Is Real-time Phishing Eliminated with FIDO?
Social Engineering Downgrade Attacks against FIDO Protocols

Enis Ulqinaku*, Hala Assal†, AbdelRahman Abdou‡, Sonia Chiasson‡, Srdjan Capkun*

*Department of Computer Science, ETH Zürich, Switzerland
†Department of Systems and Computer Engineering, Carleton University, Canada
‡School of Computer Science, Carleton University, Canada

Abstract
FIDO’s U2F is a web-authentication mechanism designed to mitigate real-time phishing—an attack that undermines multi-factor authentication by allowing an attacker to relay second-factor one-time tokens from the victim user to the legitimate website in real-time. A U2F dongle is simple to use, and is designed to restrain users from using it incorrectly. We show that social engineering attacks allow an adversary to downgrade FIDO’s U2F to alternative authentication mechanisms. Websites allow such alternatives to handle dongle malfunction or loss. All FIDO-supporting websites in Alexa’s top 100 allow choosing alternatives to FIDO, and are thus potentially vulnerable to real-time phishing attacks. We crafted a phishing website that mimics Google login’s page and implements a FIDO-downgrade attack. We then ran a carefully-designed user study to test the effect on users. We found that, when using FIDO as their second authentication factor, 55% of participants fell for real-time phishing, and another 35% would potentially be susceptible to the attack in practice.

1 Introduction
Fast Identity Online (FIDO) is driven by an industry alliance with the goal of reinforcing web authentication by “reducing the world’s over-reliance on passwords” [5]. The alliance now comprises 42 members including Amazon, Apple, Arm, Google, Microsoft, PayPal, as well as financial corporations like American Express, Mastercard, Visa, and Wells Fargo.

FIDO’s U2F standard defines cryptographic challenge-response protocols where a dongle with a private key can prove its identity to a pre-registered website. The dongle interacts with a user’s device through a Universal Serial Bus (USB) port, or wirelessly using Near-Field Communication (NFC) or Bluetooth (BLE). Such dongles are now manufactured by many companies, including Yubico and Feitian Technologies.

The technology resists exposing the secret key, comparable to some Physically Unclonable Function (PUF) technologies [3]. The challenge-response computations are performed on the dongle itself, and the private key never leaves the dongle. U2F thus enjoys relatively high resistance to the common cases of malware that run on the user’s machine. Physical theft of the dongle compromises its defence, however, such attacks are not scalable and cannot be performed remotely.

In U2F, the domain (string) in the browser’s address bar is a function of the challenge-response protocol. The browser sends that string to the dongle. In case of phishing [85, p.269], the domain will be that of the attacker’s website. Thus, an attacker relaying the result of the challenge-response from the browser to the legitimate website does not gain access; the response will not match the website’s expectation. U2F is therefore a strong defender against phishing attacks [60], including the devastating real-time phishing attacks that undermine various Two-factor Authentication (2FA) alternatives [50, 57]. In real-time phishing, attackers relay the One-Time Password (OTP) (generated on the user’s phone or sent over SMS) on the fly to the legitimate website. The FIDO alliance highlights the abilities of its suite of technologies in handling phishing [4]: “This security model eliminates the risks of phishing, all forms of password theft and replay attacks”, “[the] built-in phishing resistance and ease-of-use give it the potential to drive widespread adoption”.

We show that FIDO could nonetheless be downgraded to weaker options, enabled by websites that allow users to setup second-factor alternatives to FIDO. These are typically configured to account for, e.g., dongle loss, malfunction, or other reasons where a user simply wants to avoid using the dongle (e.g., grant access to a remote spouse). Despite extensive design efforts to empower users with a complete mental model, and previous literature showing the high usability and likeability of FIDO [30], we show how clever social engineering tactics enable a real-time phishing attacker to impersonate FIDO users, requiring neither malware nor dongle theft.

*An extended version of this work is available in [83]. Author’s copy, published at USENIX Security 2021 https://www.usenix.org/conference/usenixsecurity21/presentation/ulqinaku

†FIDO assumes a trusted browser.
We construct a real-time phishing attack which targets FIDO users and works as follows. When the legitimate website prompts the adversary to insert the U2F dongle, the adversary likewise prompts the user on their phishing website. As the user inserts their dongle, the adversary asks the legitimate website to use an alternative method, and prompts the user to submit the OTP of that method on its phishing website. We posit that users can perceive this as an additional, third authentication factor, on top of the dongle they just inserted, thus interpreting more steps as higher security [51,91]. On the phishing website, the adversary simply ignores the dongle’s response, and relays the user-submitted OTP of the alternative method to the legitimate website, hence gaining access.

We inspected Alexa’s Top 100 websites to verify if they allow choosing alternatives to FIDO during login. We found that all websites that support FIDO (23 out of 100) allow choosing weaker alternatives. Hence, their users remain potentially vulnerable to real-time phishing despite using FIDO. Ironically, most of these websites force users to first register an alternative 2FA method before being able to register FIDO as a second factor. Google’s Advanced Protection [32] program accepts logins only with security keys, however recovery at scale remains challenging for such accounts (Sec. 8.1).

In this paper we approach two research questions, (RQ1) How susceptible are users to phishing attacks when using FIDO? (RQ2) How do users detect phishing attacks when using FIDO? By implementing a website that mimics real-time phishing of Google’s login form, and through a carefully-designed user study of 51 participants, we found that only 10% of participants are unlikely to fall for (general) phishing in practice. They detected our phishing attempts early in the study, e.g., from the phishing email or the phishing URL, before reaching our downgrading FIDO part. Had they missed the regular phishing indicators, it is unclear whether these participants would detect our downgrade attack in practice.

Contributions. This paper contributes new social engineering attacks that allow an adversary to downgrade FIDO to weaker 2FA alternatives. Such alternatives are vulnerable to real-time phishing, which is the primary attack that FIDO is designed to exhaust. By allowing such downgrade, FIDO’s defence against real-time phishing is only partial. The presented methodology of evaluating the effectiveness of the new attacks can be of independent interest to future researchers.

None of our attacks exploit weaknesses in the FIDO standards, APIs, or cryptographic protocols themselves. The core enabler is rather the availability of authentication alternatives. So long as users are allowed to login using weaker alternatives, attackers also can always leverage them. In general, it is necessary to either allow alternative login methods to hardware tokens, or implement non-weaker account-recovery mechanisms to account for token losses/malfunctions. Manual recovery is costly [61]. And with adversaries now capitalizing on an ongoing pandemic [41], and a global work-from-home pattern, it becomes increasingly important to make sure promising defences like FIDO are not undermined.

2 Background

Two-Factor Authentication. 2FA is a widely deployed strategy to strengthen password authentication. It usually requires users to enroll a second factor (e.g., smartphone or special hardware) to their accounts during registration. Afterwards, upon submitting the correct password for login, the user is asked to prove possession of the second factor. To do so, most 2FA schemes require the user to submit an OTP displayed, or confirm a prompt, on their phone [11,48,50,57,58,82]. To enhance user experience (reduce inconvenience of a method) and availability (access to the user’s account), online services typically allow users to enroll more than one 2FA alternative per account.

Threat Model: Real-time phishing. Existing 2FAs protect users from password compromise but they largely remain vulnerable to real-time phishing. In real-time phishing, the user interacts with the malicious page posing as the genuine website, while the adversary authenticates simultaneously on the real website by relaying victim’s credentials. The attack is relatively easy from a technical perspective, and very effective in practice [19]. However, it is very challenging to be prevented because it mostly exploits human mistakes. Prompt notifications enhance user experience, however, they put the burden onto users to detect ongoing attacks and risk user habituation [7]. Automated tools, e.g., Evilginx [34], make real-time phishing easy to deploy and largely scalable.

FIDO Specification. FIDO alliance aims to reduce the reliance on passwords, while preserving usability. FIDO assumes three trusted and cooperating components: i) relying party, which is the server where the user authenticates; ii) client, which typically is the browser; and iii) authenticator, which is the device the user possesses. The key advantage of FIDO compared to other 2FA schemes is that the browser provides the authenticator with the domain of the visited website. Therefore, if the user falls for phishing, the browser communicates the malicious domain to the authenticator, which signs a message that is invalid to the honest server.

The alliance published three specifications [5]: (1) U2F covers use cases where the authenticator is used as a second factor; (2) UAF, which is known as “passwordless authentication”; (3) FIDO2, which is the latest specifications covering use cases of both U2F and UAF. Unless specified, “FIDO” herein refers to all three specifications described above.

FIDO2 includes WebAuthn API and the Client to Authenticator Protocol (CTAP2). CTAP2 triggers browsers to display a prompt window, which includes the domain name, when a website tries to communicate with the dongle. CTAP2 is backward compatible and supports U2F functionalities, which
Table 1: All 23 FIDO-enabled websites in Alexa’s top 100 allow weaker 2FA alternatives to be registered alongside FIDO.

<table>
<thead>
<tr>
<th></th>
<th>Support FIDO</th>
<th>Do not support FIDO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>allow alternatives</td>
<td>do not allow alternatives</td>
<td></td>
</tr>
<tr>
<td>FIDO partner</td>
<td>14</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Others</td>
<td>9</td>
<td>0</td>
<td>62</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>0</td>
<td>77</td>
</tr>
</tbody>
</table>

do not trigger the prompt. An attacker can use the latter to avoid the browser prompt, or even exploit it to their favor (see Sec. 4.1).

3 Problem Statement

Different login means affect the security of users’ accounts [11, 37]. Reports from Google [20] and Microsoft [62] show that multiple 2FA schemes are widely deployed as alternative logins (users select the 2FA challenge in every login attempt), or recovery mechanisms. Except FIDO, none of the common 2FA is secure against real-time phishing. Previous work on FIDO focused on its usability [14, 24, 30, 73]; limited work questioned its security in real-world deployments, where alternative 2FA and secure recovery are necessary. In practice, account recovery is an expensive operation for service providers [61], and remains vulnerable to social engineering attacks [29, 76]. FIDO specifications focus on authentication, but provide only general recommendations for recovery [31].

To measure the extent by which weaker 2FA alternatives to FIDO are used, we manually inspected Alexa’s top 100 websites, reviewing documentation for websites’ FIDO authentication policy (when available), and creating accounts on those websites that offer public access to test their policy in practice. Results are shown in Table 1; 23 websites (10 organizations) allow choosing alternatives to FIDO. Users of these sites are thus potentially vulnerable to real-time phishing, even when using FIDO. More disturbing, most of these sites force users to first register an alternative 2FA before enrolling their FIDO dongle, and as we show this practice undermines the added security of FIDO.

Google’s Advanced Protection [32] is the only known program where weaker 2FA alternatives are not supported. The program is opt-in (and account recovery does not scale easily to millions of users, see Sec. 8.1), thus not included in Table 1.

4 Downgrading FIDO via Social Engineering

Our attack starts as a typical real-time phishing (see Fig. 1), with the user on the phishing website and the attacker on the legitimate website at the same time. After relaying the user’s credentials (Step 2 in Fig. 1), the attacker is presented with the FIDO-prompt page from the legitimate site (Step 3), and in turn displays a FIDO-prompt page to the user (Step 4).

At this point, the attacker waits until the user authorizes their FIDO token to interact with the attacker’s page through the browser (Step 5). The attacker can leverage standard API functions (e.g., u2f.register and u2f.sign for U2F), so that the attacker is notified when such an authorization-for-interaction occurs. When the browser communicates the result of the challenge-response, the attacker ignores the result of this interaction because all they need to know is that the user has inserted the token. The attacker then chooses, on the legitimate website, to use an alternative second factor method from the list pre-configured by the (victim) user on the website (Steps 6–9), and displays a page prompting the user for that same method (Step 10). Depending on the website, this step can simply be presented to the user without any indication as to whether her FIDO-trial was successful. In our phishing implementation below (Sec. 4.1), we show how Google’s default message to users helps our (attacker’s) cause. Upon getting the token from the user (Step 11), the attacker forwards it on to the legitimate website (Step 12), hence gaining access.

Timing and ordering notes. In Fig. 1, Steps 6–9 can vary between websites; some present the user with options; others may choose for the user. These four steps (i.e., 6–9) must however occur quickly so that the page in Step 10 is displayed to the user right after the user’s FIDO authorization in Step 5. To speed-up displaying the OTP prompt to the user (Step 10), the attacker can initiate Steps 6–9 before 5, so that the OTP prompt (Step 10) is ready immediately after the user’s autho-
zation. However, the delay between Steps 9 and 12 must also be kept small before the website’s OTP token expires. All such steps can be automated, thus delays can be kept minimal.

(a) FIDO prompt generated by the browser. (b) Google Authenticator prompt

Figure 2: Screenshots of the Google’s login page.

Reflections on Step 10. A key element in this attack occurs in Step 10, where the user will be prompted for another authentication factor after using the FIDO token. We believe that seeing three login steps (password + FIDO + OTP) likely sends a false signal to the user that this login trial is even more secure than with two factors (password + FIDO). Google’s OTP page (see Fig. 2), for example, has the sentence “This extra step shows that it’s really you trying to sign in”. When an attacker displays that page after its fake FIDO-prompt, the user would interpret it as an “extra” beyond password + FIDO, but it is intended (by the legitimate site) as extra to only the password. In our implementation, we constructed this (phishing) page with the statement as-is. The “extra step” here enables our attack, as it helps attackers downgrade FIDO to other methods.

Variations to Step 10. Depending on the design of the legitimate website, variations other than presenting a page with an alternative authentication (Step 10) immediately after the FIDO prompt may be more effective in tricking the user. For example, similar to approaches discussed in previous research [77], the attacker may display: “due to technical error, we are unable to process your FIDO token at this time”, or “our FIDO-handling service is currently down, please use another method”. The latter avoids the use of FIDO APIs altogether, so alert messages familiar to the user in the browser-displayed FIDO-prompt box (where attackers have no control over the message within) are avoided.

4.1 Attack Implementation

In preparation for running a user study to test the effectiveness of this attack, we implemented a phishing website that behaves in the manner explained above. The website targets Google’s login page. Details of the user study, including ethical considerations, are discussed in Sec. 5.

Our complete phishing pages are available in [83, Fig. 5]. We obtained the domain two-step.online as our phishing domain, got a Let’s Encrypt certificate for the domain, and placed our phishing pages inside a google.com directory on our server. We intentionally opted for a domain with valid words in a non-traditional TLD such as .online for two reasons: (1) we could get a TLS certificate without being flagged as suspicious [72], and (2) users that do not understand how URLs work but might have a look at it would not be alerted as google.com is present [81, 92]. Our index.html page would get periodically blocklisted every few days, and so we hid it such that it is only accessible through a randomly generated alpha-numeric string. The page would thus only be reachable by a link, which would be emailed to potential victims. While on the phishing website, the browser’s URL bar would have a green padlock icon with the URL:


 Corresponding PHP code at the start of index.php reads:

```php
<?php
    header(‘Access-Control-Allow-Origin: *’);
    if (htmlspecialchars($_GET[“acc”]) != “8FkuX...”){
        echo "This is index.php!";
        exit(0);
    }?>
```

where `acc` is the variable containing the random string.

When implementing our phishing pages, shown in Fig. 3, we did not borrow content from Google’s website; we neither pre-downloaded content from Google to upload to our pages, nor linked to Google content from our pages. The former is not quite straightforward because Google employs code obfuscation techniques on its webpages (e.g., to thwart phishing attacks); the latter was avoided to evade potential phishing detection through analyzing our server’s requests to Google’s web-content [67]. The only object we downloaded and uploaded onto our server was Google’s logo (image). Note that creating our phishing page would be feasible for any attacker with moderate web programming experience. Our implementation of Google’s pages resulted in fewer than 2K lines of combined PHP/JavaScript/HTML/CSS code.

Recall from Sec. 2, the authentic FIDO prompt is typically displayed outside of the attacker-controlled area of the browser to prevent attackers from replicating the prompt within the content pane; note, e.g., for Chrome, the top tip of the box overlapping the URL bar (see Fig. 2a). Recall also that browsers capture the domain from the URL bar and display it to the user within the FIDO-prompt box. It is thus helpful (to the attacker) to use API functions that do not display this box to the user, yet gets the browser to notify the webserver that a dongle was inserted. For Step 10
(a) Attacker’s FIDO prompt.  
(b) Attacker’s Google Authenticator prompt.

Figure 3: Screenshots of the phishing page.

(Fig. 1), we used the `u2f.register` function, which does not display browser-generated prompts. With this function, communications with the user are left to the website developer (i.e., through standard HTML and JavaScript).\(^3\) As an attacker, we do not control the legitimate displayed message; it is browser-generated. So we implemented a mimicry of the Chrome-generated FIDO prompt as a gif image that looks like Chrome’s box,\(^4\) with a message identical to the authentic one: “Use your security key with google.com” (Fig. 3a).

The gif had an animated indeterminate progress bar, almost similar (visually) to Chrome’s authentic one (Fig. 2a). Since it was an image, it was fully contained within the browser’s content pane, located vertically at pixel 0 (top-most point).

Finally, since our aim is only to test the effect of our attack on participants in Sec. 5 (i.e., we do not want to actually steal credentials), we did not implement back-end communication between our phishing website and Google.

5 Evaluation Methodology

We considered several study designs before ultimately developing our methodology, including carefully considering legal concerns [44, 45] and ethical issues as summarized by Finn & Jakobsson [26]. While it does add realism, we dismissed the idea of using participants’ real credentials in a field study without a priori consent because previous work [44] has shown that it can lead to participants feeling violated even after learning that no personal data was compromised. A variation of this approach [45], where a clever study design has participants entering their real credentials to a legitimate site rather than a researcher-controlled site was technically not viable for our particular attack.

Another line of phishing studies [19, 46, 56, 71] ask participants to classify pages as phishing or legitimate without submitting any credential, usually to measure the effectiveness of security indicators. This approach did not align with our intended goal of measuring the effectiveness of FIDO protocol against our phishing attack, thus was not a viable study design. Yet another approach [67, 86] to studying phishing is to analyze logs from service providers to investigate the occurrence of real phishing attacks. To the best of our knowledge the downgrade attack presented in this paper is novel and there are no reports that it has been exploited in the wild, thus logs would be unhelpful. Focusing on user perspectives instead, qualitative studies [16, 46] aim to understand users’ attitudes and reasoning regarding phishing, but these subjective accounts do not provide objective measures of the effectiveness of particular attacks.

We next considered phishing studies based on role playing. Such studies typically use fictitious scenarios to simulate the experience of a user that receives legitimate and unsafe emails. The tasks are typically easy and can be performed by an average user, thus participants do not need the proper experience for the role. Prior studies have used various roles: a university worker (who receives 6 phishing emails out of 14 in total) [78], a political campaign volunteer (3 malicious emails out of 8) [28], a company employee (5 malicious emails out of 19) [53], or a user doing online shopping (gets a phishing email after each online purchase) [22]. Such studies have limitations because participants may behave less securely with mock credentials [77]. However, role playing experiments do not raise ethical concerns and they would allow us to perform a realistic downgrade attack during experiments.

We next considered phishing studies based on role playing. Such studies typically use fictitious scenarios to simulate the experience of a user that receives legitimate and unsafe emails. The tasks are typically easy and can be performed by an average user, thus participants do not need the proper experience for the role. Prior studies have used various roles: a university worker (who receives 6 phishing emails out of 14 in total) [78], a political campaign volunteer (3 malicious emails out of 8) [28], a company employee (5 malicious emails out of 19) [53], or a user doing online shopping (gets a phishing email after each online purchase) [22]. Such studies have limitations because participants may behave less securely with mock credentials [77]. However, role playing experiments do not raise ethical concerns and they would allow us to perform a realistic downgrade attack during experiments.

After careful consideration of the technical requirements of our attack, and the ethical and legal implications of exploiting participants’ personal credentials, we decided to design a role playing experiment followed by a semi-structured interview. We believed that the combination of the two methods would minimize the limitations of either individual method by providing an opportunity for cross-checking our data.

\(^3\)Note that even if a browser-generated box was used, users may already be oblivious to the messages displayed within that box.

\(^4\)An adversary can generate similar prompts for other browsers.
5.2 Study Design

We recruited participants using flyers, university mailing lists, and social media posts. Participants visited a webpage describing the study, its duration, and the compensation before scheduling the interview. The study advertisement generically explained that the purpose of the study was to evaluate and improve the usability of email clients. To eliminate any later doubt by participants about the safety of their legitimate credentials, participants did not use their own email accounts. We provided user accounts and credentials (with a randomly-generated strong password) created specifically for this study. However, to maintain ecological validity, we designed a study scenario that indirectly encouraged participants to think about the security of these accounts.

We ran the study in-person concurrently in two international cities near the end of 2019, one in North America (Ottawa, Canada), below suffixed with -N, and one in Europe (Zürich, Switzerland), -E. To maintain consistency in both cities, we carefully documented the study protocol and had the two researchers running study sessions follow this common protocol. Participants were monetarily compensated for their time, receiving CHF20 in Zürich, and $10 in Ottawa. Participants first completed a demographics questionnaire then they went through the study scenario, during which they were asked to think-aloud (i.e., to describe their thought process out loud). We next gathered feedback from participants through a semi-structured interview.\(^3\) We designed the interview questions to indirectly gauge participants’ awareness of the phishing attempts. Following previously established notions of determining participants’ thoughts [29], we asked:

“If we told you that 50% of our participants access fake websites during their study sessions, do you think you are one of them? Why/Why not?”

We did not ask participants about each email, one-by-one, whether it was a phishing attempt to avoid making them overly vigilant, and potentially biased to answer “yes”. However, we still allowed participants to go back to the emails and check them during the interview, should they ask to do so.

At the session’s end, the researcher provided participants with a debriefing form, explaining the true purpose of the study and answered any questions they had. Each session lasted approximately an hour, throughout which the researcher took notes to provide insight into participants’ thought processes (e.g., if participants hesitated when opening links (phishing or legitimate), if they hovered over the links to view the URL, and any comments they had on the emails). Study sessions were audio-recorded, and the interview portion was transcribed for analysis, and the researchers referred back to the audio recordings for more context when needed. The study received IRB approval in both cities.

5.2.1 Study Scenario

Participants were asked to role play Jordan Hart, a new employee in a technology company on her/his first day of work. They were provided with their company gear: a laptop, smartphone, and a security key (the FIDO dongle). Participants were asked to read and sign the employee on-boarding information sheet (Appendix A), a common practice in industry. This sheet outlined the company policy with respect to safeguarding company information and avoiding scams and phishing attacks, as well as explaining FIDO keys and their associated security benefits in language adapted from Google’s Security and identity products pages [33]. The sheet listed Microsoft Outlook as the company’s primary email provider, included Jordan’s Outlook account credentials (username and password), and provided their Google services credentials. We created real Microsoft and Google accounts. The sheet also included the names and email addresses of Jordan’s manager, IT manager, and HR person, from whom Jordan would receive emails. We created real Microsoft email accounts for each of them. To make sure participants were comfortable using the FIDO key, the researcher–acting as the IT manager–asked participants to login to their email with the key as a second factor, and explained how to use the Google Authenticator app (pre-installed on Jordan’s smartphone) in case of technical difficulties (Appendix A).

Jordan’s Microsoft email inbox contained 15 emails [83, Appendix F], divided into 5 folders, one for each day of the week [83, Fig. 7]. Participants were asked to assume that they login to their Outlook account daily, handle emails received that day (as tagged), logout and shutdown their laptop before going home, and come back the next day to do the same steps. The researcher simulated shutting down the laptop when indicated by the participant by logging-out of their email and clearing the browser cache after finishing each day’s emails. Participants used Google Chrome.

5.2.2 Emails

Four of the 15 emails were phishing, containing a link to our phishing website (Sec. 4.1). Such emails were spearphishing (targeted). We chose to have such a high number of phishing emails in the first week of employment to give participants a higher-than-normal chance to recognize our phishing attack. In Sec. 6.2, we explain how detecting a single phishing email suffices for us to count the participant amongst those who did not fall for our attack.

We used PHP’s mail function to send out the phishing emails using a spoofed source email address. To ensure realism, these emails included errors like grammatical mistakes and typos, mimicking typical phishing emails. Non-phishing emails were sent from the authentic email accounts of the company’s employees (Jordan’s manager, IT manager, and HR person) through the email web client. All emails were sent at once before we started recruiting participants, and sim-

---

\(^3\)The interview script is detailed in [83, Appendix E].
We used the Qualitative Content Analysis Methodology [47] to analyze qualitative data collected throughout the study (e.g., post-testing interview scripts, and researchers’ notes including participant’s comments on the content of emails and whether they entered their credentials to the website linked in the email). We developed an analysis matrix to cover the main topics relevant to our research questions. The matrix comprised of four categories with which we coded our data: identifying phishing links, participants’ perception of FIDO, their perception of 2FA, and their security attitude and awareness. We then followed an inductive analysis method, and performed open coding to look for interesting themes and common patterns in the data using NVivo software. Themes irrelevant to our research questions are not discussed herein. As recommended by previous work [13], and similar to previous research (e.g., [16]), data was coded by a single researcher with considerable experience in Human-Computer Interaction (HCI), security, and qualitative data analysis, so that this researcher would perform rigorous analysis by being immersed in the data. Two researchers met regularly to discuss the codes and interpret the data. To verify the reliability of our coding, we had a second researcher code 25% of the data individually. We calculated Cohen’s Kappa coefficient [15] to determine inter-rater reliability, which indicated “almost perfect agreement” [54] ($\kappa = 0.82$).

6.1 How to assess phishing susceptibility?

To identify participants who could be victims to our attack in practice, we need a mapping between their behaviour in the study and their attack susceptibility in practice. Simply classifying those who submit their credentials to one of our phishing links as potential victims may not be accurate because: (1) participants may not be as keen to protect their study credentials as they would their own, and (2) participants may think they need to process all emails regardless of their suspicion because this is what the study is asking them to do. We took measures to reduce the impact of both points, e.g., through emphasizing the importance of security to the persona’s employer, and integrating actions with interview responses. Only two participants mentioned they were not paying attention because they thought it was the study’s instructions, highlighting the importance of our measures.

Participants’ actions (during the study) and awareness of the attacks (in the study) may or may not align. We use both parameters, i.e., actions and awareness, to assess each participant’s phishing susceptibility in practice; combining both parameters yields four possible cases, which are summarized in Table 2 (first 3 columns). According to these two parameters, we will rate each participant as susceptible, potentially susceptible, or not susceptible to our phishing attacks in practice (column 4 in the table). (Details on how we determine a participant’s awareness of the attacks in our study can be found in Sec. 6.2 below.)

Normally, a participant who is unaware of our phishing attempts would submit their credentials to our phishing website. This is Case 1 in the table. A vigilant participant would normally refrain from submitting their credentials, and confirm their awareness of phishing attempts in the post-study
interview—Case 4.

Cases 1 and 4 are straightforward; we classify the former as “susceptible to phishing”, the latter as not. We classify Cases 2 and 3 as “potentially susceptible to phishing”. In Case 2, although they did not submit credentials, participants are unaware of any phishing attempt. In Case 3, participants are classified as aware, yet they submit credentials.

6.2 How to determine awareness of phishing?

Determining participants’ awareness from the interview is not trivial. Participants responses’ varied greatly. For example, to the above question (“If we told you that 50%...”), some participants answered affirmatively, but only name examples of non-phishing emails. Others answered affirmatively, but said they did not remember which ones were phishing. We also had participants who first denied being in the 50% that accessed fake sites, then hesitated, alternating between “yes” and “no”, then changed their minds, and gave a few true phishing examples. And there were participants that provided an immediate affirmative response, reconsidered, and finally decided there were no phishing emails. We thus ignored their direct ‘yes/no/maybe’, and instead relied on objective portions of their comments to assess awareness, as described next.

Figure 4: Determining awareness of our phishing attempts.

Any participant who (i) identified at least one phishing email or (ii) named a true phishing indicator is classified as aware-of-phishing-attempts, regardless of what else was said during the interview. Participants classified as unaware-of-phishing-attempts included those who: just denied being in the 50%; affirmed being in the 50%, but gave only examples of non-phishing emails; or affirmed but gave only false phishing indicators. Figure 4 shows this criteria, alongside common example responses in our study that we discarded because the awareness criteria was met. By true phishing indicators, we mean the website’s URL, and commonly agreed upon (though non-robust) signs of phishing emails [40], like typos, lack of context, and grammatical mistakes. Unencrypted email is an example of a false phishing indicator.

Conservative classification of attack awareness. Following the criteria in Fig. 4, we classified participants as aware-of-phishing-attempts even in situations where it is hard to tell whether they were truly aware. Thus we provide an upper bound on awareness. For example, a participant who named a true phishing indicator, yet asserted seeing no phishing emails is still classified as aware-of-phishing-attempts. Classified likewise is a participant who gave an example of one phishing email, mistakenly identified two non-phishing emails, and asserted there were no other phishing emails (i.e., missing three others). We used conservative criteria for two reasons: (1) we increase certainty that participants classified as unaware-of-phishing-attempts would most likely be unaware of similar attempts in practice, and (2) participants may have forgotten which emails were truly phishing by the time they reach the post-study interview (there were 15 emails in total). We purposefully avoided showing each of the 15 emails to participants and asking them which were phishing to avoid priming. Our hypothesis here is that, if during the study, a participant suspected a phishing attempt, they would recall that and indicate it in a manner captured by the criteria in Fig. 4.

Table 2: Classifying participants’ susceptibility to our phishing attack in practice, from their study behaviour.

<table>
<thead>
<tr>
<th>Case</th>
<th>Participant</th>
<th>submitted credentials</th>
<th>Susceptible</th>
<th>Results # %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Unaware</td>
<td>Yes</td>
<td>Yes</td>
<td>28 55</td>
</tr>
<tr>
<td>2</td>
<td>Unaware</td>
<td>No</td>
<td>Potentially</td>
<td>1 2</td>
</tr>
<tr>
<td>3</td>
<td>Aware</td>
<td>Yes</td>
<td>Potentially</td>
<td>17 33</td>
</tr>
<tr>
<td>4</td>
<td>Aware</td>
<td>No</td>
<td>No</td>
<td>5 10</td>
</tr>
</tbody>
</table>

Conservative classification of attack susceptibility. We determined susceptibility based on two factors that we first assessed independently: awareness and submission of credentials. “Awareness” is not per email, but per participant. So even if a participant named one phishing email but missed all others (or asserted there were no others), we still classify them as aware-of-phishing-attempts. When we check whether this participant submitted credentials to our phishing website, we do not match the phishing email they fell for in the study with the email they named in the interview. For example, a participant who noticed only one phishing email, E2, is classified as aware-of-phishing-attempts, even if they asserted there were no others. If this participant submits credentials upon clicking the link in any phishing email (E2 or another), we classify them as “potentially susceptible”, not as “susceptible”. One would argue that this is a “susceptible” participant because an email successfully phished their credentials. Being conservative, we opt to use any minor indication that a participant might notice similar attacks in practice as grounds for avoiding classification as “susceptible”. By classifying only the most blatant case as “susceptible”, we provide a lower bound for susceptibility to attacks.

Examples of aware participants. From our analysis, the following are examples classified as aware-of-phishing-
I just went to hotmail, the outlook website which I very often + 2) were classified as “unaware of phishing attempts”, and And I look[ed], everything looks like pretty work-related and (Case 1) of participants would be susceptible to our phishing Table 2 summarizes the results; 57% of participants (Cases 1 everything seemed legitimate enough and seemed business-y. I got from Sam Logan on the Friday was definitely a phishing (50%)[...] I was taking it for granted that the emails I was very rapid in going over the emails. She did not click on any (44). We thus discuss the amalgamated results, within the context of the two research questions in Sec. 1.

**Examples of unaware.** P10-N said, “I don’t think so. [...] Everything seemed legitimate enough and seemed business-y. And I look[ed], everything looks like pretty work-related and exactly related to what the e-mail said it would be. Yeah. It wasn’t like I just clicked on a link and it really brought me to some random page or something, it was related to what the e-mail was saying. So it seems legitimate to me.”. P13-N said, “I just went to hotmail, the outlook website which I very often go. And I logged in from there. So I think it seemed fine.”

7 Results

Through our data analysis, we looked for an effect of location by comparing data from the North American and European cities where the studies were conducted; we found no clear distinctions between the two groups. Our qualitative analysis did not reveal themes distinct to either city and we found no statistically significant difference between the two groups’ susceptibility to phishing attacks ($X^2(2, N = 51) = 1.64, p = .44$). We thus discuss the amalgamated results, within the context of the two research questions in Sec. 1.

7.1 Phishing Susceptibility with FIDO (RQ1)

Table 2 summarizes the results; 57% of participants (Cases 1 + 2) were classified as “unaware of phishing attempts”, and all but one of these participants (P12-E) submitted credentials to our phishing website. Given our conservative measures in classifying susceptibility, our results suggest that at least 55% (Case 1) of participants would be susceptible to our phishing attacks in practice. The one participant in Case 2, P12-E, was very rapid in going over the emails. She did not click on any phishing link, and also ignored several non-phishing links. She gave very short, non-informative, responses in the post-study interview. When asked why she did not click on links in the emails, she simply said, “There is no particular reason”. In contrast, 43% of participants (Cases 3 + 4) were classified as aware of phishing attempts, but only the Case 4 participants (10% of all participants) are likely to detect the discussed phishing attempts in practice. A Fisher’s Exact Test suggests that participants who were aware of phishing attempts were less likely to submit their credentials than those who were unaware ($p = .047, N = 51$). Contrarily, Fisher’s Exact Tests showed that neither gender ($p = .60, N = 51$), generation⁶ ($p = .34, N = 51$), nor having work experience ($p = .23, N = 51$) significantly influence susceptibility to phishing.

7.1.1 Takeaway

Our focus in the present paper is to determine user’s susceptibility to phishing, particularly while using FIDO. We noticed that all participants who appear to have detected and avoided our phishing attempts (Case 4) would have done so also without using FIDO. The phishing indicators they mentioned, and the reasons they discussed as to why they avoided submitting credentials to our phishing site are not related to FIDO. Likewise, those whom we classified as susceptible to phishing are susceptible despite using FIDO. That is, using FIDO did not protect them from our downgrade attacks. Essentially, what we were looking for in this research is cases of users who would have fallen for phishing without FIDO, but have not because of using FIDO. We found none.

7.2 Phishing detection while using FIDO (RQ2)

When we asked participants if they had accessed fake websites during their session, participants were evoked to think about the emails more deeply, and discuss reasons they used to classify emails as phishing or safe. Through our analysis of qualitative data, we found that participants relied mostly on general phishing indicators for determining whether the emails they received were phishing. Table 3 summarizes reasons why participants classified an email as phishing (seven), and reasons for classifying an email as safe (eleven).

7.2.1 Reasons for classifying emails as phishing

We grouped the phishing indicators discussed by participants into two categories: technical, and non-technical. The two technical indicators were suspicious URL, and that the repeated login prompts were unusual behaviour. P2-N explains, “It wants [me] to log onto Google even though I was already logged on to Google on just another tab [...] This was not a thing I noticed at the beginning when I was doing the experiment [...]. Now that I’m thinking about it. Yeah, makes sense, right? Like why are they asking you to log onto Google again when you’re already logged onto Google?”

⁶We assigned generation labels, based on participants’ ages, as in [49].
Although the identified technical indicators can alert users, they are not ideal from a usability perspective. Users do not always check the URL bar [42], and even if checked, users do not necessarily know the correct URL [6]. Users may also lack a proper understanding of the structure of URLs to be able to assess their legitimacy [6, 19, 92]. In addition, it is unlikely that typical end users would know that a website would only require users to re-login if the session cookie has expired (the three participants discussing the repeated login indicator had studied Computer Science or Computer Security). In fact, repeated login prompts can provide users with a false sense of security, because the system appears more secure if it asks for two factors at every login. P22-N explained, “I think when I clicked on the one that didn’t ask [...] for the two factor authentication, it just went straight to the [Google] Doc then I [thought] something bad had happened. Because every other time [...] it would ask me to sign in again [...] but that one suddenly didn’t. And it just makes me kind of feel like ‘uh-oh what happened?!’”

Participants also discussed non-technical indicators, such as poor grammar and styling, and the presence of a hyperlink in the email. Other non-technical indicators stem from the participant’s knowledge of the sender, including that the tone of the email was unexpected, the language of the email did not match that of the sender (sender language consistency), and that the email was out of context. P16-E said, “If it is just, like my boss sending a book to download, and we talked about it, it’s fine. But if it is a random book, then it’s weird.”

Non-technical indicators rely exclusively on users’ judgement, are inconclusive, and are relatively easy to manipulate. For example, while context appears to alert some users against phishing (as described herein and in previous work [46]), attackers can adjust their techniques to increase their phishing emails’ credibility [84]. In fact, an email can have an expected tone, a language that appears consistent with the legitimate sender’s, relevant context, proper grammar and styling, no hyperlink, and yet belong to a phishing campaign.

None of our participants mentioned relying on FIDO to identify phishing emails, and rightfully so. It is not quite clear how FIDO can be used as a complete phishing indicator—it may benignly fail due to technical errors, e.g., broken dongle.

7.2.2 Reasons for classifying emails as safe

When examining reasons why participants identified emails as safe, we found that they again relied on technical and non-technical indicators (Table 3). As many participants ended-up misclassifying emails as safe, such reasons may have misguided participants. In these cases, participants erroneously interpreted the URL (linked in the email) as a safety indicator. For example, some participants concluded they were safe from phishing because opening the links in the emails did not lead to obviously malicious behaviour (e.g., popups). Others \(n = 3\) indicated they “felt more secure with 2FA” (P23-E) and were protected against phishing because they were using FIDO (using FIDO/2FA). Interestingly, requiring the use of Google Authenticator (part of our attack) gave some participants a false sense of security. P21-N said, “I had to put in the information [code] as well and I felt secure: the company even took me to verify everything [using the Google Authenticator] to make sure that it was secured.” These participants either viewed the authenticator an additional factor or assumed it was part of FIDO, and some even thought FIDO was more secure because of the authenticator. P10-E explained, “If you have to use the authentication app on the phone, with the changing number always, it is really difficult for someone to hack your system to find this kind of information.”

Non-technical reasons also emerged for classifying emails as safe, based on context and the user’s expectations (e.g., their social and professional life). For example, users expect to receive emails from their institutions’ official communication channels, which gives these emails credibility. This finding mirrors that of Conway et al. [16] where participants felt more secure at work. We also found, similar to previous work [6], that participants relied on quickly inspecting the login interface or the content (e.g., a Google Sheet) to which the link in the email redirects, and comparing it to their expectations. P18-N explained, “I think I did a little bit of due diligence when I signed in, so [I] should be OK. Like I checked when I was logging in [that] I was logging in to the right thing. Most of the things that came up were Gmail and Outlook. The only one document that [I] opened was a Google document which did ask for my authentication” (as part of our attack).

In summary, we highlight that participants cited FIDO as one of the reasons for classifying as safe, when they have in fact fell for our phishing attack. As such, FIDO did not only fail to protect them, but it potentially gave them a false sense of security. FIDO can be relied upon as a safety indicator when it works successfully, without having the user authorize any other factors (beyond the initial password).

7.2.3 Takeaway

Despite using FIDO, we noticed that none of the participants have relied, or indicated that they would rely, on FIDO for detecting phishing attempts. In contrast, we had three participants who said they were secure because they used FIDO in all their logins, even when some of these were accompanied by other authentication factors. Evidenced by our attacks, the proper usage would be to refuse to login with alternative methods if a user has enabled FIDO. Seeing a FIDO-only login is practically opposite to using FIDO alongside other factors—the former prevents downgrade attacks, the latter enables them. We found no evidence that any of our 51 participants understood this concept.
We provide practical insights regarding potential defenses, A straightforward countermeasure to the downgrade attack of politicians or journalists. The program is opt-in and users e.g., the program achieves this for critical accounts, enables FIDO. This would have mitigated situations where presented herein is to disable alternative 2FA methods if a user based on our study results and our analysis of the attack itself.

8.1 Disable Weaker Alternatives
A straightforward countermeasure to the downgrade attack presented herein is to disable alternative 2FA methods if a user enables FIDO. This would have mitigated situations where our participants thought the extra factor was a feature rather than an indicator of attack. Google’s advanced protection program [32] achieves this for critical accounts, e.g., those of politicians or journalists. The program is opt-in and users must register at least two security keys, one for daily use,7 and others as backup. Google does not detail the recovery process in case both keys are unavailable, but states that “it may take a few days to verify it’s you and restore your access”. This delay poses a major trade-off for users.

Limitation: non-scalable recovery. Doefler et al. [20] report that challenges requiring security keys have a lower pass rate than device-based ones. So, if alternatives were disabled, more users would need the recovery process. On the other hand, such recovery adds significant costs to service providers, and does not scale to millions of users [61]. Disabling weaker FIDO alternatives comes at the cost of non-scalable recovery.

Limitation: usability impact. Previous literature [14, 24, 25]...
recently reported that users have difficulties enrolling security keys into their accounts, and are concerned about being locked out if keys are lost. Registering multiple keys can enhance the user experience but may be costly for users, which might be a barrier to some users. Moreover, service providers tend to facilitate user onboarding and enhance overall experience by offering a variety of channels to connect to its backend, e.g., browsers, native apps on different OSes, or third-party software such as email clients. Disabling FIDO alternatives can degrade usability because channels that do not support FIDO should then be dropped—otherwise, the attacker connects to the server through such channels.

## 8.2 Risk Based Authentication

Risk-based Authentication (RBA) refers to a set of server-side techniques to assess the risk of an authentication attempt, and block malicious ones [35, 79, 89]. Secure IP geolocation [1, 2], device, network, user agent, and installed plugins are examples of metadata that RBA systems analyze for deciding the risk score of a login attempt. A low risk attempt (e.g., same user agent and same IP address) gives confidence to the server that the honest user is authenticating. For higher risk requests, the server challenges the user to provide additional factors, or restricts user’s access depending on the provider’s policy [90].

**Limitation: mimicry of user’s attributes/behaviour.** A recent study [12] shows that attackers have already developed malicious tools that can circumvent RBA defenses. Such tools are made available as public services. Campobasso and Allodi [12] reveal that attackers collect necessary data from victims on top of their credentials, so they can bypass RBA defenses. Similarly, an adversary performing real-time phishing can adapt such tools to bypass RBA mechanisms on-the-fly. This adversary has a connection with the victim’s browser, and may be able to mimic attributes/behaviours to the legitimate website [3], or execute the JavaScript code (related to RBA analysis) directly on the victim user’s browser.

## 8.3 Browser Hints

The recent WebAuthn API [8] instructs browsers to always show a prompt window when a website interacts with the authenticator during both registration and authentication. The prompt is part of the user consent process, which means that the user agrees (by tapping the authenticator device) to complete the request displayed on the prompt. The prompt itself contains a short message, and browsers display it as a native popup that extends slightly above the address bar. For example, Google Chrome captures the TLD and second-level domains of the website (e.g., google.com), and displays them to the user within the prompt box alongside the message:

### Use your security key with google.com

Mozilla Firefox includes the fully qualified domain name (e.g., accounts.google.com) in a callout panel as:

### accounts.google.com wants to authenticate you using a registered security key. You can connect and authorize one now, or cancel.

Since the prompt contains a short message and the website’s domain, rendered in boldface in Firefox, it can potentially alert visitors of a phishing website.

**Limitation: users’ susceptibility to social engineering.** Relying on users to notice the domain mismatch is not a reliable countermeasure for three reasons. First, FIDO promises to relieve users’ from the burden of detecting phishing, hence security should not depend on prompts or visual indicators. Second, previous research [6, 7, 19, 81] have shown that users typically do not pay attention or understand browser hints related to security. Third, the adversary can use the U2F API to interact with the device, which does not trigger such prompt windows (as we did in our implementation—Sec. 4.1) and although not tested in our study, previous work shows that users easily overlook the absence of security cues [77].

## 8.4 Secure Login and Recovery Alternatives

Doefler et al. [20] discuss Google’s categorization of login, second factor authentication, and account recovery methods. Methods of comparable security are placed in the same category, and should be allowed depending on the account’s security status. Such a status could possibly be indicated by the user’s security configuration (e.g., enabled 2FA, configured robust recovery methods).

**Promising countermeasure.** It appears that a viable countermeasure to the attacks discussed herein is: when FIDO is enabled, only enable authentication (or 2FA) alternatives that provide resilience to similar attacks that FIDO resists. Suitable candidate alternatives include other FIDO protocols. For example, a phone-based authenticator through FIDO2 can serve as an alternative to physical security keys. This should be recommended/enforced by service providers (websites). Intuitively, a user choosing to register a security key for login is implicitly requesting resilience to advanced attacks (e.g., real-time phishing). To that user, a service provider should only allow alternatives of equivalent defence capabilities.

Login and recovery are two sides of the same coin—recovery alternatives must also match the security level of the user-configured login methods. For example, for a FIDO-enabled account, recovery through a secondary email that has weaker security undermines the security of that account. Hammann et al. [37] discuss how account-access graphs could help users and service providers discover vulnerable paths.

---

8 At the time of this writing, security keys from Yubico (a popular vendor and FIDO Alliance partner) cost around $20.
8.5 User Education

Many participants in our study relied on wrong phishing indicators. Several reported that once they click a link in an email, they wait to see if the visited page is rendered correctly. Participant P9-E said, “[...] if the website looks fine, I mean the front page, I am not suspicious”. Similarly, P20-E mis-classified the phishing website as legitimate: “It’s the same because it looks the same up here [refers to logo section], and I would be trusting it’s fine”. Others rely solely on the email itself; participant P18-E said: “I decide before whether to click or not, and once I click it, it’s opened (done)”. When asked if she continues checking the visited website, she added: “Not really”. This is not new; the fact that phishing and similar social engineering tactics rely on users’ lack of understanding or incomplete mental models is well established [6, 19]. So long as authentication relies on user actions, education remains a plausible (yet somewhat ineffective) countermeasure.

Limitation: impracticality. When available, education can help users form reasonable mental models of phishing and may help them detect some attacks. This is an incomplete countermeasure, however, because it assumes that users are capable of continuously devoting all of their attention to this task and that all attacks will have user-noticeable indicators. Security is rarely users’ primary task [88] and humans are incapable of remaining highly vigilant 100% of the time. As we saw in our study, some attacks include sufficient contextual indicators, either by design (e.g., spear phishing) or by coincidence, to trick even an attentive user who is actively evaluating for security [6].

Educational campaigns or marketing efforts that promote security keys as phishing resistant can further contribute to users forming incorrect mental models [69], and can thus have adverse effects as users develop a false sense of security and become less attentive to attacks.

9 Related Work

Phishing is a widely-studied attack vector from the social engineering category. It uses effective techniques to take over accounts, fooling even knowledgeable users [19, 26, 43–45].

2FA schemes. The industry and the academic community has developed several 2FA schemes [48, 50, 57, 58, 68] to protect users’ accounts. However, real-time phishimg is still very effective to bypass 2FA and automated tools [34] make such attacks simpler, cheaper, and easy to scale. Previous works [23, 63] report that phishing is widely employed and preferred by malicious actors, even at hack-for-hire services [63].

FIDO is based on public key cryptography [17] and its benefits are already demonstrated in a company setting [55]. The protocol itself is considered secure and it is promoted by the industry as being foolproof phishing-resistant [33]. The research community has focused on the usability aspects of FIDO [18, 70, 74] but have not questioned its security in real-world deployments. However, the necessity for alternative 2FA is already emphasized on previous studies [20, 73] because users cannot always complete the FIDO step. On the users side also, the possibility of being locked out is reported as the main obstacle for using FIDO in daily routine [24, 30].

Anti-Phishing ecosystem. Service providers, browser vendors, and other entities have developed an ecosystem to detect and prevent phishing, however adversaries adapt their tools continuously and evade such systems [64–66, 95]. Oest et al. [67] report that a phishing campaign is detected nine hours after the first victim, hence spear-phinshings that target individuals are much more difficult to be prevented by the ecosystem. Another line of work [25, 39, 59] try to detect malicious websites based on the URL analysis. Email providers have developed frameworks to filter out phishing emails before reaching users [21, 38], however attackers still find their way to their target’s inbox [63]. To limit the consequences of password reuse [10, 27], prior works have proposed frameworks that allow servers to learn when a password is compromised [80, 87], while [52] shows that secure implementation of critical protocols, such as TLS is not trivial for developers.

Client side. Password managers are a possible countermeasure to phishing attacks because credentials are released only if the user visits the correct domain. Blanchou and Youn [9] were among the first to report vulnerabilities in password managers. Others [36, 75] describe the challenges of designing and implementing secure extensions, while [93] reported that spoofing the sidebar is effective in phishing the master password as well. Yang et al. [94] measured the effectiveness of browser indicators, while [71] show that users lose the ability to detect phishing some period after training.

10 Concluding Remarks

OTP-based 2FA schemes are amongst the most common phishing defences. Being replayable [3], they fail to defend against real-time phishing, where the adversary relays user-submitted OTPs to the legitimate site in real-time. The FIDO alliance has designed challenge-response mechanisms with browser involvement, which enables the inclusion of a website’s URL in the challenge. Relaying the response becomes useless, and real-time phishing is thus defeated. U2F is one such standard, where the response is computed on a hardware token. To handle token loss/malfunction, websites often allow/force users to register 2FA alternatives to FIDO. All FIDO-supporting websites in Alexa’s top 100 adopt the practice. We ran a user study to test whether a phishing attack that downgrades FIDO to weaker alternatives is effective. Although the study tested U2F tokens, findings (particularly regarding downgrade effectiveness) can extend to other relevant FIDO specifications. We make the following four remarks.

User studies that evaluate attacks must be gracefully executed. Evaluating attacks through user studies is challenging. Participants may fall for said attacks during the study, not
because of successful deception, but rather due to participants’ lack of investment in protecting assets or misinterpretation of study requirements. If participants’ actions were the sole metric, we would have misidentified 88% (instead of 55%) of participants as susceptible to our attacks (Table 2). Such studies followed by semi-structured interviews delicately gauging explanations of participants’ actions, without calling attention to said attacks, provide richer results.

Even with FIDO, users remain susceptible to real-time phishing that downgrades FIDO to weaker alternatives. Most participants failed to detect our phishing attacks. Those who succeeded (10%) have done so without FIDO’s help. We found no case where a participant was close to fall for would either expose users to lockouts due to token losses, or who succeeded (10%) have done so without FIDO’s help. do not understand how FIDO protects them. typically prioritize usability for everyday use. performed less frequently, whereas standard 2FA schemes continue to make users potentially susceptible to other social tacks, websites should only allow alternatives of comparable security. Many of the other countermeasures we explored as a three-factor login, giving an increased sense of (false) security rather than arousing suspicion. The effect of such attacks in practice is exacerbated by two points: (1) users can become less careful seeing more factors, and (2) reassuring wording on login pages (e.g., Google’s statement on 2FA pages “This extra step shows that it’s really you trying to sign in”).

Despite understanding how to use FIDO [30], users do not understand how FIDO protects them. While discussing how they detected our attacks, no participant mentioned relying on FIDO. FIDO protects users when login is granted after using only FIDO, not after using FIDO plus other factors. The former prevents real-time phishing and downgrade attacks, the latter enables them. As it is counter-intuitive, no participant appears to have assimilated this concept.

Enabling only FIDO alternatives to FIDO is an effective countermeasure. To address the necessity of allowing alternatives to FIDO’s U2F, without enabling downgrade attacks, websites should only allow alternatives of comparable security. Many of the other countermeasures we explored would either expose users to lockouts due to token losses, or continue to make users potentially susceptible to other social engineering variations. Relevant FIDO specifications that are also resilient to real-time phishing (e.g., CTAP2) appear to be suitable alternatives from a security perspective.

Actionable Takeaways. We call upon the FIDO alliance, its industry partners, and the security research community to undertake and promote the following two actionable items as applicable: (a) to pursue efforts to inform, educate, and design technologies that persuade users who have configured security keys to only use such keys for login; and (b) develop new recovery schemes that are phishing-resistant and scalable to millions of users. For the latter, such schemes may prioritize security guarantees over usability, as recovery is normally performed less frequently, whereas standard 2FA schemes typically prioritize usability for everyday use.

Acknowledgments

Abdou acknowledges funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery Grant. Chiasson acknowledges funding from the NSERC Discovery Grants and Canada Research Chairs programs. We thank the anonymous reviewers and our shepherd, Kent Seamons, for their insightful feedback. We also thank Sebastian Navas Chaparro for his assistance in collecting our Canadian data sample.

References


A Employee On-boarding Information

NanoTech

Employee Onboarding Information

Welcome to NanoTech! We're excited to have you!

At NanoTech, we're committed to protecting the company, employees, and customers' resources, internal and external networks, and sensitive data.

Our goals:
- Safeguard NanoTech confidential information, employee information, and our customers' confidential information
- Ensure uninterrupted and efficient operations at NanoTech
- Protecting NanoTech against scammers, including phishing attacks
- Comply with industry, regulatory, and customer requirements

Your role:
- Report theft, loss, or unauthorized disclosure of NanoTech information
- Report attempts for stealing NanoTech information, including suspicious phishing emails
- Adhere to copyright, trade secret, patent and IP laws
- Log off from your email account(s) at the end of your work day

As part of your onboarding process, you'll receive your work devices and credentials. Reach out to the IT department if you have any issues.

You will be using two-factor authentication to authenticate network resources, computer resources, and Google services. For authentication, you will need your password and your security key. If you lose your security key or if it was damaged, you can log in to your account using other backup mechanisms, e.g., using the Google Authenticator app already installed on your work phone. In case of loss or damage to the security key, please reach out immediately to the IT department to replace your key.

Why the FIDO security keys?
A FIDO security key is a phishing-resistant two-factor authentication (2FA) device. FIDO keys use cryptography to provide two-way verification: it makes sure that you are logging into the service you originally registered the security key with, and the service verifies that it's the correct security key as well. This provides superior protection to code-based verification, like SMS and one-time password (OTP).

We've already registered your security key with your accounts! Start inventing!

Please sign here to indicate that you have read this information and agree to adhere to it.

Signature/Date

NanoTech

Jordan Hart
Email: jordan.hart540@hotmail.com
Password
(Email is the primary method of correspondence in the company)

Google services account
Username: jordan.hart540@gmail.com
Password
(You will need to use two-factor authentication for Google services)

Manager: Alex James (alexjames123@hotmail.com)
IT manager: Sam Parker (sam.parker000@hotmail.com)
HR correspondence: Sam Logan (sam.logan2019@hotmail.com)