shown in Table 5.

Table 5. Methods to influence the attitudes and behaviors of victims to conduct social engineering attacks.

Methods to Influence Attitudes and Behaviors	Description
Impression/ commitment [54,55]	The self-presentation theory highlights the fact that every individual presents a likable impression, both internally and to other people. An individual might put a lot of effort into creating a desirable image. Such efforts can be an opening for hackers to conduct SE attacks. For example, an individual's behavior can be influenced if the social image of that person is threatened.
Cognitive dissonance theory [56,57]	The theory highlights the inner conflict when an individual's behaviors and beliefs are not aligned with each other. Such conflict can influence cognitive biases, i.e., decision-making. A malicious actor can exploit these cognitive biases for extracting confidential information.

Table 5. *Cont.*

Methods to Influence Attitudes and Behaviors	Description
Behavior affects attitude [58]	The act that once an individual has agreed to a minor request, he/she is more likely to comply with a major request, is known as the foot-in-the-door effect. In essence, people build an image by performing a minor favor; to maintain this image, they tend to agree on the next favor. SE attackers can use such behavior to initiate an attack.
Bystander effect [59]	The bystander effect defines human behavior involving an individual who is reluctant to help when bystanders are present. In SE attacks, victims may be tempted into a specific situation while in a group and exploited later in a private chat to acquire personal or confidential information.
Scarcity/time pressure [41]	In SE attacks, the attacker uses scarcity to enforce a feeling of panic or urgency. This panic/urgency can influence the decision-making abilities of the victim. Due to this confusion, the attacker can persuade the victim into making decisions deemed desirable by the attacker.

3.4. Trust and Deception

On social media and in virtual networking environments, users exhibit trust levels based on their engagement on the virtual platforms [60]. The higher the engagement on virtual platforms, the higher the level of trust in them. This level of engagement can be measured in many ways, i.e., number of friends or connections, posts, groups followed, etc. Users that show high levels of social or virtual network engagement are more exposed to SE attacks [61]. In SE attacks, trust and deception can further be classified into sub-domains, as shown in Table 6.

Table 6. Sub-domains of trust and deception to conducting social engineering attacks.

Sub-Domains of Trust and Deception	Description
Trust/relation [4]	Building trust is one of the most crucial parts of SE attacks. An attacker can use multiple means to develop a trusted relationship with the victim. Methods such as influence, persuasion, likeness, reward, etc., can be used to build a trusted relationship with the target. As per the research, once a relationship of trust is developed, the victim does not feel hesitant to be vulnerable in front of the trusted individual. Such a relationship can be a risk-taking behavior that could assist in a SE attack.
Deception/fraud [55,62,63]	Deception is an intentional act based on strategic interaction by the deceiver. The interpersonal deception theory (IDT) suggests that humans believe that they can identify deception but a majority of the time they do not. The IDT also highlights that the malicious actor takes every action based on a strategic plan to manipulate the victim's behavior. A SE attack based on deception can rely on several methods, e.g., lying, creating false fiction or narratives, partial truths, dodging questions, giving the impression of misunderstanding, etc.

3.5. Language and Reasoning

Language is not only the most common method of communication, but it is also a means of processing, generating, and expressing thoughts. The process of social interaction using language is very similar to a programming language process. Humans hear the words as input, process those words, and generate a response as output. Seeking appropriate words to explain the context of the subject or feeling is important. This implies that crafting and elaborating information for SE attacks relies heavily on the language being used for interacting with the victim, i.e., the language cognition is exploited [64,65]. Table 7 provides an overview of language and reasoning sub-domains associated with SE attacks.

Effective cyberattacks based on SE rely on human vulnerabilities. The attackers exploit human behavior, knowledge, emotions, cognition, personal traits, human nature, etc. Table 8 highlights the human vulnerabilities associated with SE attacks. A single SE attack can be conducted using multiple influence methods. A single influence method can exploit several human vulnerabilities. This multidimensional interconnection between SE attacks, influence methods, and human vulnerabilities makes SE-based cyberattacks challenging for security professionals.

The mapping in Table 8 provides a simple yet elaborate perspective on the working of SE attacks. Such mapping can assist greatly in identifying the vulnerabilities and building an effective security infrastructure to mitigate them. The highlighted interconnection

between human vulnerabilities and SE attacks may not be sufficient to represent every age group or human behavior; however, it can provide a broader perspective to individuals and organizations on understanding and countering these SE-based cyberattacks.

Table 7. Sub-domains of language/reasoning linked with social engineering attacks.

SUB-Domains of Language/Reasoning	Description
Framing effect/cognitive bias [66]	The phenomena of reflecting cognitive biases, i.e., expressing opinions and decision-making, are influenced by the way a question is asked. This cognitive bias based on language-framing leads to decision manipulation. For example, beef labeled "75% lean" is more preferred by customers as compared to the label "25% fat". In SE attacks, the cognitive biasing of a victim is exploited by using a pre-planned language framework.
Complicating the thinking process [67–69]	Language plays an integral role in the thought process for social interaction. This reliance can create an opportunity to invoke 'thinking confusion' through language. For example, to induce thinking confusion, an attacker can engage his/her victim in a statement with non-grammatical or unclear meaning. Such statements can tempt the victim into acting based on presumption, i.e., a statement "that's touching to hear" may result in the victim touching his ear. Another example can be an incomplete statement, such as "I can't hear", which could encourage the victim to check or adjust the equipment.

Table 8. Association between social engineering-based cyberattacks, method of influence, and human vulnerability.

No.	Attack	Methods to Influence Victims	Human Vulnerability Exploited
1	Impersonation, business email compromise (BEC), clone phishing	Moral influence social responsibility, similarity, persuasion using authority/credibility, interpersonal deception theory (IDT), complicating the thinking process, curiosity	Being helpful charity, kindness, trying to be acceptable in social norms
2	Spear phishing, pretexting, smishing phishing, whaling phishing, vishing phishing deepfake	Similarity, persuasion using authority/credibility, impression/commitment behavior affects attitude, scarcity, deception/fraud, complicating the thinking process	Being helpful, being obedient to authority, helping nature, panic negligence
3	Social media phishing, scareware, reverse-engineering attack, deepfake	Group influence, informative influence/normative influence, social exchange, theory/reciprocity norm, moral influence/social responsibility	Friendly nature, negligence, trustful nature, credibility
4	Waterhole attack, deepfake	Persuasion using authority/credibility, framing effect/cognitive bias, complicating the thinking process	Curiosity, greed, excitement fear

3.6. Countering Social Engineering-Based Cyberattacks

Human awareness can be identified as a key factor in countering SE attacks. SE methods focus on hacking humans by targeting human cognitive biases rather than machines. The methods to counter SE-based cyberattacks can be seen in Table 9. The ML-based approaches to counter and detect SE-based cyberattacks are discussed separately in Section 5. Based on recently published work, most researchers find the following methods to be highly effective to counter SE attacks. Based on Table 9, it can be concluded that training and educating individuals on cybersecurity and SE attacks can play a significant role. In Table 9, the checkmark highlight the countermeasures suggested by the referred study.

Several researchers have highlighted the role of well-defined policies to counter cyberattacks. Policies that can be implemented to avoid and manage an event of a data breach or SE-based cyberattacks. Organizational policies are further categorized into two main groups cybersecurity and communication policies. Cybersecurity policies are defined specifically for cyberattacks. Such policies may include instructions on avoiding illegal software, use of personal devices on the company network, steps to follow in case of cyberattacks, documentation of a cyberattack, human resources (HR) procedures for third party vendor privileges and access, critical area access management, password management, organizational security infrastructure, etc.

Table 9. Suggested method to counter social engineering attacks.

No.	Training	Cyber Security Policies	Communication Policies	Company Equipment	Spam Filter/ Antivirus/ Firewall	Encrypted Communication	Password/Data Management	Incident Report
[8]	✓	✓	✓		✓	✓	✓	
[33]	✓	✓	✓		✓	✓		
[58]	✓				✓			✓
[62]	✓	✓			✓	✓	✓	✓
[70]	✓							
[71]	✓	✓	✓					
[72]	✓	✓	✓	✓				
[73]	✓	✓						
[74]	✓		✓	✓				
[75]	✓	✓	✓					
[76]	✓	✓			✓			
[77]	✓	✓		✓				
[78]	✓	✓						
[79]	✓	✓			✓	✓	✓	
[80]	✓		✓		✓			

A well-defined cybersecurity policy can limit many data breaches or cyberattacks. Moreover, organizational policies for official (or in some cases, unofficial) communication should be implemented. The reason to implement a separate set of policies for communication is based on the high number of SE attacks that exploit communication norms to conduct cyberattacks (i.e., Table 1). An organization should define clear policies for official communication within and outside the organization. Such policies may include a process of approval and validation for connecting a personal device to the organizational network, what level of information can be shared on emails, SMS, or calls, procedures to validate the authenticity of suspicious email, SMS, or calls, communication methods in case of working from home, etc. Organizational communication policies can play an integral role in avoiding any cyberattack based on SE. The importance and need of appropriate antiviruses, firewall, spam filters, and updated software patches cannot be underestimated and must be installed on both organizational and personal systems. Some of the research and survey papers on SE attack mitigation have encouraged organizations to provide company equipment to employees, as equipment managed by the organization's IT department can easily be updated with security software and can be checked frequently for malicious software. Due to recent advancements in ML, the ML-based approaches are discussed separately in the following section.

3.7. Machine Learning-Based Countermeasures

Machine learning (ML) is another domain that can play an important role in countering SE-based cyberattacks. For phishing-based attacks, ML models can be trained to identify patterns and language in emails, SMS, malicious links, and even calls using natural language processing (NLP) [58,71]. However, the continuous evolution of phishing characteristics can be a concern for ML-based methods. In this section, some of the most significant ML methods to counter SE-based cyberattacks are presented. The section also highlights the existing concerns with the discussed ML methods.

3.7.1. Deep Learning

Deep learning (DL)-based approaches can also play a vital role in countering SE-based cyberattacks since DL has been an effective approach used to counter a wide range of malware, phishing attacks, traffic analysis, spam detection, intrusion detection, etc. [81–86]. For instance, deep neural networks (DNNs) in DL are inspired by the human brain. As more data are fed to a DNN, it gradually becomes better at detecting malicious dialog. Due to this reason, Google is also using neural networks to detect spam emails [72]. When it comes to phishing, DNN-based solutions can be highly effective. In [87], the authors proposed a hybrid model based on DNN and long short-term memory (LSTM)

to identify phishing web links. The authors used NLP to select features and character embedding-based features for the DNN-LSTM model to identify phishing website links. The model was trained on two datasets—Ebbu2017 and a secondary dataset that was based on several internet resources. Even with the high detection rate by the proposed model, the authors stated the concerns on the datasets used. The datasets used may lack the precise representation of real work attacks. The papers [88,89] also presented DL-based approaches to counter SE-based attacks initiated through Twitter, DNS, URL, and email. The authors presented elaborate insight into the origination of ransomware based on different case studies. However, the absence of datasets representing sophisticated SE-based attacks may hold the key to enhanced DL to counter SE-based cyberattacks. This concern is highlighted by K. Simar, et al. in [90]; in their study, they presented a DL-based approach to structure the unstructured data generated by different internet sources. However, handling misinformation can be a concern in the proposed approach. When validating the authenticity of the source, publishing an event or article may still be questionable. Nonetheless, if the source validation process can be enriched, the proposed approach can play an integral role in improving the DL-based approach against SE attacks.

3.7.2. Reinforcement Learning

Reinforcement learning (RL), another aspect of ML, is also a method used to counter SE-based cyberattacks. In [91], the authors proposed a cyber-resilient mechanism (CRM) to counter online threats, including uncertain real-time scenarios. The model used the feedback architecture of RL to define policies, as the system observes the online actions. However, the model requires online observations to learn and adapt unknown attack methods used in SE-based cyberattacks. In [92], the authors used the RL-based greedy approach. The authors used predefined attack and defense approaches (i.e., Petri net) to train the RL model. Based on the experimentation results, the authors concluded that the model gradually improved its performance to identify a cyberattack. The goal of the paper was to highlight the potential of RL to counter cyberattacks. The main concern for the RL-based approach is the observation of an unlimited range of human behaviors. As time goes by, the data related to human behavior will grow exponentially, making it difficult to track and store information [93].

3.7.3. Natural Language Processing

One of the most convenient tools in ML to counter phishing-based cyberattacks is NLP [94]. NLP with ML has played an important role in countering phishing attacks [95]. Several NLP processes, e.g., information extraction, text categorization, and machine translation, are inspired by DL [96]. The NLP relies on five main features to identify phishing emails or online links. Those features are email body characteristics, email subject, uniform resource locator (URL) characteristics, hidden script (i.e., JavaScript, pop-up on click activity, etc.) characteristics, and sender characteristics. A paper by Tim Repke et al. [97] used the word-embedded technique with DL to analyze email text for human mode identification. This technique is not used to identify phishing, but it can be useful for identifying abnormalities in the usual email text, which can help in identifying email phishing, i.e., impersonation, BEC, clone phishing, etc. The only concern for the ML-based NLP model is the dependency on the surface text of an email. If the structure of dialog or a sentence is altered, it becomes difficult for the model to identify as phishing [96]. However, the key challenge for ML to counter SE-based cyberattacks is the absence of an attack pattern or a methodology that can identify the multidimensional approach to SE-based cyberattacks [86]. As discussed in Section 3, SE-based attacks exploit human vulnerabilities, making them beyond the traditional approaches of security in computer science. Therefore, the psychological decision-making and cognitive biases involved in SE-based cyberattacks are ongoing concerns for ML-based approaches [98]. To achieve a higher detection rate, the researcher should explore the linguistic features of SE and integrate cognitive and psychological factors into ML-based approaches.

4. Discussion

The manipulation of human behavior and emotions in cyberattacks presents an unknown and challenging variable for security experts. The prime reason behind this variable can generally be characterized as culture. Culture can play an important role in influencing human behavior, beliefs, morals, decisions, and attitudes [99,100]. Even with technological advancements in security, humans can be exploited for their vulnerabilities. Based on Table 9, it can be concluded that the most recent publications agree that raising the awareness of cybersecurity through training and education is essential. Such awareness can help in decreasing cyberattacks based on SE. On the other hand, some studies [101–103] highlight that despite appropriate training and policies, human vulnerabilities can still be exploited via SE attacks. For example, not every individual working in an organization has basic knowledge of computer security. Training an employee with no prior computer knowledge can be costly, time-consuming, and may not be very effective [104]. Such employees are highly vulnerable to several phishing-based SE attacks. Another study [105] concluded that an individual's self-efficiency also plays an important role in avoiding SE attacks. The connection between an individual's self-efficiency and vulnerability to SE attacks is further explored in some research publications.

Papers [1,4,106–108] categorized vulnerabilities based on behaviors displayed by an individual on social media and in daily life. The papers then highlighted the behavioral traits that can be targeted by SE attacks. The studies also highlighted that social psychology can be used to reinforce security policies. The mapping in Table 8 also provides an abstract view of how behaviors can be used to identify exploitable vulnerabilities for SE attacks. In addition, Table 8 maps an extensive range of behavioral human traits on specific SE-based attacks, whereas most of the recent papers focused on mapping phishing-based attacks on human behavior. It can be observed that there is a clear gap between sophisticated SE attacks and existing countermeasures. The lack of effective approaches to prevent and avoid SE-based cyberattacks is an ongoing challenge for security experts. To efficiently counter these SE attacks, multidimensional countermeasures based on human vulnerabilities and technical components are necessary. ML-based approaches show high efficiency in countering SE-based cyberattacks; however, there is still a need for further improvement in ML-based methods, as discussed in Section 5. With that in mind, this paper provides a broad outline of the components interconnecting human vulnerabilities with SE-based cyberattacks. Additionally, this research can provide readers and researchers with an appropriate understanding of the available countermeasures, including ML-based approaches against cyberattacks based on SE. Such understanding can assist organizations in defining policies to proficiently counter such attacks. Understandably, several research papers have highlighted the impacts, approaches, and motivations behind SE attacks. However, to the best of our knowledge, there is a lack of research on mapping human behavior on specific SE attacks. The association of human behavior to specific SE attacks may hold the key to enhanced countermeasures against such attacks. So far, a handful of researchers have worked on associating behavioral approaches to explicit SE-based attack vulnerabilities. The association between attacks and human vulnerabilities presented in Table 8 can play an integral role in improving the approaches to counter SE attacks.

This paper presents a theoretical understanding of cyberattacks based on SE. The key concern of evaluating human vulnerabilities in cybersecurity is the viewpoint of the observer. A cybersecurity professional might have a different perspective on human vulnerabilities as compared to a person with a background in human psychology. Most of the influence methods and human vulnerabilities discussed in the paper have been used to conduct cyberattacks; however, a few of them are based on theoretical concepts and case studies. Such theoretical concepts need further studies and testing to improve the understanding of human behavior under SE attacks. The analysis of human vulnerabilities based on theoretical concepts can be considered a limitation of this paper. Human behavior and vulnerabilities are also dependent on factors such as the working environment, age, educational background, work experience, etc. On the other hand, theoretical analysis

plays a key role in providing grounds to improve and conduct further studies. In ML-based approaches, in particular, a better understanding of human behavior can lead to improved models of NLP- and DNN-based countermeasures. Despite the existing countermeasures and efforts, the World Economic Forum emphasized that SE-based cyberattacks are among the most concerning security aspects for organizations in 2022 [109]. Table 10 provides a summary of the topics covered in the paper.

Table 10. Summary of topics covered in the paper.

Section	Section Summary
Section 1	In this section, readers are introduced to the idea and working of SE attacks. To highlight the importance of cyberattacks based on SE, some of the most recent and prominent attacks are presented to the readers in the section.
Section 2	This section covers the most common types of SE attacks and their methodologies, i.e., phishing, dumpster diving, scareware, water hole, reverse SE.
Section 3	In this section, the methods to influence or exploit human vulnerabilities to conduct SE attacks are discussed. The section also provides the interconnection between SE attacks, methods of influence, and human vulnerabilities. The mapping of SE attacks, methods of influence, and human vulnerabilities plays a key role in understanding and countering cyberattacks based on SE.
Section 4	This section presents the reader with recent research on methods to counter SE attacks. The section also provides readers with an elaborate understanding of different countermeasures proposed to counter SE attacks, including the most prominent ML-based methods. The section also covers existing concerns about ML-based countermeasures.
Section 5	In this section, concerns over recently proposed methods to counter SE attacks are discussed. The section also emphasizes the need for a multidimensional approach to counter SE attacks. The limitations of the paper are also highlighted in this section.

5. Conclusions

Cyberattacks based on SE are major threats to organizations and individuals. As highlighted in the paper, human vulnerabilities play a key role in initiating SE attack cycles. Recent SE attacks have highlighted that exploiting the human factor to conduct a cyberattack is a highly efficient approach. As a result, security through technology is no longer the sole solution for an organization or an individual. In organizations, the responsibility of mitigating cybersecurity threats is a shared task between the IT department and every employee. One approach to reducing the exploitation of human vulnerabilities is to improve security awareness. However, only providing awareness against SE-based cyberattacks is not sufficient. At the organizational level, a systematic approach to identifying vulnerable employees can play a significant role in minimizing cybersecurity threats, i.e., by analyzing the security awareness of the workforce, maintaining effective means of communication (regarding cyberattack threats), routine system updates, and appropriate security infrastructure. A cybersecurity approach involving human factors is a step towards an efficient, robust, and resilient cybersecurity framework. This research paper can provide grounds for analyzing and constructing a security framework involving human factors. The paper provides systematic knowledge of SE attacks, methods of attacks, human vulnerabilities, mapping of attacks on human vulnerabilities, and recent research on mitigating SE attacks; thus, this paper provides a structural perspective to help understand how SE attacks work. For future work, we plan to design a systematic flow of steps to follow in case of a SE-based cyberattack or threat. The flow will also help in identifying vulnerabilities in implemented security infrastructure and employee awareness of SE cyberattacks.

Author Contributions: M.A.S. (Murtaza Ahmed Siddiqi), W.P. and M.A.S. (Moquddam A. Siddiqi) wrote the paper and conducted the research. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) NRF-2022R1A2C1011774 and the 2022 Yeungnam University research grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no conflict of interest to report regarding the present study.

Abbreviations

The following abbreviations are used in this manuscript:

SE	social engineering
BEC	business email compromise
RAT	remote access Trojan
ML	machine learning
DSD	distributed spam distraction
GANs	generative adversarial networks
ANNs	artificial neural networks
TPB	theory of planned behavior
IDT	interpersonal deception theory
SMS	short message service
NLP	natural language processing
DL	deep learning
DNN	deep neural network
LSTM	long short-term memory
RL	reinforcement learning
CRM	cyber-resilient mechanism

References

1. Abroshan, H.; Devos, J.; Poels, G.; Laermans, E. Phishing happens beyond technology: The effects of human behaviors and demographics on each step of a phishing process. *IEEE Access* **2021**, *9*, 44928–44949. [CrossRef]
2. Siddiqi, M.A.; Mugheri, A.; Oad, K. Advanced persistent threats defense techniques: A review. *Pak. J. Comput. Inf. Syst.* **2016**, *1*, 53–65.
3. Wang, Z.; Zhu, H.; Sun, L. Social engineering in cybersecurity: Effect mechanisms, human vulnerabilities and attack methods. *IEEE Access* **2021**, *9*, 11895–11910. [CrossRef]
4. Albladi, S.M.; Weir, G.R.S. Predicting individuals' vulnerability to social engineering in social networks. *Cybersecurity* **2020**, *3*, 7. [CrossRef]
5. Saudi Aramco Confirms Data Leak after Reported Cyber Ransom. Available online: https://www.bloomberg.com/news/articles/2021-07-21/saudi-aramco-confirms-data-leak-after-reported-cyber-extortion (accessed on 6 August 2021).
6. Marriott Discloses Data Breach Possibly Affecting over 5 Million Customers. Available online: https://edition.cnn.com/2020/04/01/business/marriott-hack-trnd/index.html (accessed on 10 August 2021).
7. Marriott Data Breach FAQ: How Did It Happen and What Was the Impact? Available online: https://www.csoonline.com/article/3441220/marriott-data-breach-faq-how-did-it-happen-and-what-was-the-impact.html (accessed on 7 July 2021).
8. Hughes-Larteya, K.; Li, M.; Botchey, F.E.; Qin, Z. Human factor, a critical weak point in the information security of an organization's Internet of things. *Heliyon* **2021**, *7*, 6522–6535. [CrossRef]
9. Siddiqi, M.A.; Ghani, N. Critical analysis on advanced persistent threats. *Int. J. Comput. Appl.* **2016**, *141*, 46–50.
10. Americans Lost $29.8 Billion to Phone Scams Alone over the Past Year. Available online: https://www.cnbc.com/2021/06/29/americans-lost-billions-of-dollars-to-phone-scams-over-the-past-year.html (accessed on 8 August 2021).
11. Widespread Credential Phishing Campaign Abuses Open Redirector Links. Available online: https://www.microsoft.com/security/blog/2021/08/26/widespread-credential-phishing-campaign-abuses-open-redirector-links/ (accessed on 11 October 2021).
12. Twitter Hack: Staff Tricked by Phone Spear-Phishing Scam. Available online: https://www.bbc.com/news/technology-53607374 (accessed on 10 August 2021).
13. Shark Tank Host Barbara Corcoran Loses $380,000 in Email Scam. Available online: https://www.forbes.com/sites/rachelsandler/2020/02/27/shark-tank-host-barbara-corcoran-loses-380000-in-email-scam/?sh=73b0935a511a (accessed on 7 October 2021).
14. Toyota Parts Supplier Hit by $37 Million Email Scam. Available online: https://www.forbes.com/sites/leemathews/2019/09/06/toyota-parts-supplier-hit-by-37-million-email-scam/?sh=733a2c6e5856 (accessed on 7 October 2021).
15. Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. Available online: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 (accessed on 11 October 2021).
16. Google and Facebook Duped in Huge 'Scam'. Available online: https://www.bbc.com/news/technology-39744007 (accessed on 15 October 2021).
17. Facebook and Google Were Conned out of $100m in Phishing Scheme. Available online: https://www.theguardian.com/technology/2017/apr/28/facebook-google-conned-100m-phishing-scheme (accessed on 12 October 2021).
18. Govindankutty, M.S. Is human error paving way to cyber security? *Int. Res. J. Eng. Technol.* **2021**, *8*, 4174–4178.

19. Siddiqi, M.A.; Pak, W. Optimizing filter-based feature selection method flow for intrusion detection system. *Electronics* **2020**, *9*, 2114. [CrossRef]
20. Human Cyber Risk—The First Line of Defense. Available online: https://www.aig.com/about-us/knowledge-insights/human-cyber-risk-the-first-line-of-defense (accessed on 12 August 2021).
21. Pfeffel, K.; Ulsamer, P.; Müller, N. Where the user does look when reading phishing mails—An eye-tracking study. In Proceedings of the International Conference on Human-Computer Interaction (HCII), Orlando, FL, USA, 26–31 July 2019.
22. Gratian, M.; Bandi, S.; Cukier, M.; Dykstra, J.; Ginther, A. Correlating human traits and cyber security behavior intentions. *Comput. Secur.* **2018**, *73*, 345–358. [CrossRef]
23. Dhillon, G.; Talib, Y.A.; Picoto, W.N. The mediating role of psychological empowerment in information security compliance intentions. *J. Assoc. Inf. Syst.* **2020**, *21*, 152–174. [CrossRef]
24. 12 Types of Phishing Attacks and How to Identify Them. Available online: https://securityscorecard.com/blog/types-of-phishing-attacks-and-how-to-identify-them (accessed on 16 August 2021).
25. Social Engineering Attack Escalation. Available online: https://appriver.com/blog/201708social-engineering-attack-escalation (accessed on 11 September 2021).
26. Cross, M. *Social Media Security: Leveraging Social Networking While Mitigating Risk*, 1st ed.; Syngress Publishing: Rockland, MA, USA, 2014; pp. 161–191.
27. Grover, A.; Berghel, H.; Cobb, D. *Advances in Computers*; Academic Press: Burlington, MA, USA, 2011; Volume 83, pp. 1–50.
28. Malin, C.H.; Gudaitis, T.; Holt, T.J.; Kilger, M. Phishing, Watering Holes, and Scareware. In *Deception in the Digital Age: Exploiting and Defending Human Targets through Computer-Mediated Communications*, 1st ed.; Academic Press: Burlington, MA, USA, 2017; pp. 149–166.
29. Malin, C.H.; Gudaitis, T.; Holt, T.J.; Kilger, M. Viral Influence: Deceptive Computing Attacks through Persuasion. In *Deception in the Digital Age: Exploiting and Defending Human Targets through Computer-Mediated Communications*, 1st ed.; Academic Press: Burlington, MA, USA, 2017; pp. 77–124.
30. Social Engineering: What You Can Do to Avoid Being a Victim. Available online: https://www.g2.com/articles/social-engineering (accessed on 26 August 2021).
31. Social Engineering Technique: The Watering Hole Attack. Available online: https://medium.com/@thefoursec/social-engineering-technique-the-watering-hole-attack-9ee3d2ca17b4 (accessed on 26 August 2021).
32. Shi, Z.R.; Schlenker, A.; Hay, B.; Bittleston, D.; Gao, S.; Peterson, E.; Trezza, J.; Fang, F. Draining the water hole: Mitigating social engineering attacks with cybertweak. In Proceedings of the Thirty-Second Innovative Applications of Artificial Intelligence Conference (IAAI-20), New York, NY, USA, 9–11 February 2020.
33. Parthy, P.P.; Rajendran, G. Identification and prevention of social engineering attacks on an enterprise. In Proceedings of the International Carnahan Conference on Security Technology (ICCST), Chennai, India, 1–3 October 2019.
34. Irani, D.; Balduzzi, M.; Balzarotti, D.; Kirda, E.; Pu, C. Reverse social engineering attacks in online social networks. In Proceedings of the Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA), Berlin, Germany, 7–8 July 2011.
35. Albahar, M.; Almalki, J. Deepfakes: Threats and countermeasures systematic review. *J. Theor. Appl. Inf. Technol.* **2019**, *97*, 3242–3250.
36. Chi, H.; Maduakor, U.; Alo, R.; Williams, E. Integrating deepfake detection into cybersecurity curriculum. In Proceedings of the Future Technologies Conference (FTC), Virtual Platform, San Francisco, CA, USA, 5–6 November 2020.
37. Gass, R.H. *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed.; Elsevier: Houston, TX, USA, 2015; pp. 348–354.
38. Myers, D. *Social Psychology*, 10th ed.; Mc Graw Hill: New York, NY, USA, 2012; pp. 266–304.
39. Mamedova, N.; Urintsov, A.; Staroverova, O.; Ivanov, E.; Galahov, D. Social engineering in the context of ensuring information security. In Proceedings of the Current Issues of Linguistics and Didactics: The Interdisciplinary Approach in Humanities and Social Sciences (CILDIAH), Volgograd, Russia, 23–28 April 2019.
40. Foa, E.B.; Foa, U.G. *Handbook of Social Resource Theory*, 2012th ed.; Springer: New York, NY, USA, 2012; pp. 15–32.
41. Wang, Z.; Zhu, H.; Liu, P.; Sun, L. Social engineering in cybersecurity: A domain ontology and knowledge graph application examples. *Cybersecurity* **2021**, *4*, 31. [CrossRef]
42. Collins, N.L.; Miller, L.C. Self-disclosure and liking: A meta-analytic review. *Psychol. Bull.* **1994**, *116*, 457–475. [CrossRef] [PubMed]
43. Hacking Human Psychology: Understanding Social Engineering Hacks. Available online: https://www.relativity.com/blog/hacking-human-psychology-understanding-social-engineering/ (accessed on 2 September 2021).
44. Ferreira, A.; Coventry, L.; Lenzini, G. Principles of persuasion in social engineering and their use in phishing. In Proceedings of the Name of the Human Aspects of Information Security, Privacy, and Trust (HAS), Los Angeles, CA, USA, 2–7 August 2015.
45. Cialdini, R.B. *Influence: The Psychology of Persuasion*, revised ed.; Harper Business:New York, NY, USA, 2006; pp. 1–12.
46. Norton, M.; Frost, J.; Ariely, D. Less is more: The lure of ambiguity, or why familiarity breeds contempt. *J. Pers. Soc. Psychol.* **2007**, *92*, 97–105. [CrossRef]
47. Guadagno, R.E.; Cialdini, R.B. *The Social Net: The Social Psychology of the Internet*, 1st ed.; Oxford University Press: New York, NY, USA, 2009; pp. 91–113.
48. Robert, O.; Timothy, B. Distraction increases yielding to propaganda by inhibiting counterarguing. *J. Pers. Soc. Psychol.* **1970**, *15*, 344–358.

49. Siadati, H.; Nguyena, T.; Gupta, P.; Jakobsson, M.; Memon, N. Mind your SMSes: Mitigating social engineering in second factor authentication. *Comput. Secur.* **2017**, *65*, 14–28. [CrossRef]
50. Priester, J.; Petty, R. Source attributions and persuasion: Perceived honesty as a determinant of message scrutiny. *Pers. Soc. Psychol. Bull.* **1995**, *21*, 637–654. [CrossRef]
51. Mitnick, K.D.; Simon, W.L.; Wozniak, S. *The Art of Deception: Controlling the Human Element of Security*, 1st ed.; Wiley: Hoboken, NJ, USA, 2003; pp. 59–71.
52. Ajzen, I. The theory of planned behavior: Frequently asked questions. *Hum. Behav. Emerg. Technol.* **2002**, *2*, 314–324. [CrossRef]
53. Gulenko, I. Social against social engineering: Concept and development of a Facebook application to raise security and risk awareness. *Inf. Manag. Comput. Secur.* **2013**, *21*, 91–101. [CrossRef]
54. Leary, M.R. *Self-Presentation Impression Management And Interpersonal Behavior*, 1st ed.; Routledge: London, UK, 1996; pp. 25–35.
55. Montañez, R.; Golob, E.; Xu, S. Human cognition through the lens of social engineering cyberattacks. *Front. Psychol.* **2020**, *11*, 1755–1773. [CrossRef]
56. Metzger, M.J.; Hartsell, E.H.; Flanagin, A.J. Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Commun. Res.* **2020**, *47*, 3–28. [CrossRef]
57. Social Engineering as a Threat to Societies: The Cambridge Analytica Case. Available online: https://thestrategybridge.org/the-bridge/2018/7/18/social-engineering-as-a-threat-to-societies-the-cambridge-analytica-case (accessed on 20 September 2021).
58. Lahcen, R.A.M.; Caulkins, B.; Mohapatra, R.; Kumar, M. Review and insight on the behavioral aspects of cybersecurity. *Cybersecurity* **2020**, *3*, 10. [CrossRef]
59. You, L.; Lee, Y.H. The bystander effect in cyberbullying on social network sites: Anonymity, group size, and intervention intentions. *Telemat. Inform.* **2019**, *45*, 101284. [CrossRef]
60. Sherchan, W.; Nepal, S.; Paris, C. A survey of trust in social networks. *ACM Comput. Surv.* **2013**, *45*, 1–33. [CrossRef]
61. Molodetska, K.; Solonnikov, V.; Voitko, O.; Humeniuk, I.; Matsko, O.; Samchyshyn, O. Counteraction to information influence in social networking services by means of fuzzy logic system. *Int. J. Electr. Comput. Eng.* **2021**, *11*, 2490–2499. [CrossRef]
62. Campbell, C.C. Solutions for counteracting human deception in social engineering attacks. *Inf. Technol. People* **2019**, *32*, 1130–1152. [CrossRef]
63. Burgoon, J.K.; Buller, D.B. Interpersonal deception theory. *Commun. Theory* **1996**, *6*, 203–242. [CrossRef]
64. Handoko, H.; Putri, D.A.W. Threat language: Cognitive exploitation in social engineering. In Proceedings of the International Conference on Social Sciences, Humanities, Economics and Law (ICSSHEL), Padang, Indonesia, 5–6 September 2018.
65. Dorr, B.J.; Bhatia, A.; Dalton, A.; Mather, B.; Hebenstreit, B.; Santhanam, S.; Cheng, Z.; Shaikh, S.; Zemel, A.; Strzalkowski, T. Detecting asks in SE attacks: Impact of linguistic and structural knowledge. *arXiv* **2020**, arXiv:2002.10931.
66. Rodríguez-Priego, N.; Bavel, R.V.; Vila, J.; Briggs, P. Framing effects on online security behavior. *Front. Psychol.* **2020**, *11*, 2833–2844. [CrossRef]
67. Yasin, A.; Fatima, R.; Liu, L.; Wang, J.; Ali, R.; Wei, Z. Understanding and deciphering of social engineering attack scenarios. *Secur. Priv.* **2021**, *4*, e161. [CrossRef]
68. Handoko, H.; Putri, D.A.W.; Sastra, G.; Revita, I. The language of social engineering: From persuasion to deception. In Proceedings of the 2nd International Seminar on Linguistics (ISL), Padang, West Sumatra, Indonesia, 12–13 August 2015.
69. Comment of NLP and Social Engineering Hacking the Human Mind Article. Available online: https://www.hellboundhackers.org/articles/read-article.php?article_id=8%78 (accessed on 9 September 2021).
70. Alkhaiwani, A.H.; Almalki, G.A. Saudi human awareness needs. A survey in how human causes errors and mistakes leads to leak confidential data with proposed solutions in Saudi Arabia. In Proceedings of the National Computing Colleges Conference (NCCC), Taif, Saudi Arabia, 27–28 March 2021.
71. Spear Phishing: Top Threats and Trends. Available online: https://assets.barracuda.com/assets/docs/dms/spear-phishing_report_vol6.pdf (accessed on 23 May 2022).
72. Sushruth, V.; Reddy, K.R.; Chandavarkar, B.R. Social engineering attacks during the COVID-19 pandemic. *SN Comput. Sci.* **2021**, *2*, 78. [CrossRef]
73. Washo, A.H. An interdisciplinary view of social engineering: A call to action for research. *Comput. Hum. Behav. Rep.* **2021**, *4*, 100126. [CrossRef]
74. Alsulami, M.H.; Alharbi, F.D.; Almutairi, H.M.; Almutairi, B.S.; Alotaibi, M.M.; Alanzi, M.E.; Alotaibi, K.G.; Alharthi, S.S. Measuring awareness of social engineering in the educational sector in the kingdom of Saudi Arabia. *Information* **2021**, *12*, 208. [CrossRef]
75. Aldawood, H.; Skinner, G. Reviewing cyber security social engineering training and awareness programs—Pitfalls and ongoing issues. *Future Internet* **2019**, *11*, 73. [CrossRef]
76. Fan, W.; Lwakatare, K.; Rong, R. Social engineering: I-E based model of human weakness for attack and defense investigations.*Int. J. Comput. Netw. Inf. Secur.* **2017**, *9*, 1–11. [CrossRef]
77. Bakhshi, T. Social engineering: Revisiting end-user awareness and susceptibility to classic attack vectors. In Proceedings of the 13th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 27–28 December2017.
78. Sillanpää, M.; Hautamäki, J. Social engineering intrusion: A case study. In Proceedings of the 11th International Conference on Advances in Information Technology (IAIT), Bangkok, Thailand, 1–3 July 2020.

79. What Is Social Engineering? A Definition + Techniques to Watch for. Available online: https://us.norton.com/internetsecurity-emerging-threats-what-is-social-engineering.html (accessed on 16 September 2021).
80. What Is Social Engineering and How to Prevent It. Available online: https://www.avast.com/c-social-engineering (accessed on 16 September 2021).
81. Network Intrusion Detection Techniques Using Machine Learning. Available online: https://www.researchgate.net/publication/349392282_Network_Intrusion_Detection_Techniques_using_Machine_Learning (accessed on 20 May 2022).
82. Here's How Cyber Threats Are Being Detected Using Deep Learning. Available online: https://techhq.com/2021/09/heres-how-cyber-threats-are-being-detected-using-deep-learning (accessed on 4 November 2021).
83. Peng, T.; Harris, I.; Sawa, Y. Detecting phishing attacks using natural language processing and machine learning. In Proceedings of the IEEE 12th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 31 January–2 February 2018.
84. Tsinganos, N.; Sakellariou, G.; Fouliras, P.; Mavridis, I. Towards an automated recognition system for chat-based social engineering attacks in enterprise environments. In Proceedings of the 13th International Conference on Availability, Reliability and Security (ICARS), Hamburg, Germany, 27–30 August 2018.
85. Siddiqi, M.; Pak, W. An agile approach to identify single and hybrid normalization for enhancing machine learning based network intrusion detection. *IEEE Access* **2021**, *9*, 137494–137513. [CrossRef]
86. Lansley, M.; Polatidis, N.; Kapetanakis, S.; Amin, K.; Samakovitis, G.; Petridis, M. Seen the villains: Detecting social engineering attacks using case-based reasoning and deep learning. In Proceedings of the Twenty-Seventh International Conference on Case-Based Reasoning (ICCBR), Otzenhausen, Germany, 28–30 September 2019.
87. Ozcan, A.; Catal, C.; Donmez, E.; Senturk, B. A hybrid DNN–LSTM model for detecting phishing URLs. *Neural Comput. Appl.* **2021**, *9*, 1–17. [CrossRef]
88. Vinayakumar, R.; Alazab, M.; Jolfaei, A.; Soman, K.P.; Poornachandran, P. Ransomware Triage Using Deep Learning: Twitter as a Case Study. In Proceedings of the Cybersecurity and Cyberforensics Conference (CCC), Melbourne, Australia, 8–9 May 2019.
89. Vinayakumar, R.; Soman, K.P.; Poornachandran, P.; Mohan, V.S.; Kumar, A.D. ScaleNet: Scalable and Hybrid Framework for Cyber Threat Situational Awareness Based on DNS, URL, and Email Data Analysis. *J. Cyber Secur. Mobil.* **2019**, *8*, 189–240. [CrossRef]
90. Ketha, S.; Srinivasan, S.; Ravi, V.; Soman, K.P. Deep Learning Approach for Intelligent Named Entity Recognition of Cyber Security. In Proceedings of the the 5th International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS'19), Trivandrum, India, 18–21 December 2019.
91. Huang, Y.; Huang, L.; Zhu, Q. Reinforcement learning for feedback-enabled cyber resilience. *Annu. Rev. Control* **2022**, *23*, 273–295. [CrossRef]
92. Bland, J.A.; Petty, M.D.; Whitaker, T.S.; Maxwell, K.P.; Cantrell, W.A. Machine learning cyberattack and defense strategies. *Comput. Secur.* **2020**, *92*, 101738. [CrossRef]
93. Rawindaran, N.; Jayal, A.; Prakash, E.; Hewage, C. Cost benefits of using machine learning features in NIDS for cyber security in UK small medium enterprises (SME). *Future Internet* **2021**, *13*, 186. [CrossRef]
94. Sallouma, S.; Gaber, T.; Vadera, S.; Shaalan, K. Phishing email detection using natural language processing techniques: A literature survey. *Procedia Comput. Sci.* **2021**, *189*, 19–28. [CrossRef]
95. Fang, Y.; Zhang, C.; Huang, C.; Liu, L.; Yang, Y. Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism. *IEEE Access* **2019**, *7*, 56329–56340. [CrossRef]
96. Gutierrez, C.N.; Kim, T.; Corte, R.D.; Avery, J.; Goldwasser, D.; Cinque, M.; Bagchi, S. Learning from the ones that got away: Detecting new forms of phishing attacks. *IEEE Trans. Dependable Secure Comput.* **2018**, *15*, 988–1001. [CrossRef]
97. Repke, T.; Krestel, R. Bringing back structure to free text email conversations with recurrent neural networks. In Proceedings of the European Conference on Information Retrieval (ECIR), Grenoble, France, 25–29 March 2018.
98. Lan, Y. Chat-oriented social engineering attack detection using attention-based Bi-LSTM and CNN. In Proceedings of the 2nd International Conference on Computing and Data Science (CDS), Stanford, CA, USA, 28–30 January 2021.
99. Cano, J. The human factor in information security: The weakest link or the most fatigued? *Inf. Syst. Audit. Control Assoc.* **2019**, *5*, 1–7.
100. Bian, J.; Li, L.; Sun, J.; Deng, J.; Li, Q.; Zhang, X.; Yan, L. The influence of self-relevance and cultural values on moral orientation. *Front. Psychol.* **2019**, *10*, 292. [CrossRef] [PubMed]
101. Bada, M.; Sasse, A.M.; Nurse, J. Cyber security awareness campaigns: why do they fail to change behavior? In Proceedings of the International Conference on Cyber Security for Sustainable Society (ICCSSS), Coventry, UK, 26 February 2015.
102. Mortan, E.A. *Cyber Security and Supply Chain Management: Risk, Challenges, and Solutions*, 1st ed.; World Scientific Publishing: Singapore, 2021; pp. 62–63.
103. Alkhalil, Z.; Hewage, C.; Nawaf, L.; Khan, I. Phishing attacks: A recent comprehensive study and a new anatomy. *Front. Comput. Sci.* **2021**, *3*, 563060. [CrossRef]
104. Dodge, R.; Carver, C.; Ferguson, A.J. Phishing for user security awareness. *Comput. Secur.* **2007**, *26*, 73–80. [CrossRef]
105. Arachchilage, N.A.G.; Love, S. Security awareness of computer users: A phishing threat avoidance perspective. *Comput. Hum. Behav.* **2014**, *38*, 304–312. [CrossRef]
106. Ani, U.D.; He, H.; Tiwari, A. Human factor security: Evaluating the cybersecurity capacity of the industrial workforce. *J. Syst. Inf. Technol.* **2019**, *21*, 2–35. [CrossRef]

107. Sibrian, J. Sensitive Data? Now That's a Catch! the Psychology of Phishing, Chapter 3-Sensitive Data? Now That's a Catch! The Psychology of Phishing. Bachelor's Thesis, Harvard College, Cambridge, MA, USA, 2021; pp. 17–28.
108. Kabay, M.E.; Robertson, B.; Akella, M.; Lang, D.T. Chapter 50-Using Social Psychology to Implement Security Policies. In *Computer Security Handbook*, 6th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012; pp. 50.1–50.25.
109. What You Need to Know about Cybersecurity in 2022. Available online: https://www.weforum.org/agenda/2022/01/cybersecurity-2022-global-outlook (accessed on 4 April 2022).