TORPEDO: TOoltip-poweRed Phishing Email DetectiOn

Melanie Volkamer, Karen Renaud, Benjamin Reinheimer

TORPEDO: TOoltip-poweRed Phishing Email DetectiOn

Melanie Volkamer[1,3], Karen Renaud[2], and Benjamin Reinheimer[1]

[1] SECUSO, Computer Science Department, TU Darmstadt, Germany
firstname.lastname@secuso.org
[2] School of Computing Science, University of Glasgow, United Kingdom
karen.renaud@glasgow.ac.uk
[3] Karlstad University, Sweden

Abstract. We propose a concept called *TORPEDO* to improve phish detection by providing just-in-time and just-in-place trustworthy tooltips to help people judge links embedded in emails. *TORPEDO*'s tooltips contain the actual URL with the domain highlighted and delay link activation for a short period, giving the person time to inspect the URL before they click. Furthermore, *TORPEDO* consists of an information diagram to explain phish detection. We evaluated *TORPEDO* in particular with respect to its effectiveness: Compared to the worst case 'status bar'. as used in Thunderbird and Web email clients. *TORPEDO* performed significantly better in detecting phishes and identifying legitimate emails (85.17% versus 43.31% correct answers for phish). A proof of concept implementation is available as a Thunderbird Add-On.

1 Introduction

Phishing is merely a modern equivalent of a confidence trick that has been carried out for centuries: to deceive someone to derive some personal benefit. The first time that the term "phishing" was used to refer to this digital version of confidence tricking was on January 2, 1996[4]. Phishing messages offer a link embedded in an email message that entices the recipient to click. Email recipients are likely to click on links due to their widespread legitimate use. If they do click, it redirects them to a website masquerading as the real thing or downloads some malware onto their computer. Twenty years after its emergence, phishing still succeeds [11, 39].Automated detection is a powerful tool against phishing, but the fact that it takes, on average, 28.75 hours to detect new phish websites [2] means that users have to detect phishing messages themselves during the discovery window. However, many people are unable to distinguish legitimate from phish messages. Since there is no financial bar on the number of emails phishers can send, this means a sizeable number of people are snared every day.

The goal of our research was to propose a solution to reduce phishers' success in the email environment significantly (note, we studied the approach in

[4] The mention occurred in `alt.online-service.america-online`.

the email environment but it could be easily adopted to other message formats such as Facebook and Twitter messages). To achieve this goal, we needed first to understand why people fall for phishing. We thus carried out a literature review and a cognitive walk-through analysis of emails as displayed by commonly-used desktop and webmail clients. Based on our findings, we proposed a concept called *TORPEDO* (**TO**oltip-powe**RE**d **P**hish **E**mail **D**etecti**O**n) to assist users by providing a *just-in-time*, *just-in-place, trustworthy* tooltips that display the actual URL with the domain *highlighted* in bold (see Fig. 1). Furthermore, we disable the link briefly, introducing a short delay, to increase the likelihood that people will check the URL before clicking on it. Finally, we provide users with an *extended information diagram* to explain the phish detection process. An evaluation delivered significant improvements (85.17% for phish, 91.57% for legitimate emails compared to 43.31% and from 63.66% when only providing the URL in the status bar, as Thunderbird does it for instance). We implemented a corresponding Thunderbird Add-On, as a proof-of-concept.

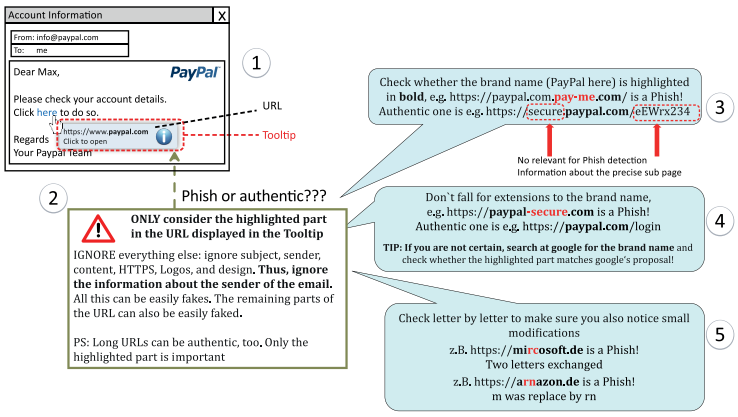


Fig. 1: Just-in-time, just-in-place, trustworthy tooltips as shown in the upper-left part. The entire figure is displayed when more information is requested.

2 Identifying ons Why People Fall for Phish

It is important to understand why people fall for phish in order to support them the better. To identify possible reasons for people falling victim to phish, we carried out a literature review and conducted a cognitive walk-through analysis.

2.1 Literature Review

A phishing email contains a number of signals that may indicate that the email is a phish, the most reliable of which is the actual URL as explained in [15] and [26].

Many people do not realise this but, just in case they do, phishers routinely obfuscate the URL to dampen down this signal (e.g. using amazon.shop-secure.com to phish amazon accounts). Our literature review revealed a number of papers in which the authors showed that the reality is different since many people focus on other signals, applying various flawed heuristics, namely:

- *The Sender:* People are likely to trust emails from friends [16] or from reputable businesses [41].
- *The Look and Feel:* People judge emails' trustworthiness based on their first impression, informed by a recognisable logo [4, 30], attractive design [30, 35], the use of their name or the provision of the company's contact details [17].
- *The Email Content:* People read the email in order to judge the trustworthiness thereof. Relied-upon indicators are the grammar and spelling quality [35, 30]. Researchers also showed that people are more likely to fall for a phish when: emotions such as excitement, fear or anxiety [4, 35] are provoked, a sense of urgency is invoked [30, 34], existing attitudes and beliefs (wanting to believe that the scammer is honest) are exploited [32], or persuasive and influencing techniques are used [35, 38]. Researchers argue that under such conditions of arousal people's decision-making abilities are impaired and they are less likely to pick up danger signals [37].
- *Wrong Parts of the URL:* Some people do look at the URL [17]. However, they misinterpret the URL due to a lack of knowledge of the semantics of URLs [9, 40]. Some people are reassured by the mere presence of `HTTPS` in the embedded link and look no further [14]. Others are reassured by the brand name being embedded 'somewhere in the URL' [17].

2.2 Cognitive Walk-through Analysis

For one week, we carefully considered emails that we received, examining the embedded URLs to identify possible challenges which could impair or encourage phish detection. We examined Thunderbird and Apple Mail as well as Web interfaces from three popular Web email clients. We made the following observations:

- *Information not provided where expected/needed:* Thunderbird[5] as well as the Web email clients, display the actual URL destination in the status bar at the bottom of the window. *Problem:* The status bar is some distance away from the user's current attentional focus and might easily be missed. The text of the email is far more prominent and thus likely to be focused on.
- *Tooltip provided by sender:* Some email senders provide a tooltip which appears when the mouse hovers over the link when using Thunderbird or the Web email clients[6] while the actual URL is still displayed in the status bar. Providing such tooltip is actually a very simple attack since the phisher only needs to provide a "title=" attribute. *Problem:* The tooltip encourages examination of the URL by appearing where the user's attention is focused.

[5] Note, this is different for Apple Mail and also for Outlook.
[6] Again, this is different for Apple Mail and for Outlook.

If the recipient relies only on the tooltip, a phisher would be successful when providing a reassuring URL.

– *Redirection:* Some email providers seem to make phish detection difficult, if not impossible. These clients do not display the actual URL in the status bar, but rather an (obfuscated) URL, a so-called *dereferer* (see Fig. 2). Web mail providers argue that they do this to protect their users (due to some checks before redirecting users to the actual web page). *Problem:* This approach makes it almost impossible for even the most security aware to detect the perfidy of the link.

Fig. 2: Status bar displays an obfuscated URL for an embedded URL

– *Tiny URLs:* Some senders of (legitimate) emails use shortened URLs to redirect the person to a different website.
Problem: From the URL it is impossible to know where a click will send people to. It is necessary to use external services to get the final destination.
– *Habituation:* While Apple Mail shows a toolbar next to the link in the email, it does so (obviously) for both legitimate and phishing emails.
Problem: While knowing that one should check the URL in the tooltip before clicking, due to habituation people are not very likely to check each URL or even a high percentage.
– *Mouse hover vs. clicking:* In order to get the relevant information in both desktop clients as well as in the Web email clients, one need to touch the link with the mouse while one must not click.
Problem: It is likely that users are not cautioned enough and instead of only moving the mouse to the link they already click before having checked.

2.3 Reasons Why People Fall for Phishing

From the above findings, the following reasons can be deduced:

1. Not being aware that the URL is the only reliable signal: making a decision based on the wrong signal.
2. Not knowing which displayed URL to trust. There are three options: embedded in email text, in the displayed tooltip, or in the status bar.
3. Not having access to the actual URL (destination) due to URLs being obscured – either because of redirection or the use of tiny URLs.
4. Not consulting the URL carefully enough before clicking due to accidental clicks and/or habituation effects.
5. Not knowing how to distinguish authentic from phish URLs.

3 TORPEDO as possible Solution

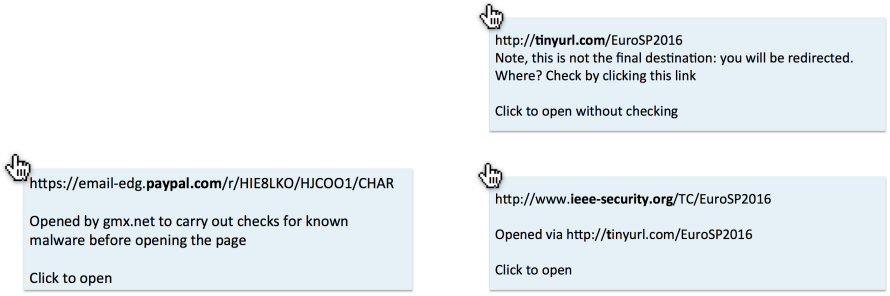
We try to address all of these reasons with *TORPEDO*. *TORPEDO* proves just-in-time and just-in-place trustworthy tooltips which contain the actual URL with the domain highlighted. It delays link activation for a short period. Furthermore, *TORPEDO* consists of an information diagram to explain phish detection, to be used together with the tooltips (when first used and on demand). We explain in the following paragraphs the different aspects and how they are supposed to address the identified reasons .

Just-in-time means that the tooltip appears when the person hovers their mouse over an embedded link. This addresses 'Reason 1' by making the reliable signal more prominent (at least compared to the status bar used by Thunderbird and the Web email clients). *Just-in-place* means that it appears right next to the link (i.e. more precisely right below the link) and only there. This addresses 'Reason 1' and 'Reason 2' by making the most important signal the most prominent one and always displaying it at the same position. *Highlighting the domain* in bold letters (similar to the highlighting in the addressbar of some Web browser) focuses attention on the most important part of the URL addressing 'Reason 5'.

Disabling the link for a short period , perhaps three seconds while providing continuous feedback in terms of a counter showing the time left to click (3, 2, 1s) increases the likelihood of people examining the link before clicking, addressing 'Reason 4'. Note, the delay is configurable to give users control. Furthermore, a white list is maintained to remember domains users have already clicked on twice before (requiring two clicks means that domains will not as easily be accidentally white listed). Whitelisted links will be activated immediately and not be subject to any delay to not annoy users.

Trustworthiness, first, requires overwriting tooltips provided by the email sender. This, together with the tooltip appearing just-in-time and just-in-place, addresses 'Reason 2'. It also addresses 'Reason 3' partially by providing the actual URL, instead of the obfuscated one the phisher wants the user to see. Fig. 3 (a) shows how we propose to handle the redirections ('redirectUrl=') aspect of 'Reason 3', i.e. providing the actual URL and informing the user that this is the second but final destination. There are two possibilities to address the 'tiny URLs' [7] aspects of 'Reason 3': (1) automatically replace these URLs by the actual one using the service from `http://longurl.org`, or (2) inform users and let them decide whether to check for the actual URL using this service. From a usability point of view the first option looks more promising; however, from a security and privacy perspective the second one is more promising (as e.g. the tool would send requests although the user does not want to visit the corresponding page). We decided to go for option (2) by default but allowing to configure option (1). Thus, users would first see the upper tooltip of Fig. 3 (b) and the other one once decided to check for the actual URL.

[7] According to Wikipedia popular shortening services are: bit.ly, goo.gl, ow.ly, t.co, TinyURL, and Tr.im. The URL is parsed accordingly. Those services are addressed.



(a) Redirection case (b) Shortened URL case

Fig. 3: Example tooltips

Information diagram, to support users in phish detection in general but in particular in checking the URL. This diagram (see Fig. 1) addresses mainly 'Reason 5'. It was iterated several times based on feedback from lay people. The diagram is shown when users initially start using our tooltips. Since users are not regularly confronted with phishing emails they may forget the rules after installation, the diagram is also available on-demand. The information diagram contains the following content while using a process approach explaining step by step what to do while considering the URL manipulation tricks introduced in [7] namely obfuscation, misleading, mangle and camouflage:

- In *Step 1*, we suggest focusing only on the URL. This addresses the fact that people do not focus on the URL ('Reason 1'). It is also explained that the remaining parts of the URL can easily be faked.
- In *Step 2* we recommend that people actually only consider the highlighted part of the URL displayed in the tooltip.
- In *Step 3*, we explain that they should check whether the brand name is highlighted (to address misleading URLs such as amazon.shop-secure.com but also obvious phishes such as IP addresses). More precisely, we explain that they should ignore the remaining parts of the URL.
- In *Step 4*, we advise them to check for extensions of the brand name such as in amzon-shopping-in-America.com. This the most difficult phishing URL to detect as some companies use such extensions in their authentic domain. We, thus, recommend that they search at Google if they are not sure.
- In *Step 5*, we recommend that they check letter by letter to identify small modifications in the domain name (to pick up mircosoft.com).

4 Evaluation

We wanted to evaluate *TORPEDO*'s effectiveness, efficiency and user-engendered confidence in terms of properly judging the authenticity of emails, as compared to the *status quo* status bar display in Thunderbird and the considered Web

email clients. To do so, we conducted a between-subjects online study launched on SoSciSurveywith participants randomly associated to one of two groups:

- **Status bar:** The group sees the URLs in the status bar.
- **TORPEDO:** The group sees the URL in a tooltip with domain highlighted in bold while having seen the information diagram.

Moreover, we formulated the following hypotheses:

- **H1 – Phish detection:** The TORPEDO group will detect more phishing emails, as compared to the status bar group.
- **H2 – Authentic eMail identification:** The TORPEDO group will identify more authentic emails, as compared to the status bar group.
- **H3 – Efficiency:** The TORPEDO group will judge emails more quickly, as compared to the status bar group[8].
- **H4 – Confidence:** The TORPEDO group will be more certain of their judgements, as compared to the status bar group.

4.1 Study Procedure

The study comprised the following three phases[9]:
Phase 1 – Welcome: General information was provided, including the goal of the study, number of phases, the estimated duration, and data protection. We explained that it was important not to seek assistance (we did not elaborate by citing Google, as this could have been counter productive). We introduced the main tasks: They should imagine that their friend Max Müller is about to work through his emails. Since Max has accounts at all the companies in question, it is important for him to know which emails are authentic and which are phish. Therefore, they were asked to help him to judge the emails based on screenshots which Max provides to them on the following pages.
Phase 2 – Judging screenshots: Participants were presented with screenshots of 16 emails (8 authentic / 8 phish) each on a different web service and in random order. The TORPEDO group got in addition the information diagram, both before starting to answer questions as well as below each screenshot. Participants were asked: Is the email authentic? Then participants were then asked: How certain are you that you properly judged the displayed emails. The TORPEDO group was also asked to comment on the information diagram.
Phase 3: Demographics: We requested demographic information.

4.2 Creation of Email Screenshots

We selected 16 web service providers based on the degree of popularity based on Alexa (see Table 1). For all of these, we determined what authentic emails look

[8] Note, on the one hand it is important that people take their time to check the URL, however in addition, if they know what to consider, they can make decisions faster.
[9] Questions were translated from German for this paper.

like (including the 'from' address). All emails addressed the 'Heartbleed-Bug'. The text recommended that the recipient change their password and provided a link to facilitate this. The text slightly differed from one email to the next but the meaning remained the same. All raising some (but not strong) pressure to change the password. Then, half of the screenshots were 'turned into' a screenshot of a phish email by modifying the URL. For the TORPEDO group a tooltip was added to display the relevant link. We decided to simulate a worst-case scenario, i.e. advanced phishing emails which can only reliably be detected by checking the URL. We wanted to investigate the difference between the status bar and tooltip, and not the impact of various different signals. All emails were personalised. We also used `HTTPS` for both phish and non-phish displays because we did not want the absence or presence of `HTTPS` to constitute a cue due to the findings in Section 2.1. Next, we considered which URL manipulation techniques to apply to get a representative set of manipulated URLs. Researchers have identified different URL manipulation classifications [7, 15, 25]. We used the categories from [7] with each type's anticipated success depending on how well users understand URLs and the thoroughness of their URL checking:

- *Obfuscate:* The phish URL is composed of an arbitrary name or IP address. Note, the brand name of the authentic website does not appear.
- *Mislead:* The phish URL embeds the authentic name somewhere (e.g. in the subdomain or the path) in order to allay suspicions.
- *Mangle:* The phish URL includes letter substitutions, different letter ordering, or misspelling e.g. `arnazon` instead of `amazon`.
- *Camouflage:* The domain name of the phish URL contains the brand name together with an extension or a different top level domain.

4.3 Ethics, Recruitment, and Incentives

Our University's ethical requirements with respect to respondent consent and data privacy were met. Participants first read an information page on which they were assured that their data would not be linked to their identity and that the responses would only be used for study purposes. Furthermore, using SoSciSurvey ensured that data was stored in Germany and thus subject to German data protection law. No debriefing was necessary. We recruited participants through a platform called Workhub, which is a German equivalent of Amazon Mechanical Turk. Every Workhub participant receives €3.

5 Results and Discussion

The demographics are summarized for both groups in Table 2. Participants in the TORPEDO group, on average, detected phishing emails 85.17% of the time and they, on average, identified legitimate emails as such 91.57% of the time. The corresponding percentages for the control group are: 43.31% for phish

Table 1: Legitimate(L) and Manipulated(M) URLs incl. type of manipulation.

Brand	**URL** (abbreviated with '...')
Postbank	L: https://banking.postbank.de/rai/login
Ebay	L:https://signin.ebay.de/ws/eBayISAPI.dll?SignIn&UsingSSL...
Xing	L: https://login.xing.com/login?dest_url=https\%3A\%2F\%2Fwww...
Google	L: https://accounts.google.com/login?hl=de
Dropbox	L: https://www.dropbox.com/s/VPrize8EppElIxxWOwETRB87 Pe733AR...
Telekom	L: https://accounts.login.idm.telekom.com/oauth2/auth?response...
Zalando	L: https://www.zalando.de/login/
MediaMarkt	L: https://www.mediamarkt.de/webapp/wcs/stores/servlet/Logo...
Facebook (Obfuscate)	L: https://www.facebook.com/login M: https://130.83.167.26/login
Flickr (Obfuscate)	L: https://login.yahoo.com M: https://www.xplan.com/signing/flickr/
Twitter (Mislead)	L: https://twitter.com/login M: https://twitter.webmessenger.com
Amazon (Mislead)	L: https://www.amazon.de/ap/signin M: https://www.amazon.de.buecherkaufen.de/ap/signing?...
DeutscheBank (Mangle)	L: https://www.deutsche-bank.de M: https://meine.cleutsche-bank.de/trxm/db/i nit.do?login...
Maxdome (Mangle)	L https://www.maxdome.de/ M: https://www.maxdorne.de/?fwe=true&force-l ogin-layer=true
Paypal (Camouflage)	L: https://www.paypal.com/signin/?country.x=DE&locale... M: https://www.paypalsecure.de/webapps/mpp/ho me
GMX (Camouflage)	L: https://www.gmx.net M: https://meinaccount.gmxfreemail.de/

detection and 63.66% for identifying legitimate emails. Note, participants, on average, detected Camouflaged URLs 72.09% of the time in the TORPEDO group and 38.37% in the status bar group, Misleading URLs 87.21% versus 30.23%, Mangled URLs 91.86% versus 37.20%) and Obfuscated URLs 89.53% versus 67.44%. Furthermore, the corresponding numbers for the answer 'I do not know' are (TORPEDO/status bar): 3.49%/6.98%, 2.91%/7.56%, 2.33%/8.14%, and 5.81%/5.23%. The descriptive data for H3 and H4 is depicted in Fig. 4.

As to the violation of homogeneity of variances for the compared groups we started our analyses with Mann–Whitney U tests for every hypothesis supplemented with an approximated effect size.

Table 2: Demographics

	# Participants	Average **Age**	Median	Youngest	Oldest	Male	IT Expert
Status bar	43	25.70	23	17	54	25	5
TORPEDO	43	27.86	26	18	60	25	2

H1 – Phish Detection: Our analysis shows a significantly improved detection rate for phish Emails for participants in the TORPEDO group, as compared to those in the statusbar group ($U = 210, p < .001$, $\eta^2 = 0.455$).

H2 – Authentic Email Identification: Our analysis shows a significantly improved identification rate of authentic Emails for participants in the TOR-

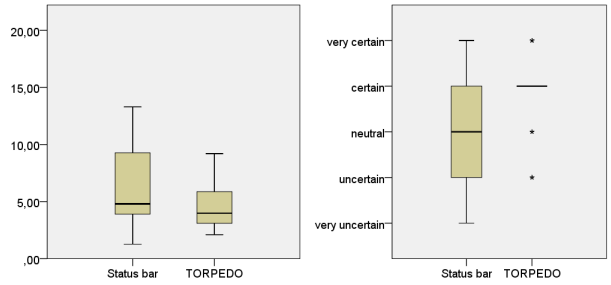


Fig. 4: Descriptive data for timing and certainty of correct decision.

PEDO group, as compared to those in the status bar group ($U = 304, p < .001$, $\eta^2 = 0.374$).

H3 – Efficiency: Our analysis shows that participants in the TORPEDO group were significantly more efficient than participants in the status bar group ($U = 676.5, p = .032$, $\eta^2 = 0.053$).

H4 – Confidence: Our analysis shows that participants in the TORPEDO group were significantly more certain about their decisions than participants in the status bar group ($U = 536.5, p < .001$, $\eta^2 = 0.155$).

Diagram Feedback. We used open coding to analyse the free text answers. We came up with the following categories: 'grateful', 'nothing to improve', 'confusing', 'too much information', 'too little information', and 'small improvements'. Most of the participants (29 from 43 in total) were happy with the diagram, not mentioning anything to improve. Three were grateful. Eight mentioned that the diagram was confusing and five added that the confusion cleared once they read it. Two participants considered the diagram to contain too much information while another two participants complained about it giving too little information (requesting more examples). Three provided small suggestions for improvement: provide a title, and reconsider the usage of terms such as phish and URL.

Discussion. The results show that, in the studied scenario, we significantly improved phish detection as well as the identification of legitimate emails with TORPEDO. The detection rates for all four phishing types increased, too. The participants in the TORPEDO group are also more confident that they made the proper decision in comparison to the status bar group. The decision making process is also significantly more efficient. Operating more quickly can, in the long run, lead to more errors being made. It is worth providing people with information such as that given in the information diagram since it is easy to apply and requires the email recipient to spend less time checking each individual email. The feedback to the information diagram showed that there is not much need to improve the diagram other than making the numbers clearer and adding a title. We acknowledge that the diagram might not provide sufficient information for

some people. In these few cases, we recommend extending our defence approach with existing proven training approaches (see Section 6).

Limitations. We acknowledge that we evaluated the approach in a best-case scenario as their primary task was security. Phishing effectiveness evaluations in field studies are not possible due to ethical and legal constraints. Lab studies also have their limitations because participants would not use a study laptop instead of their own. We also acknowledge that the URLs were displayed the entire time and not only when hovering the mouse over the link. This only partially simulates the proposed delay. Note, we only used `HTTPS` since we wanted to assess their ability to check the actual URL, not the presence or absence of `HTTPS`. Finally, we acknowledge that the sample was not representative.

6 Related Work

Researchers have proposed different ways of addressing phishing:

Automated Detection. Phishing emails can be detected either pre- or post-click. Emails can be analysed by the email provider before being forwarded to the user. This analysis includes checking the integrated URLs against several blacklists provided by companies such as Microsoft, Google and phishTank. Other checks can also be carried out. For example, to look at differences between displayed and actual domain names [12] or carry out an NLP analysis of the actual email text [36]. If the email is delivered and the person clicks, post-click detection can also occur. Web browsers or Add-ons can check the URL against various blacklists or check the web site content in combination with the actual URL. A number of different approaches for these checks have been proposed [27, 3, 28, 31]. In both pre- and post-click checks a risky situation can either lead to blocking or a warning e.g. [24, 29, 42, 40]. As a final comment, there is an inherent flaw to post-click warnings. The human tendency to consistency makes it less likely for people to even want to detect the deceptive nature of any site if they have already committed to the process [8]. They have judged the email to be legitimate. Withdrawing at this stage is unlikely. TORPEDO does not aim to replace detection approaches but to complement them in order to help people to protect themselves in case none of the checks detects the phish or it is simply during the pre-detection window [2].

Training. A number of researchers have focused on training users to spot phish [1, 5, 6, 18, 20–22, 33] but most of them address phish detection in a web browser context. Researchers have shown that training improves phish detection rates. Training has two drawbacks as compared to TORPEDO. First, people need to be aware that there is a problem and that they need to be active in dealing with it. Otherwise they will not undergo training. This problem was addressed by Kumaraguru [23] by employing the concept of teachable moments, where people are given instructions or training when they almost fall for a phish. In their scheme, instead of blocking a link they allow it, and then show them that they almost fell for a phish. Second, people may forget the information the training imparted as they are not confronted with phishing emails every day.

Again the teachable moment approach can help. However, we think providing the information graphic on demand, as and when required, is the safer approach as it might be that the next time they forget how to check the teachable moment mechanism may not be installed and they would be unprotected. Dodge *et al.* [10] report a different approach, post-click training. They send out fake phish messages and then train people who click on the links. They report a positive effect. However, this approach can only be taken within an organisation.A few participants in our evaluation had trouble understanding our diagram. For those, more exhaustive training may help them. Note, the training, as such, would be shorter than what would be required without TORPEDO.

Combining Approaches. We propose TORPEDO to complement existing approaches to address the email phishing problem. Other researchers have also proposed combining approaches to maximise phish protection. Khonji *et al.* [19] suggest a two-pronged approach, the first prong being user training, and the second being automated detection. The latter includes blacklists, machine learning and visual similarity detection. Frauenstein and Von Solms [13] propose combining human, organisational and technical measures. The first includes awareness and training, the second policies and procedures and the last one includes automated measures to detect phish.

7 Conclusion and Future Work

Phishing is a thorny issue. Trying to filter out phishing emails before they reached the end users reduces the problem. Great strides have been made in this direction but no one will claim that any automated system will catch 100% of phishing emails. So, it is up to the end users to protect themselves. The research we have presented here offers a way to support end users by deploying TORPEDO providing just-in-time, just-in-place, trustworthy tooltips; disabling links for a short period of time; detection of re-directions and tiny URLs, and providing a diagram at installation, or on demand, that encapsulates phish detection advice in a step by step fashion. This approach was evaluated and improved. We found that it highly significantly improved phish and legitimate email detection, made such detection significantly faster and led to people feeling more confident about their judgements compared to the status bar approach as used in Thunderbird and Web email clients. With 85.17% phish detection compared to 43.31% with the status bar URL display, it can only be hoped that the different email clients and email providers deploy this approach as soon as possible. Until then, people can use the *TORPEDO* Add-on we developed. As future work, we plan to conduct acceptance tests to determine whether TORPEDO will indeed be used. We also plan to extend this approach to mobile email clients.

Acknowledgement. This work was developed within the project 'KMU AWARE' which is funded by the German Federal Ministry for Economic Affairs and Energy under grant no. BMWi-VIA5-090168623-01-1/2015. The authors assume responsibility for the content.

References

1. A. Alnajim and M. Munro. An Anti-Phishing Approach that Uses Training Intervention for Phishing Websites Detection. In *6th International Conference on Information Technology: New Generations*, pages 405–410. IEEE, 2009.
2. APWG Internet Policy Committee. Global Phishing Survey: Trends and Domain Name Use in 2H2013, 2013. `http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_2H2013.pdf` Accessed 13 March 2016.
3. Z. Bar-Yossef, I. Keidar, and U. Schonfeld. Do not crawl in the DUST: Different URLs with similar text. *TWEB*, 3(1):1–31, ACM, 2009.
4. M. Blythe, H. Petrie, and J. A. Clark. F for fake: four studies on how we fall for phish. In *CHI*, pages 3469–3478. ACM, 2011.
5. G. Canova, M. Volkamer, C. Bergmann, and R. Borza. NoPhish: An Anti-Phishing Education App. In *Security and Trust Management*, pages 188–192. LNCS, 2014.
6. G. Canova, M. Volkamer, C. Bergmann, R. Borza, B. Reinheimer, S. Stockhardt, and R. Tenberg. Learn to Spot Phishing URLs with the Android NoPhish App. In *Information Security Education Across the Curriculum*, pages 87–100. Springer, 2015.
7. G. Canova, M. Volkamer, C. Bergmann, and B. Reinheimer. NoPhish App Evaluation: Lab and Retention Study. In *USEC*. Internet Society, 2015.
8. R. B. Cialdini, J. T. Cacioppo, R. Bassett, and J. A. Miller. Low-ball procedure for producing compliance: commitment then cost. *Journal of Personality and Social Psychology*, 36(5):463, APA, 1978.
9. R. Dhamija, J. D. Tygar, and M. Hearst. Why Phishing Works. In *CHI*, pages 581–590. ACM, 2006.
10. R. C. Dodge, C. Carver, and A. J. Ferguson. Phishing for user security awareness. *Computers & Security*, 26(1):73–80, Elsevier, 2007.
11. J.-P. Erkkilä. Why we fall for phishing. In *Conference on Human Factors in Computer Systems*. ACM, 2011.
12. I. Fette, N. Sadeh, and A. Tomasic. Learning to detect phishing emails. In *16th International Conference on World Wide Web*, pages 649–656. ACM, 2007.
13. E. D. Frauenstein and R. von Solms. Phishing: How an Organization can Protect Itself. In *Information Security South Africa Conference*, pages 253–268. Information Security South Africa, 2009.
14. B. Friedman, D. Hurley, D. C. Howe, E. Felten, and H. Nissenbaum. Users' conceptions of web security: a comparative study. In *CHI*, pages 746–747. ACM, 2002.
15. S. Garera, N. Provos, M. Chew, and A. D. Rubin. A framework for detection and measurement of phishing attacks. In *Recurring Malcode*, pages 1–8. ACM, 2007.
16. T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social Phishing. *Communications of the ACM*, 50(10):94–100, ACM, 2007.
17. M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y.-K. Lim. What instills trust? A qualitative study of phishing. In *FC*, pages 356–361. LNCS, 2007.
18. K. Jansson and R. von Solms. Simulating malicious emails to educate end users on-demand. In *3rd Symposium on Web Society*, pages 74–80. IEEE, 2011.
19. M. Khonji, Y. Iraqi, and A. Jones. Phishing detection: a literature survey. *Communications Surveys & Tutorials, IEEE*, 15(4):2091–2121, IEEE, 2013.
20. I. Kirlappos and M. A. Sasse. Security Education against Phishing: A Modest Proposal for a Major Rethink. *Security and Privacy*, 10(2):24–32, IEEE, 2012.
21. 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 *CHI*, pages 905–914. ACM, 2007.

22. P. Kumaraguru, Y. Rhee, S. Sheng, S. Hasan, A. Acquisti, L.-F. Cranor, and J. Hong. Getting Users to Pay Attention to Anti-phishing Education: Evaluation of Retention and Transfer. In *Anti-phishing WG*, pages 70–81. ACM, 2007.
23. P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong. Teaching Johnny Not to Fall for Phish. *Transactions on Internet Technology*, 10(2):1–7, ACM, 2010.
24. L. Li and M. Helenius. Usability evaluation of anti-phishing toolbars. *Journal in Computer Virology*, 3(2):163–184, Springer, 2007.
25. E. Lin, S. Greenberg, E. Trotter, D. Ma, and J. Aycock. Does domain highlighting help people identify phishing sites? In *CHI*, pages 2075–2084. ACM, 2011.
26. J. Ma, L. K. Saul, S. Savage, and G. M. Voelker. Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs. In *15th SIGKDD*, pages 1245–1254. ACM, 2009.
27. S. Marchal, J. François, T. Engel, et al. Proactive discovery of phishing related domain names. In *Attacks, Intrusions, and Defenses*, pages 190–209. LNCS, 2012.
28. M.-E. Maurer and D. Herzner. Using Visual Website Similarity for Phishing Detection and Reporting. In *CHI*, pages 1625–1630. ACM, 2012.
29. M.-E. Maurer, A. D. Luca, and S. Kempe. Using data type based security alert dialogs to raise online security awareness. In *SOUPS*, page 2. ACM, 2011.
30. R. Naidoo. Analysing Urgency and Trust Cues Exploited in Phishing Scam Designs. In *10th International Conference on Cyber Warfare and Security*, page 216. Academic Conferences Limited, 2015.
31. P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta. PhishNet: Predictive Blacklisting to Detect Phishing Attacks. In *INFOCOM*, pages 1–5. IEEE, 2010.
32. J. J. Rusch. The "social engineering" of Internet fraud. In *Internet Society Annual Conference*. Internet Society, 1999.
33. S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge. Anti-Phishing Phil: The Design and Evaluation of a Game That Teaches People Not to Fall for Phish. In *SOUPS*, pages 88–99. ACM, 2007.
34. F. Stajano and P. Wilson. Understanding scam victims: seven principles for systems security. *Communications of the ACM*, 54(3):70–75, ACM, 2011.
35. University of Exeter School of Psychology. The psychology of scams: Provoking and committing errors of judgement, University of Exeter, 2012.
36. R. Verma, N. Shashidhar, and N. Hossain. Detecting phishing emails the natural language way. In *ESORICS*, pages 824–841. Springer, 2012.
37. A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao. Why do people get phished?Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, 51(3):576–586, Elsevier, 2011.
38. J. Wang, R. Chen, T. Herath, and H. Rao. An empirical exploration of the design pattern of phishing attacks. *Inform. Assurance, Security & Privacy Services*, Emerald Publishers, 2009.
39. Webroot. Webroot 2015 Threat Brief. `http://www.webroot.com/shared/pdf/Webroot_2015_Threat_Brief.pdf` Accessed 13 March 2016.
40. M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *CHI*, pages 601–610. ACM, 2006.
41. Z. Xu and W. Zhang. Victimized by Phishing: A Heuristic-Systematic Perspective. *Journal of Internet Banking and Commerce*, 17(3), ARRAY Development, 2012.
42. Y. Zhang, S. Egelman, L. F. Cranor, and J. Hong. Phinding Phish: Evaluating Anti-Phishing Tools. In *NDSS*. School of Computer Science, Internet Society, 2007.