Mind your SMSes: Mitigating Social Engineering in Second Factor Authentication

Hossein Siadati
New York University
hossein@nyu.edu

Toan Nguyen
New York University
toan.v.nguyen@nyu.edu

Payas Gupta
New York University
payasgupta@nyu.edu

Markus Jakobsson
Agari
mjakobsson@agari.com

Nasir Memon
New York University
memon@nyu.edu

ABSTRACT

SMS-based second factor authentication is a cornerstone for many service providers, ranging from email service providers and social networks to financial institutions and online marketplaces. Attackers are not slow to capitalize on the vulnerabilities of this mechanism, using social engineering techniques to coerce users to forward authentication codes. We demonstrate one social engineering attack for which we experimentally obtained a 50% success rate against Google's SMS-based authentication. At the heart of the problem is the messaging associated with the authentication code, and how this must not have been developed with security against social engineering in mind. Pursuing a top-down methodology, we generate alternative messages and experimentally test these against an array of social engineering attempts. Our most robust messaging approach reduces the success of the most effective social engineering attack to 8%, or a sixth of its success against Google's standard second factor verification code messages.

Keywords

Phishing, 2-factor authentication, 2-step verification, SMS, verification code forwarding attack, human factors, warning

1. INTRODUCTION

It is increasingly recognized that the human factor constitutes the weakest link for the security of individuals and organizations. In the last ten years, social engineering has become a staple in the arsenal of criminals, whether petty thieves or nation state backed actors. Social engineering is used to commit fraud; to trick users to install malware; to willingly share sensitive corporate information; and to perform transfers of large sums of money [31].

A common use of social engineering, in the form of phishing attacks, is to steal user credentials. During the last few years, it has also increasingly been used to defeat out-of-band authentication methods. To authenticate a user initiating a critical action—e.g., a password reset, user account setup or a fund transfer—an out-of-band authentication system generates a one time random secret (also known as "Verification Code" or "Security Code") and sends this to the user by a separate channel (for example, on a mobile phone using a number on file for the user). On receiving the code, the user relays it back to the service provider using the original channel (for example, the web channel which was used for password reset). If the user-entered code matches what was sent, the user is authenticated and the requested action is completed. The assumption is that the second channel will be hard for an attacker to access even if the attacker has some knowledge of user credentials.

The main objective of out-of-band authentication is to build a second layer of authentication that can be used even when a user's credentials are fully or partially known to an attacker. Knowledge of user credentials is a realistic threat given the frequency of large-scale data breaches [30], the success rates of targeted phishing attacks and the prevalence of malware [34, 53, 55].

Social engineering attacks against out-of-band authentication essentially trick the user to relay the verification code to the attacker. Symantec has reported an increase in the number of such attacks [54] and previous work shows that a social engineering attack on out-of-band authentication methods can have success rates of 25% [51]. Balduzzi et al. has found 22 instances of such attacks using a mobile honeypot in China [7]. Such attacks, called *Verification Code Forwarding Attack* (VCFA), begin with the attacker triggering a verification code to be sent to a user from a service provider. Then by sending an appropriate message to the user, the attacker dupes the user to send the code to him. This is typically done by posing as the service provider. Once the attacker receives the code, he can complete the action on the user's account. Figure 1 shows the steps of a VCFA. VCFA is particularly serious since many service providers use SMS-based authentication in contexts where there is a greater than normal risk, whether due to a suspected account takeover, a large-value transaction request, or a combination of these.

In the above scenario, there are two messages that can be tailored. One is the message the service provider sends to the user (step 2) to complete the out-of-band authentication which we call the *verification* message and the second by the attacker (step 3), which we call the *attack* message. When you consider verification messages, different service providers appear to have made different design choices. For example Google's SMS-based second factor authentication (2FA) epitomizes simplicity: "Your Google verification code is 654722". Google is in good company with Microsoft ("Microsoft account verification code: 1505"), Wells Fargo ("Use Wells Fargo verification code 817882") and many others. Although these messages are simple, we identify social engineering pitches capable of coercing 50% of users to share this code by conducting a systematic large scale user study. In comparison, SMS-based verification messages containing a warning— such as that used by AT&T ("AT&T Free Msg: Your Temporary password is 987876. If you did not request a temporary password, call 1-800-ATT-2020. We'll never contact you to ask for this password.") are slightly more robust against attack, with a 34% success rate for our most effective attack.

In this paper we investigate how social engineering attacks, specif-

Published in Computer & Security 65 (2017) 14–28
Link: https://doi.org/10.1016/j.cose.2016.09.009

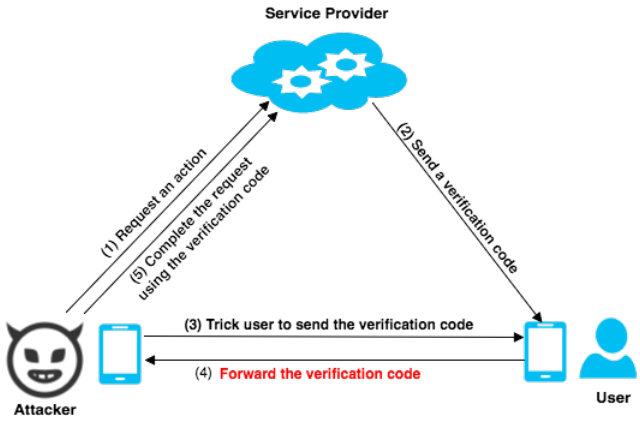


Figure 1: Verification Code Forwarding Attack scenario. Step 1: The attacker triggers the delivery of a verification code (e.g., using password reset feature). Step 3: He tricks the user to forward the verification code to him. Step 5: Upon receiving the code from the victim, the attacker completes the password reset request and takes over the user's account.

ically VCFA on SMS based out-of-band authentication could be mitigated by improving the messages used during authentication process. This paper mainly focuses on designing better verification messages, and we test our design against a variety of systematically generated and tested attack messages. In order to perform a systematic study of the effectiveness of verification messages in mitigating social engineering attacks, we first identify what attack messages that seem most likely to coerce users. Then, we turn to developing and testing a collection of verification messaging alternatives, and to enunciate the principles underlying the most robust approaches.

We find that when a warning "Please ignore this message if you did not request a code" *precedes* the authentication code, stands up to social engineering attempts[1] better than any other tested method, resulting in a susceptibility to attack of just 8%.

Our main contribution lies in the notable reduction of the expected yield as seen by a cunning attacker employing social engineering to steal verification codes. However, of potential independent interest are the light-weight experimental techniques we developed and used to identify likely contenders— whether for most successful attacker approach or most robust verification message— and for a realistic testing of the exact yields of various attacks, when applied to various verification messages. These light-weight tools allow us to concurrently argue about hypothetical approaches in two dimensions, corresponding to what the good guys should do in light of what the attackers might try, and to select the verification message with the greatest robustness against attack.

2. DESIGNING ATTACK MESSAGES

As can be seen from the previous section, VCFA does not require much knowledge about the target. The attacker only needs to know the email address and phone number associated with the account, and both can be easy to acquire. Data breaches are one source of such information. For example, the Anthem breach [41] involved the loss of email addresses and phone numbers, as well as other personal information, of 80 million customers. Malware is another means of collecting such information. Specifically, smartphone malware such as Asacub [34], Ackposts [53], and Godwon [55] are known for harvesting contact lists among other valuable information. Moreover, vulnerabilities in messaging apps have been shown to enable attackers to harvest large amounts of contact information in a short interval of time [35, 40, 29].

Given the ease with which an attacker can gain access to information to launch an attack, the main challenge lies in tricking users to send their verification codes. However, it turns out that users are quite susceptible and easy to deceive. For example, previous work has shown that the following message sent by the attacker soon after a victim receives a verification code due to action triggered by the attacker, achieves a success rate of 25% for the attacker [51]:

> Please verify that your phone is still associated with your Gmail account by replying to this message with the code we have just sent to you.

One reason for the ease of deceiving users is due to the shortcomings of the messages used by service providers. This begs the question - can better messages be designed that may be more robust against attacks? However, any such verification message must resist different types of attack messages. Hence in the rest of this section we first study different strategies an attacker might apply.

Social engineering messages should justify why the user has received an unwanted verification code, and the reason that she should send the code back. To study the extent to which VCFA can be successful, we systematically generated a number of attack messages using three categories of storylines:

1. **Compliance:** Companies contact customers to update them about new policies and terms of service. Users expect to be contacted by service providers if they misuse an online service.
2. **Service:** Users are persuaded by promotions from service providers (e.g., larger email inbox), or by communication messages from their friends offered via service provider (e.g., voicemail).
3. **Security:** Companies usually communicate with users concerning the security of their account. Users are asked to change their passwords, notified on changes in security settings of their accounts, and contacted if an abnormal activity, such as login from a new computer, is observed. Therefore, users expect to receive such messages from service providers.

Further, previous work has studied persuasion tools for phishing [3, 44, 13]. Abraham et al. [3] have classified phishing persuasion techniques to "Curiosity, empathy, and excitement", "Fear", and "Greed". We have used these classes of persuasion techniques for the *tone* of phishing messages in the previously listed classes.

Combining storylines with different classes of persuasion, we generated 16 different attack messages (See Table 1) to test attacks on SMS-based out-of-band authentication systems. We first tested them against the Google SMS-2FA system (using a field experiment explained in Section 4) and found significant variation in success rates for the attacker. For example, we found a variation of VCFA attack which can yield a success rate of 50% (See section 4) against Google SMS-2FA system. In any case, these messages formed the basis of our study as we examined the efficacy of designed verification messages to resist deceptive attack messages. The next sections explain the principles and the procedures that helped us to identify effective countermeasure messages.

3. FIXING VERIFICATION MESSAGES

[1]Whereas we only tested this method in the context of SMS-based 2FA, we have no reason to believe that the identified principles would not apply to other security-related messaging contexts.

Table 1: Generated attack messages. We approximated the phishing ratio of each message using an online experiment. We selected the attack messages with phishing rate above 20% (highlighted in the table) for further field experiments.

#	Message	Category
x_1	We have received a report of harassment related to your Gmail account. Please reply with the verification code we just sent to you to receive the details privately.	Compliance
x_2	We have received a complaint of abuse of your Gmail account. Please reply with the verification code we just sent to you to receive the details privately.	Compliance
x_3	You have a voicemail on Google Voice. To listen, please reply with the message code we just sent to you.	Service
x_4	You have a private message on Google Voice. To listen, please reply with the message code we just sent to you.	Service
x_5	Your Gmail storage is expiring soon. To continue receiving new emails, please reply with the confirmation code we just sent to you.	Service
x_6	We are verifying your Gmail account. To continue your Gmail service without any interruption, please reply with the verification code we just sent to you.	Service
x_7	To continue using SMS-based verification for your Gmail account, please reply with the verification code we just sent to you.	Service
x_8	You have received a voice mail with delivery confirmation. Respond with the code to have it delivered.	Service
x_9	We are processing your request. To cancel the requested action on your Gmail account, you may reply with the verification code we just sent to you within 30 minutes.	Security
x_{10}	Your Gmail account has been accessed from Nigeria. If you do NOT want to authorize further access, please reply with the confirmation code we just sent to you.	Security
x_{11}	We received a request from you to delete all your Gmail messages from last month. To cancel, please reply the verification code we just sent to you.	Security
x_{12}	Did you request a password reset for your Gmail account? To cancel, please reply with the verification code we just sent to you.	Security
x_{13}	Your account has been limited due to a login attempt from an unknown location. To block the access, please reply with the verification code we just sent to you.	Security
x_{14}	Your password has been reset. Use the one-time code sent in the previous SMS. Log in now <*hyperlink*> using the one-time code in the previous SMS and select a new password.	Security
x_{15}	There was a password reset request for your Gmail account. To cancel, please forward the verification code we just sent to you.	Security
x_{16}	Did you request a password reset for your Gmail account? Delete this message if you did. Otherwise, send "Cancel" + the verification code we just sent to you.	Security

It has been previously established that communicating security risks to users is crucial, specifically when the hazard can not be mitigated automatically and the users need to make the final decision [14, 22]. Approaches to communicate security with users include *Warnings*, *Notices*, *Status indicators*, *Training*, and *Policy* [14]. In the context of verification code forwarding attack, unforgeable *notices* and *status* indicators (e.g., SSL certificate to validate the sender of SMSes) need modifications of SMS service infrastructures and SMS applications. *Training* is slow to utilize, and *policies* are not applicable in the context of this attack. However, *warning* is a scalable approach to communicate the security of verification codes with users over the SMS platform.

Warnings are shown to be more effective when they are placed proximate (in time and space) to the hazard [57, 24]. This suggests that the **verification code component (m)** and the **warning component (w)** should be delivered in one SMS. These two components together form the **new verification message (wm or mw).**

In this section, we discuss the challenges in designing warnings to reduce the susceptibility of users to VCFA (Section 3.1) and lay out the principles and framework for warnings (Section 3.2).

3.1 Challenges

The design of warnings has been explored well in the HCI domain (see, e.g., [57]). There are several studies on the design and evaluation of warnings of anti-phishing tools [38, 18, 20, 10], SSL warnings [52, 22, 5, 21], software updates [56], and anti-malware [6]. Two goals of warning messages are *Comprehension* and *Adherence* [21] where the former makes the user understand the risk, and the latter convinces the user to take the recommended action.

While existing design principles are useful for designing a warning component in verification messages, they are not sufficient. This is because SMS-2FA and SMS-based communications pose new challenges in designing warnings. Below, we list three major challenges:

1. **Undetectable attack vector:** Using existing SMS-2FA, there is no practical way for the service provider to detect a potential VCFA. Indeed, attackers as well as the owner of an account can initiate a password reset given the account ID (e.g, an email address) and phone number associated with the account. Consequently, a verification message can not be tailored differently for a malicious request as compared to a benign request. Therefore, the *same* warning component will be presented to users whenever a verification code is delivered. This is different from warnings for communicating security in other attack scenarios where some heuristics are initially used to detect a potential attack and warning is shown only if an attack is observed. For example, SSL warning message is shown to the user only when the SSL certificate provided by a website is invalid, and phishing warning is shown only if a website is potentially malicious.
2. **Passive warning communication:** In the traditional hazard situation, warnings *actively* communicate with the user at the time when a potential risk is observed. For example, in Chrome browser, if a user wants to visit a website with an invalid SSL certificate, a warning message first blocks the user from accessing the website. User needs to opt-in if she wants to proceed with visiting the website. On the other hand, SMS-2FA can not actively engage with the user to warn about risky behavior. Instead, the service provider has to deal with the attack by a passive warning.
3. **Limited UI capabilities:** Salience of a warning message represents the extent to which it gains attraction against a competing field of visual stimuli. A salient warning increases the readability, comprehension, recall, and compliance to warning [57, 11]. Font size and color, contrast, borders, pictorial symbols, and special effects (such as flashing lights) increase the salience of a warning message. However, *SMS user interface* does not support any of these visual effects. It is limited

to undecorated text without options for font color and size, and service providers preferably limit their communications to 160 characters (i.e. one SMS message) to minimize the costs.

New design principles that are able to tackle these challenges have to be devised and used in the warning components (in combination with current verification code component) against VCFA.

3.2 Design Principles

We now describe the principles that were used for designing warning components with the goal of reducing the susceptibility of users to VCFA. A number of these principles were identified to address specific challenges faced with communication in the context of SMS-based verification, and others are known and important principles that were tailored for this context.

- **Abuse-proof:** A warning component should be abuse-proof in two forms. *First*, the warning should be free from loopholes so the attackers can not misuse, subvert, or overwrite its intended meaning. For example, a warning component such as "Google will not send a follow-up message" (intended to disprove following attack messages) is *abusable* since it is unclear to recipients what the follow-up message is. After all, if the attacker incorporates the same text in his VCFA message then this may help the attacker – for example, "Your high-security message is pending. Respond to this SMS with the verification code for it to be delivered. Google will not send a follow-up message." *Second*, an attacker should not be able to launch a new attack by sending an SMS containing a spoofed verification message. For example, a warning that requests users to call a phone number or click on a link to report an unwanted verification message, is abusable; attackers can send a spoofed verification message with a malicious URL or a phone number to defraud users by alternative means.
- **Worry-free:** Embedding the security warning in the verification message should satisfy two seemingly conflicting properties. One one hand, the message should deal with the problem of *false alarm effect* created by presenting the warning component both in normal and attack scenarios. Otherwise, this can result in users perceiving the warning as less credible, perceiving threats as less intense or less probable, and overestimating their capability to cope with the danger. Since less intense warnings are affected less by this phenomena [12], SMS-2FA warning should be made gentle.
  On the other hand, the warning should help users to make the right decision and take the correct action when a VCFA occurs. Fortunately, in the event of a VCFA, the required action is *no-action* and does not demand strong emotions to be carried out, e.g., fear. In other words, it is sufficient that the user should ignore the message sent through attackers' request; according to the existing SMS-2FA, even changing the account's password does not provide any added protection because the attacker can repeatedly request a password reset even after the password has been changed.
- **Actionable and practical:** Only describing the nature of the hazard present in a VCFA is not sufficient to induce the required action. Instead, a warning message should be actionable and offer explicit directives or instructions on how to avoid potential hazards [57]. An action could be either to contact someone or even to ignore a message. For example, Google's warning – "Google will never text or call you to ask for a verification code" that is prompted in the voice-based verification – is informative, but not actionable. Moreover, the requested action should be practical and easy to accomplish. For example, it is not practical for average users to verify the sender's phone numbers (to identify if they are actually communicating with the service provider).
- **Concise and clear:** The warning component should be concise so most people take the time and effort to read it [57]. Moreover, in SMS-based communication, cost is another incentive to suggest concise messages (up to 160 characters), which should be "readable", "specific", and "Jargon-free" [21].

We considered these principles while creating warning components. We also use them to compare the merit of warnings qualitatively as a step for selecting more effective ones. Finally, we evaluate the effectiveness of a few selected warnings using a field experiment.

3.3 Instances of Warning Components

Principles for designing warnings are suitable to evaluate the quality of warning components, but they lack the capability to generate them. In this section, we present the framework we used to *generate* warning components. The framework comprises a list of *primitives* (i.e., atomic facts) that can help users to make the right decision at the time of a VCFA. The list of primitives is created by analyzing the elements involved in, and the process of SMS-based verification. For each primitive, we generate a warning that communicates the atomic fact with the users. To avoid complexity, we mostly focus on warnings with only one primitive. However, in a few instances, we found it suitable to mix more than one primitives to create warnings. Below, we list six primitives that we used for generating warnings.

1. **Verification code is sensitive:** There is a strong correlation between the use of security features and the risk perception of users [19]. Users' awareness of sensitivity of the verification code can change their attitude towards protecting the verification code. Therefore, a warning component emphasizing the sensitivity of the verification code may help users to think diligently before sharing their code with an imposter.
2. **The requested action (e.g., password reset) could not be completed without the verification code:** A large proportion of users believe that receiving an unwanted verification code means that somebody is hacking their account [51]. The fear induced by this knowledge prepares users for abuse by attackers. The knowledge about the fact that the requested action by an attacker (e.g., password reset) can not be completed without knowledge of the verification code can be helpful to assure users that the situation is under control. Indeed, because of this characteristic of SMS-2FA, users do not need to be worried or take immediate action when VCFA is observed. A stronger variation of this primitive is the fact that the verification code is only available within a limited time interval and will expire if not used.
3. **The verification code should be entered in the same UI as user's requests for the code:** This primitive is related to the correct process of using the verification code. In fact, it implies that (I) the user should have requested the code, and (II) the verification code should be entered on the same page from where she has requested the code. In comparison, in VCFA, an attacker lures the user to send the verification code that the user has not requested. Therefore, reminding users of the correct process might help her/him not to fall for phishing.
4. **Service providers will not contact to ask for verification code:** This primitive is related to the expected behavior from service provider and the correct verification process. Indeed,

Table 2: Generated warning components, primitive(s) they use, and the design principles they qualify as No (✗), Somehow (⊖) and Yes (✓). The final selected warning components evaluated by field experiment are highlighted.

#	Message	Primitive(s)	Abuse-proof	Concise/Clear	Worry-free	Actionable
w_1	If you did not request this code, beware of imposters asking you for this code.	5	✓	⊖	✗	✗
w_2	Google will not send a follow-up message.	4	✗	✓	✓	✗
w_3^*	Google will never text or call you to ask for a verification code.	4	✓	✓	✓	✗
w_4	Do not share this code via SMS or over phone call.	1,3,4	✓	⊖	✓	✓
w_5	You did not request this code? Do not enter this code anywhere and this code will expire automatically.	1,5	✓	⊖	⊖	✓
w_6	If you did not request this code, verify the phone number you are communicating with.	5	✗	⊖	✓	✗
w_7	If you did not request this code, do not use it and it will expire automatically after 10 minutes.	2	✓	⊖	✓	✓
w_8	Please ignore this message if you did not request a code.	2	✓	✓	✓	✓
w_9	Ignore this message if you do not want to proceed, or you did not request a code.	2,5	⊖	✗	✓	✓
w_{10}	Ignore this message if you do not want to proceed.	2	✗	⊖	✗	✓
w_{11}	Delete this message if this is not for you.	1,5	✓	⊖	✓	✗
w_{12}	This is a sensitive code. Delete this message if this is not for you or you do not want to proceed.	1,2	✓	⊖	✓	✗
w_{13}	Enter this code only on the Google's website where you requested the code.	3	✓	✓	✓	✓
w_{14}	Don't use this code if you do not want to proceed.	2	✗	⊖	✓	✓
w_{15}	If you did not request this code, ignore this message and your account will not be affected.	2,5	⊖	⊖	⊖	✓
w_{16}	If you did not request this code, ignore this message and no action will be taken.	2,5	⊖	⊖	✓	✓
w_{17}	We will cancel your password reset request automatically if you do not use this code.	2	✗	✓	✓	⊖

service providers will not contact the user to ask for verification code. Instead, the only way to use the verification code is to enter it into the website where the code has been requested. Being aware of this fact may help users to ignore follow-up messages by attackers asking for the code.

5. **A phishing attack is possible:** Awareness about threats is pursued by phishing training campaigns and is believed to be an important factor in curbing phishing attacks [39]. Similarly, the awareness about the fact that the attackers could be posing as service providers to launch an attack to steal the verification code and harm users maybe useful for users protection.

Based on the above primitives, we generated several warning components. The list of selected warnings chosen for further tests is shown in Table 2.

4. METHODOLOGY

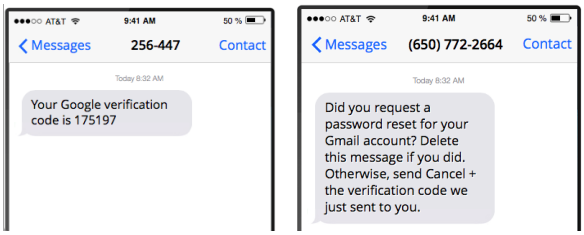
In this section, we describe our methodology to achieve the goal of finding an effective verification message that reduces the susceptibility of users to VCFA. For this, we first found an effective attack message using two experiments (E0 and E1), and then evaluated the new verification messages against the most effective attack message using E2 experiment. The components of and the sequence of experiments are shown in Figure 3.

4.1 Exp. E0: Online Experiment

We used an online survey to filter out the least effective attack messages. The survey showed two snapshots of a phone, one with a verification SMS from Google (Figure 2(a)) and the other a phishing message from an attacker (Figure 2(b)).

We asked the participants to select what they would do if they receive the two displayed SMS messages, one following the other. We provided the following options:

1. I will delete both of the messages
2. I will reply quickly to the second message
3. I will reply to the second message whenever I am free



(a) Google's verification message (b) Content of verification code forwarding attack message

Figure 2: Two snapshots of phone screens presented for online survey. The participants were asked what they would do if they receive these two messages one after the other. Participants who chose options indicating that they would reply with a code were counted as phished users.

4. Other

We counted the subjects that selected options 2 or 3 as phished users as these options signify intention to send the verification code to the attackers. We surveyed 800 subjects from MTurk to test 16 variations of attack messages (subjects were randomly and almost equally assigned to the variations). Using the results of this experiment, we eliminated the attack messages with yield below 20% from further evaluation in the next experiment. The phishing messages that we tested as well as the selected messages for the field experiment (highlighted) are shown in Table 1.

4.2 Field Experiments

In this section, we provide details of two field experiments E1 and E2. We conducted E1 to find the most effective attack message out of all that we generated, and conducted E2 to evaluate the success of new verification messages in reducing the susceptibility of users to VCFA. These experiments were between subjects and there was no overlap between the subjects of the experiments.

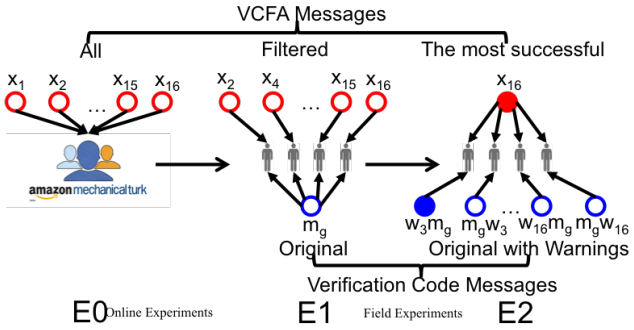


Figure 3: We used E0 to filter out the least effective attack messages among 16 attack messages. Field experiments E1 and E2 were respectively used for finding the most effective attack message (among nine attack messages), and evaluating the new verification messages. We used an online survey in E0, and conducted a field experiment in E1 and E2. x_i's are the attack messages; w_i's are the warning components; and m_g is the standard Google verification message.

4.2.1 Designing a VCFA Experiment

We followed a phishing experiment design that closely resembles the verification code forwarding attack [51]. For this purpose, we used two phone numbers, *phone A* and *phone B*, with Mountain View/CA area code (area code of Google's headquarters). During the experiment, *phone A* was used to send verification codes that supposedly were coming from a legitimate service provider (Google in our experiment), and *phone B* was used to send the phishing messages from attackers and receive users' replies. We detail the recruitment process and the experiment's procedure in this section.

- **Step 1:** *Subject Recruitment* – We recruited participants for our experiment by posting an advertisement for a user study on Craigslist [1] (New York/ Job, etc., and New York/ Community/Volunteers) and via Amazon Mechanical Turk (US Only) [2]. Due to the nature of the study, we only recruited participants who were between the age of 18 and 65. We did not disclose the use of phishing or any other details of the user study and only described it as a "study concerning the usage of communication devices (e.g., phone, computers, etc.)". The subjects gave us consent to use the information that they provided, including email addresses and phone numbers, to contact them for the purpose of the study.
  As the first step of the study, we asked the subjects to fill out a survey over the phone by calling and replying to a number of questions to an interactive voice response (IVR) system. The IVR system asked a total of eight questions about their age, gender, number of devices, and online activities. Along with recording the subjects' responses, the IVR system stored the caller ID of the participants. We used these phone numbers as the contact point of subjects for the experiment. We identified those who called from landlines or VOIP phones, using Twilio's phone lookup service, and eliminated them from the rest of the study because they could not receive SMSes. However, all participants who took the IVR survey were given a chance to win a raffle of five $20 Starbucks gift card and one $200 gift card.
- **Step 2:** *Sending a verification message* – In this step, we sent a verification message from the *phone A* to the subjects' phone number. The content of the verification messages varied for different group of subjects depending on the treatment under study. However, in order to avoid likely biases from the verification code in the message, we used the same verification code (i.e., "312985") for all treatments. We sent these messages on Wednesdays and Tuesdays between 3-5 pm.
- **Step 3:** *Sending a phishing message* – 30 seconds after the subjects received the verification code, we sent a phishing message to each participant using the *phone B* (attacker's phone). The content of phishing message varied based on the treatment under study. These messages aimed to deceive the users to send the verification code. We measure the success of these attack messages based on the percentage of subjects who sent the verification code to *phone B*.
- **Step 4:** *Online Exit Survey* – We debriefed the participants a day after we sent the VCFA phishing messages. This delay was applied mainly because we did not want to tip off potential victims ahead of time (some participants sent the code after 4 hours). We also invited them to fill out an online survey mostly concerned with the the SMS-based 2FA and their experience with SMS messages they received during the experiment.

Exp. E1*: Identifying an Effective Attack Message.*

We designed E1 to identify the most effective attack message among the messages that we selected in E0. We estimated their phishing rate using the phishing experiment described in Section 4.2.1. In the phishing experiment, we experimented with the selected attack messages against Google's verification message: m_g = *Your Google verification code is 312985*. We kept the verification code constant for all the subjects to avoid potential biases.

Exp. E2*: Identifying the Success of Warning Components Against the most Effective Attack Message Found in* E1 .

We designed E2 to evaluate the effectiveness of the new verification messages against the most effective attack message. To create the new verification messages, we combined the warning components that were selected in Section 3.3 (see Table 2) with the Google's verification message (m_g). We used the new verification messages as the first message in the VCFA experiment. We used the most effective message found in E1 as the attacker's message in this experiment.

It was not clear upfront whether the warning should be placed in the front of the message or at the end of the message. While some researchers find it more effective to place the warning before instructions [58], others disagree [43]. To evaluate both of the ordering methods, we created two variations of verification message for every warning; a) *Original message preceding Warnings:* by appending warning components to m_g e.g. $m_g.w_3$ forming "Your Google verification code is 312985. Please ignore this message if you did not request a code." b) *Warnings preceding original message* by appending m_g to warning components e.g. $w_3.m_g$ forming "Please ignore this message if you did not request a code. Your Google verification code is 312985." In total, we tested nine variations, four messages where warnings are preceded by m_g message and four messages where m_g is preceded by warnings. Moreover, we also tested the original m_g message against the most effective attack message found in E1.

4.2.2 Ethical Considerations

We followed recommendations from the literature of designing ethical and accurate phishing experiments [33, 23]. First, we con-

ducted experiments on real and diverse populations rather than only students, in field studies instead of "closed-lab" environments which might alert subjects about the real study and likely affect the outcome of the experiments. Our recruited subjects were not aware of the phishing experiments they were being tested on. Instead, they expected to receive an online survey about their computer usage (the exit survey). Thus, the result of our experiment could reflect the real success rate of VCFA in the wild. Second, we did not tamper with subjects' accounts and did not trigger real verification messages from the service provider (in this case, Google). We created and sent verification messages from our own phone number. No harm or service disruption was made to real subjects' accounts. Finally, our study was approved by the Institutional Review Board of our institution.

5. EVALUATION

In this section we present the results obtained from our experiments. We also discuss the effects of the content and order of the two components of the verification messages sent in E2, and study the confounding factors that may influence the results and conclusions. Our approach for evaluation is not only statistical, but we also support our findings based on the recurrent trends in the results of the experiments. In addition, we report some other interesting findings about user attitudes towards the SMS-2FA and differences among participants who were successfully phished and those who were not.

We recruited a total of 334 subjects using the procedure described in section 4.2.1. Out of the 334 participants, 100 were chosen randomly for experiment E1 and 234 for E2.

5.1 Evaluating the Success of Experiments

Success of a phishing attack in our experiments is computed based on the *phishing ratio*, i.e. the number of subjects who sent the verification code they received to the attacker.

Success of attacks in E1 .

We tested nine different attack messages against Google's standard verification message on a total of 100 subjects. The attack messages demonstrate different success rates as shown in Table 3(a). The attack message x_{16} : "*Did you request a password reset for your Gmail account? Delete this message if you did. Otherwise, send "Cancel" + the verification code we just sent to you.*" was the most effective attack message with a phishing rate of 60%. We use this message for evaluating our designed verification messages.

Success of verification messages in E2 .

In E2, we used x_{16}, the most effective attack message from E1, to evaluate the performance of new verification messages we designed. Each new verification message is composed of a warning component (w_i) and Google's standard verification message (m_g). To study the effect of the order of warning and verification code components, we experimented both cases of (I) code preceded by the warning ($w_i.m_g$), and (II) warning preceded by code ($m_g.w_i$). Google's standard verification message (m_g) was also included in the experiment separately as the control group.

Table 3(b) shows the result of the experiment for each verification message. We found that the verification message $w_8.m_g$: "*Please ignore this message if you did not request a code. Your Google verification code is 312985*" has the lowest phishing rate as compared to the others, and can reduce susceptibility of the users to VCFA dramatically; Out of 26 participants, only 2 (8%) fell for VCFA. This is six times less than the phishing rate of 50% for the

Table 3: Phishing rate for experiments E1 and E2. In E1, a number of generated attack messages were experimented against Google's standard verification message m_g to find the most *effective attack message*. In E2, performance of our new and designed verification messages were evaluated against the attack message (i.e. x_{16}) found in E1. m_g in our case was "Your Google verification code is 312985".

(a) E1

Verification Message	Attack Message	# Subjects	% Phished
m_g	x_2	15	33
m_g	x_4	10	10
m_g	x_5	10	40
m_g	x_8	10	30
m_g	x_{11}	10	30
m_g	x_{13}	15	13
m_g	x_{14}	10	10
m_g	x_{15}	10	10
m_g	$x_{16}*$	10	**60**

(b) E2

Verification Message	Attack Message	# Subjects	% Phished	% Avg. Phished
$m_g.w_3$	x_{16}	26	34	36
$w_3.m_g$	x_{16}	26	38	
$m_g.w_8$	x_{16}	26	31	20
$w_8.m_g$	x_{16}	26	**8***	
$m_g.w_{13}$	x_{16}	26	50	36.5
$w_{13}.m_g$	x_{16}	26	23	
$m_g.w_{16}$	x_{16}	26	23	17
$w_{16}.m_g$	x_{16}	26	11	
m_g	x_{16}	26	50	50

m_g i.e. 13 out of 26 subjects. We use Bonferroni corrected pairwise comparison, and compare different verification messages using Fisher's Exact test. Specifically, Fisher's Exact test with Bonferroni correction between m_g and $w_8.m_g$ shows that there is a high statistical significant difference between the phishing ratio of two verification messages ($p - value < 0.005$). On a similar note, there is also a striking statistical difference between phishing ratio of m_g and $w_{16}.m_g$ ($p - value < 0.005$). For most of other verification messages, except $m_g.w_{13}$ result shows that warnings are effective in reducing the susceptibility of users to phishing attack.

To effectively evaluate the role of the warnings regardless of their position in the verification message, we compute the average success rate of phishing attack for pairs of the verification messages with the same warning (e.g., $m_g.w_3$ and $w_3.m_g$). The average phishing ratio for all pairs is shown in Table 3(b). As it can be noticed, warnings w_{16} and w_8 perform better than other warnings regardless of their position in the verification message. Therefore, generating appropriate warnings is important.

5.2 Effect of Ordering of Warning Components in E2

To understand the effect of relative position of the warning component and the verification code component within a verification message, we formed two groups of messages: a) group of messages that *start* with a warning, and b) group of messages that *end* with a warning. As we can see in the Table 3(b), for the former, the success rate of phishing ranges from 8% to 38% whereas for the latter it is between 23% and 50%.

Chi-squared test between the ratio of the two groups shows that there is a statistically significant difference between the two groups and there is an effect of the placement of the warning component ($\chi^2 = 5.43, p < 0.05$). In other words, in the verification messages, we found that warnings appearing before the verification code are more effective.

5.3 Factors Affecting Phishing Outcome in E2

Various hypotheses were put forward in the course of our experiments to account for the confounding factors in the phishing

outcome in E2. Specifically, we studied the effect of gender, age, and previous usage of SMS-2FA mechanisms on the susceptibility of the users to phishing.

5.3.1 Gender

Phishing studies have shown that women are more susceptible to phishing emails [32, 50]. In the experiment E2, our participants pool consists of 97 males and 137 females (59% females). We assigned similar ratio of men and women to all verification messages we tested. The result of the experiment shows that a similar percentage of males (27%) and females(29%) fall for phishing (Figure 4). Chi-squared test between the phishing ratio of the two groups does not show a significant effect of the gender on the success of VCFA ($\chi^2 = 0.10, p = 0.74$).

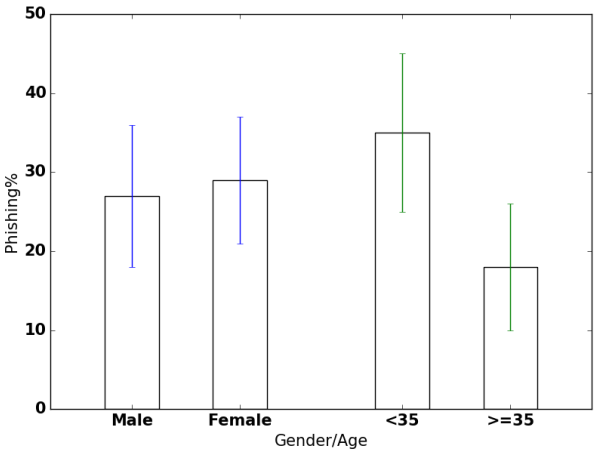


Figure 4: Phishing rate based on gender and age in E2 experiment. The difference between the performance of participants in younger vs. older groups is significant.

5.3.2 Age

Previous phishing studies show that younger people (18-25) are more susceptible to phishing [50]. This is attributed to several factors including lower level of education, fewer years of experience with the Internet, less exposure to security training material, and less aversion to financial risk [50]. We studied the relation between age and success of VCFA. Since we captured age of subjects in the online exit survey, we had data for only a subset of the subjects (154 –69%– of participants in the experiment took the exit survey). Out of 154, 97 participants were between the age of 18 and 35 and rest 57 were more than 35 years old. Out of 57 younger participants, 35% participants fell for VCFA and 18% fell for phishing among older participants. Chi-square test between the ratio of the two groups shows the significance of the effect of the group ($\chi^2 = 4.5, p < 0.05$). One might speculate that the difference between percentage of phishing is because older people are less familiar with SMS-based 2FA. This is not the case in our experiment; as declared in the exit survey, 77% (44 out of 57) of the users over 35 years old had used SMS-2FA before. Similarly, 82% (80 out of 97) of users below 35 had used SMS-2FA before. Therefore, the difference between the performance of different age groups potentially means that younger people are more susceptible to phishing of this type. This is consistent with the findings of other researchers [50].

5.3.3 SMS-2FA Usage

Only 1% (3 out of 157) of the participants from the online exit survey did not have a Gmail account. 20% of users had not used SMS-based 2FA before. We compared the percentage of subjects who fell for phishing between group of people who have used SMS-based 2FA and those that have not used it. We learned that similar percentage of subjects i.e. 30% of those who have not used SMS-based 2FA, and 29% (35 out of 122) of those who had used it fell for phishing. Therefore, we do not see any effect of previous experience and usage of 2FA.

5.4 Response Time

In our experiments, the response time of users to the attack messages were different. While some users replied quickly (within one minute), others were slower (the most delayed reply was sent after 4 hours and 24 minutes). The overall median time of sending a reply was 3 minutes, and 87% of phished users replied within an hour after they received the code, which is in range of Google's verification code expiration [16]. However, for computing the phishing rates in our experiments, we did not consider the response delays. Attackers can potentially decrease the delay of a solicited response by sending at different times of the day or repeating the attack right after they get a reply from the victim. We have not tested these methods and leave them as a future work.

5.5 "Cancel" and Other Reply Messages

A subset of users replied with messages other than verification codes. For example, a handful sent "Cancel" with wrong code or sent the verification code to Google's phone number (*phone A*). We are aware of this because both phone A and phone B were in our control, though in reality this may not be the case. A relatively large number of users (N=34) sent only "Cancel" without a verification code. To confirm their actual intention, we asked subjects in the online exit survey the reason of not sending the verification code. 55% (19 out of 34) of the participants who only sent "Cancel" without code confirmed that they misunderstood the request. Remaining users mentioned reasons including the suspiciousness of the sender of messages and sensitivity of the code as the reason for not sending the code along with "Cancel". The attacker can potentially improve the attack's yield by tweaking the message to remove the confusion.

6. DISCUSSION AND LIMITATIONS

In this section, we discuss other possible approaches to mitigate social engineering attack on out-of-band authentication. Furthermore, we talk about best practices and recommendations for improving the security of SMS-2FA with respect to social engineering attacks. Finally we discuss the limitations of this work.

Other possible approaches.

Our proposal to address the problem of out-of-band phishing by designing better verification messages does not preclude other countermeasures. Machine learning techniques can be used to automatically detect potentially malicious verification code requests. One aspect of this can be profiling users based on their locations using the capabilities of SMS service providers. This may limit the choices of the geographical locations for the attackers.

Majority of users don't verify the sender of verification codes. Specifically, in the online exit survey that we ran after the field experiments, 73% of users mentioned that they did not verify the phone number of sender they received during the experiments. This is inspiring though, that verifying the sender of SMSes makes a significant difference; while only 20% of the group of users who verified the sender fell for phishing, 32% of the other group fell for

phishing. As a result, adding a naming directory or mechanism that enables users to verify the sender of the verification messages can potentially make users less susceptible.

Table 4: Example SMS-2FA messages in use.

Providers	2FA-Message
Google	Your Google verification code is [6 digit]
Microsoft	Microsoft account verification code: [4 digit]
Wells Fargo	Use Wells Fargo verification code [6 digit]
Yahoo	Your Yahoo verification code is [8 letters]
AT&T	AT&T Free Msg: Your temporary password is [6 digits]. If you did not request a temporary password, call 1-800-ATT-202. We'll never contact you to ask for this password.
Twitter	Your Twitter confirmation code is [6 digits].
Dropbox	Your security code is [6 digit]. Happy Dropboxing!
Facebook	[6 digits] is your Facebook password reset code, or reset your password here: [URL]
Instagram	Use [6 digits] to verify your Instagram account.
Telegram	Telegram code [5 digits]
Snapchat	Snapchat Code: [6 digits]. Happy Snapping!
Wallet One	Enter the password [6 digits] It is valid for 30 min
OptionFair	Hello! Your PIN is: [6 digits].
AddisonLee	Enter this code: [4 digits] to complete your registration and you're ready to go
Telecom	Welcome to our HotSpot. Your SMS-Code is [7 digits]
Wasabee	Welcome to Wasabee. Your registration code is: [6 digits]
Zoosk	Welcome to Zoosk! Enter [5 digits] on website to validate your account.

Best practices for improving the security of SMS-2FA.

By studying verification messages of well-known service providers (see Table 4), we noticed that most of them don't use a warning to protect users against social engineering. Warnings used by some of the service providers do not qualify according to the design principles we established in this work. As an example, AT&T opens up a bigger threat to users by asking them to call a phone number if they receive an unwanted verification code. This in turn can be used to launch another type of social engineering attack (e.g., the attacker can send a message with a fake verification code and include a malicious phone numbers to launch a Vishing attack [26]). Several other such attacks have been discussed in the recent past [28, 27, 7]. As a result, service providers who use or provide SMS-based 2FA should potentially redesign their verification messages to thwart VCFA.

Moreover, service providers and the security research community should increase awareness about the possibility of VCFA that can affect millions of users in a dramatic way. Service providers are recommended to warn users about the possibility of such attacks, providing guidance on how users can verify the phone numbers they receive SMSes from (potentially by providing a list of phone numbers that the service provider uses to send the verification codes), and also provide guidance for users on appropriate and effective action in face of VCFA. However, this is not full-proof as there exists techniques to spoof the sender's number while delivering the message.

As for now, there is no practical way for users to rollback a process (either login or password reset) triggering an unwanted verification code. The best action of the user against VCFA seems to be "no-action" (changing the password is not useful because the attacker can request the verification code again knowing the email address only). However, it is not natural to ask users to take no action upon perceiving a risk. Therefore, service providers are recommended to design a process for users to see critical actions requested on their account and be able to cancel them while the attackers are waiting for the verification code. Alternatively this can be used as a security report mechanism usable by service providers to flag the attacking IP addresses or to increase security measures on an account.

Limitations.

Our findings should be seen as strong indications that messaging matters – however, there may very well be attacks with higher yield than those we have tested. Our experiments, for example, did not involve causing service disruptions to users to make the need for a 2FA more credible. Such attacks, which no doubt would have a higher yield than the attacks we simulated, are also more complex to mount, and have the potential of harming the subjects. In contrast, our experiments did not pose this risk.

Similarly, we do not in any way lay claims of having identified the service provider messages with the best ability to defuse attacks – we have simply identified the great potential for improvements, and the risk of complacency. However, to further verify the efficacy of our proposed countermeasure against attack messages, we tested the best countermeasure identified based on E2 (i.e. $w_8.m_g$) against the attack message from a recent work [51]. The attack message we tested against the new verification message was: "*Please verify that your phone is still associated with your Gmail account by replying to this message with the code we have just sent to you.*". We found that only 4% (1 out of 26 subjects) fell for phishing, showing a dramatic decrease in comparison with 25% phishing rate of the same attacker against Google's standard message as reported in [51].

7. RELATED WORK

In this section, we present an overview of state-of-the-art attacks on SMS-based 2FA. We also relate our work with studies on anti-phishing message design.

Attacks against SMS-based 2FA.

Since SMS-based 2FA started to gain popularity among service providers, numerous attacks against this mechanism have been suggested or observed in the wild.

The first attack class exploited numerous vulnerabilities in cellular network infrastructure, including the use of weak encryption [9, 8, 17, 45] and the fundamental flaw that allows mobile devices to connect to rogue base stations [25, 15, 4], to eavesdrop verification codes. Decrease in cost of cellular network eavesdropping devices has made this class of attack class become more popular among less powerful attackers, besides nation state backed actors [15].

The second and possibly the most popular class of attacks on SMS-based 2FA observed in the wild used malware to steal verification codes and bypass the entire authentication process [16, 36]. The first step of this type of attack is to install malware on a user's device such as a PC or mobile phone. Using cross platform infection techniques, these malwares can then infect other user's devices and can completely circumvent 2FA authentication process. In contrast, in VCFA, the attacker does not need to control a user device which makes this attack more scalable.

Another class of attacks on SMS-based 2FA exploited the vulnerabilities in voicemail protection from many mobile carriers to steal verification codes [37, 49]. Many service providers allow verification codes to be delivered by a phone call instead of an SMS. An attacker triggers a phone call delivering a verification code to vic-

tim's phone number, often when the victim's phone is in no-disturb mode (e.g., at night) so the call is redirected to voicemail. He then connects to victim's voicemail inbox to listen and grab the code by spoofing victim's number which can be done fairly easy [46]. This is due to the vulnerability on how mobile carriers identify a caller [46] and also the fact that many users don't set up a PIN or use a default PIN for their voicemail inbox [48].

Although less popular, phishing attacks on SMS-based 2FA are on the rise and many real-world incidents have been recorded. "Operation Emmental" [42] has used a method to launch a general phishing attack to access 2FA protected bank accounts using a combination of social engineering techniques and a malware installed on user's phone to steal One-Time-Passwords (OTP) issued by the bank and delivered by SMS. Citizen Lab has documented growing instances of a phishing campaign targeting Iranian's diaspora and Western activists that tricks victims to enter credentials and verification codes to a fake Gmail login page [47]. Symantec also reported an emerging phishing attack similar to VCFA which lured users to forward verification codes to scammers [54]. Previous work has experimentally measured the success rate of VCFA (25%)[51]. In this paper, we systematically studied VCFA in terms of attacker strategies and showed that attackers can significantly increase their success rate by tweaking their attack messages. We also found potential verification messages that can reduce success rate of the attacks dramatically.

Anti-phishing message design.

The success of VCFA and similar attacks can be attributed to the lack of warning in the content of verification messages, as shown in Table 4. Warning message as an effective tool to combat phishing attacks have been extensively studied in the literature and are considered a valuable part of the anti-phishing tool set [38, 18, 20, 10], browser SSL warnings [52, 22, 5, 21], software updates [56], and anti-malware [6]. Principles for the design of effective warning messages were also introduced in numerous studies [57, 11, 21]. We followed these principles but also introduced new principles to communicate security in the very specific scenario of SMS communication.

8. CONCLUSIONS

In this work, we examined the problem of social engineering attacks on SMS-based 2-factor authentication. By means of experiments, we observed that an aggressive attacker could lure as many as 50% of users to forward him their verification code. We demonstrated that the SMS message containing the verification code sent by the service provider can play a critical role in mitigating these attacks. We developed principles for designing abuse-proof verification messages to reduce the susceptibility of users in forwarding the verification code to the attacker. We evaluated a number of such messages using a field experiment. The results of our experiment demonstrate the potential of better messaging. They revealed that the success of the attack could be reduced to just 8%. Study of other social engineering approaches against out-of-band authentications and possible remediations remain future work.

9. REFERENCES

[1] Craigslist. https://newyork.craigslist.org.

[2] Mturk. https://www.mturk.com.

[3] S. Abraham and I. Chengalur-Smith. An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society*, 32(3):183–196, 2010.

[4] Z. Ahmadian, S. Salimi, and A. Salahi. New attacks on umts network access. In *Wireless Telecommunications Symposium, 2009. WTS 2009*, pages 1–6. IEEE, 2009.

[5] D. Akhawe and A. P. Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *Usenix security*, pages 257–272, 2013.

[6] H. Almuhimedi, A. P. Felt, R. W. Reeder, and S. Consolvo. Your reputation precedes you: History, reputation, and the chrome malware warning. In *SOUPS*, pages 113–128, 2014.

[7] M. Balduzzi, P. Gupta, L. Gu, D. Gao, and M. Ahamad. Mobipot: Understanding mobile telephony threats with honeycards. In *Proceedings of the 11th ACM SIGSAC Symposium on Information, Computer and Communications Security*, ASIA CCS '16, New York, NY, USA, 2016. ACM.

[8] E. Barkan, E. Biham, and N. Keller. Instant ciphertext-only cryptanalysis of gsm encrypted communication. *Journal of Cryptology*, 21(3):392–429, 2008.

[9] A. Biryukov, A. Shamir, and D. Wagner. Real time cryptanalysis of a5/1 on a pc. In *Fast Software Encryption*, pages 1–18. Springer, 2000.

[10] C. Bravo-Lillo, L. Cranor, S. Komanduri, S. Schechter, and M. Sleeper. Harder to ignore? revisiting pop-up fatigue and approaches to prevent it. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 105–111, 2014.

[11] C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter. Your attention please: designing security-decision uis to make genuine risks harder to ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 6. ACM, 2013.

[12] S. Breznitz. *Cry wolf: The psychology of false alarms*. Psychology Press, 2013.

[13] R. B. Cialdini. The psychology of persuasion. *New York: Quill William Morrow*, 1984.

[14] L. F. Cranor. A framework for reasoning about the human in the loop. *UPSEC*, 8:1–15, 2008.

[15] A. Dabrowski, N. Pianta, T. Klepp, M. Mulazzani, and E. Weippl. Imsi-catch me if you can: Imsi-catcher-catchers. In *Proceedings of the 30th Annual Computer Security Applications Conference*, ACSAC'14, pages 246–255, New York, NY, USA, 2014. ACM.

[16] A. Dmitrienko, C. Liebchen, C. Rossow, and A.-R. Sadeghi. On the (in) security of mobile two-factor authentication. In *Financial Cryptography and Data Security*, pages 365–383. Springer, 2014.

[17] O. Dunkelman, N. Keller, and A. Shamir. A practical-time related-key attack on the kasumi cryptosystem used in gsm and 3g telephony. *Journal of Cryptology*, 27(4):824–849, 2014.

[18] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1065–1074. ACM, 2008.

[19] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner. Are you ready to lock? In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 750–761. ACM, 2014.

[20] S. Egelman and S. Schechter. The importance of being earnest [in security warnings]. In *Financial cryptography and data security*, pages 52–59. Springer, 2013.

[21] A. P. Felt, A. Ainslie, R. W. Reeder, S. Consolvo,

S. Thyagaraja, A. Bettes, H. Harris, and J. Grimes. Improving ssl warnings: comprehension and adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2893–2902. ACM, 2015.

[22] A. P. Felt, R. W. Reeder, H. Almuhimedi, and S. Consolvo. Experimenting at scale with google chrome's ssl warning. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2667–2670. ACM, 2014.

[23] P. Finn and M. Jakobsson. Designing ethical phishing experiments. *IEEE Technology and Society Magazine*, 26(1):46–58, Spring 2007.

[24] J. P. Frantz, T. P. Rhoades, S. L. Young, and J. A. Schiller. Assessing the effects of adding messages to warning labels. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 44, pages 818–821. SAGE Publications, 2000.

[25] N. Golde, K. Redon, and R. Borgaonkar. Weaponizing femtocells: The effect of rogue devices on mobile telecommunications. In *NDSS*, 2012.

[26] S. E. Griffin and C. C. Rackley. Vishing. In *Proceedings of the 5th annual conference on Information security curriculum development*, pages 33–35. ACM, 2008.

[27] P. Gupta, M. Ahamad, J. Curtis, V. Balasubramaniyan, and A. Bobotek. M3AAWG Telephony Honeypots: Benefits and Deployment Options. Technical report, 2014.

[28] P. Gupta, B. Srinivasan, V. Balasubramaniyan, and M. Ahamad. Phoneypot: Data-driven understanding of telephony threats. In *22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2014*. The Internet Society, 2015.

[29] S. Gupta, P. Gupta, M. Ahamad, and P. Kumaraguru. Abusing phone numbers and cross-application features for crafting targeted attacks. *arXiv preprint arXiv:1512.07330*, 2015.

[30] T. Hunt. Pwned websites list. https://haveibeenpwned.com/PwnedWebsites. [Online; accessed 22-May-2016].

[31] ic3. Internet crime complaint center(ic3). http://www.ic3.gov/media/2015.aspx, 2015. Accessed: 2016-05-17.

[32] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.

[33] M. Jakobsson and J. Ratkiewicz. Designing ethical phishing experiments: a study of (ROT13) ronl query features. In *Proceedings of the 15th international conference on World Wide Web*, pages 513–522. ACM, 2006.

[34] Kaspersky. Asacub android trojan: From information stealing to financial fraud. http://www.kaspersky.com/about/news/virus/2016/Asacub-Android-Trojan-From-Information-Stealing-to-Financial-Fraud, 2016. Accessed: 22-May-2016.

[35] E. Kim, K. Park, H. Kim, and J. Song. I've got your number. In *Information Security Applications*, pages 55–67. Springer, 2014.

[36] R. K. Konoth, V. van der Veen, and H. Bos. How anywhere computing just killed your phone-based two-factor authentication. 2016.

[37] krebsonsecurity. Attackers Hit Weak Spots in 2-Factor Authentication. http://krebsonsecurity.com/2012/06/attackers-target-weak-spots-in-2-factor-authentication/, 2012. [Online; accessed 22-May-2016].

[38] P. Kumaraguru, Y. Rhee, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 905–914. ACM, 2007.

[39] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong. Lessons from a real world evaluation of anti-phishing training. In *eCrime Researchers Summit, 2008*, pages 1–12. IEEE, 2008.

[40] S. Kurowski. Using a whatsapp vulnerability for profiling individuals. *Open Identity Summit, GI-Edition-Lecture Notes in Informatics (LNI)-Proceedings*, 237:140–146, 2014.

[41] latimes. Anthem hack exposes data on 80 million; experts warn of identity theft. http://www.latimes.com/business/la-fi-anthem-hacked-20150204-story.html#page=1, 2015. Accessed: 22-May-2016.

[42] T. Micro. Finding Holes Operation Emmental. http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-finding-holes-operation-emmental.pdf, 2014. [Online; accessed 22-May-2016].

[43] J. M. Miller, J. P. Frantz, and B. W. Main. The ability of two lay groups to judge product warning effectiveness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 37, pages 989–993. SAGE Publications, 1993.

[44] D. Modic. Willing to be scammed: How self-control impacts internet scam compliance. 2012.

[45] C. Mulliner, R. Borgaonkar, P. Stewin, and J.-P. Seifert. Sms-based one-time passwords: attacks and defense. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 150–159. Springer, 2013.

[46] T. Register. Reg probe bombshell: How we HACKED mobile voicemail without a PIN. http://www.theregister.co.uk/2014/04/24/voicemail_still_easy_to_hack/. [Online; accessed 22-May-2016].

[47] J. Scott Railton and K. Kleemola. London Calling: Two-Factor Authentication Phishing From Iran. https://citizenlab.org/2015/08/iran_two_factor_phishing/, 2015. [Online; accessed 22-May-2016].

[48] S. N. Security. How phone hacking worked and how to make sure you're not a victim. https://nakedsecurity.sophos.com/2011/07/08/how-phone-hacking-worked/. [Online; accessed 22-May-2016].

[49] S. Shah. How I bypassed 2-Factor-Authentication on Google, Facebook, Yahoo, LinkedIn, and many others. https://shubh.am/how-i-bypassed-2-factor-authentication-on-google-yahoo-linkedin-and-many-others/. [Online; accessed 22-May-2016].

[50] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 373–382. ACM, 2010.

[51] H. Siadati, T. Nguyen, and N. Memon. Verification code forwarding attack (short paper). In *International Conference on Passwords*, pages 65–71. Springer International Publishing, 2015.

[52] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F.

Cranor. Crying wolf: An empirical study of ssl warning effectiveness. In *USENIX Security Symposium*, pages 399–416, 2009.

[53] Symantec. Android.ackposts. https://www.symantec.com/security_response/writeup.jsp?docid=2012-072302-3943-99, 2012. Accessed: 22-May-2016.

[54] Symantec. Password recovery scam tricks users into handing over email account access. http://www.symantec.com/connect/blogs/password-recovery-scam-tricks-users-handing-over-email-account-access, 2015. [Online; accessed 22-May-2016].

[55] Versprite. Android infostealer - godwon - analysis. http://versprite.com/og/android-infostealer-godwon-analysis/, 2012. Accessed: 22-May-2016.

[56] R. Wash, E. J. Rader, K. Vaniea, and M. Rizor. Out of the loop: How automated software updates cause unintended security consequences. In *SOUPS*, pages 89–104, 2014.

[57] M. S. Wogalter, V. C. Conzola, and T. L. Smith-Jackson. Research-based guidelines for warning design and evaluation. *Applied ergonomics*, 33(3):219–230, 2002.

[58] M. S. Wogalter, G. A. Fontenelle, and K. R. Laughery. Behavioral effectiveness of warnings. 29(7):679–683, 1985.