What.Hack: Learn Phishing Email Defence the Fun Way

Zikai Alex Wen
Yiming Li
Reid Wade
Jeffrey Huang
Amy Wang
Cornell University
Ithaca, NY 14850, USA
zw385@cornell.edu
yl564@cornell.edu
rmw244@cornell.edu
jh868@cornell.edu
aw545@cornell.edu

CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA.
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3048412

Abstract
As information security systems become increasingly sophisticated and reliable, humans are rapidly becoming the weakest link in the security pipeline. Technological countermeasures can only be deployed if the humans depending upon them are aware of how to use them, and hackers are beginning to take advantage of the knowledge gap that exists in this area. The recent DNC hackings during the 2016 US presidential election are evidence of this, as staff were tricked into sharing passwords which granted access to confidential information by fake Google security emails. Vulnerabilities such as these are due in part to insufficient and tiresome training when it comes to information security. A potential solution is the introduction of more engaging training methods, which teach information security in an active and entertaining way. To this end, we introduce the game *What.Hack* to teach information security and defense methods for social engineering threats.

Author Keywords
Educational Game; Usable Security

ACM Classification Keywords
K.3.1 [Computers and Education]: Computer Uses in Education; K.6.5 [Security and Protection]: Hacking Defence

Table 1 Rules in effect since Round X:

Round 1

a. Senders must be from the Big Red bank.

Round 2

(New: Trusted List)

b. Revoke a.

c. Senders must be from the trusted list.

Round 3

(New: URL Section)

d. Revoke c.

e. Senders not on the trusted list should not send URLs.

Round 4

f. Revoke e.

g. Senders not on the trusted list should not send URLs start with a purely numeric address.

Round 5

h. URL address must be identical to its pop-up address.

Round 6

(New: Attachment Section)

i. Senders not on the trusted list should not send attachments.

Introduction

Phishing is the act of deceiving people into divulging information or unintentionally installing malware on their computers by sending the victim(s) counterfeit emails [5]. These counterfeit emails work by misleading the victim into thinking they come from a legitimate source. For example, a phishing email can link to an imitation of the PayPal login screen. Victims who believe the link is legitimate will enter their login credentials to the fake site, unwittingly giving the hackers access to their PayPal account. In addition to financial gain, government-backed hackers may disrupt elections by phishing specific persons who are affiliated with powerful institutions. In the 2016 US election, John Podesta, the chairman of Hillary Clinton's campaign, clicked on the change password link in a phishing email intended to look like a Google warning [9]. His action immediately unlocked some or all of his emails to the hacker.

To repel phishing attacks, phishing defence technology has evolved rapidly. Recent automatic systems apply machine learning to classify phishing emails. However, these automated approaches are not foolproof [1]. There remains a non-negligible probability of users receiving phishing emails and these users must decide whether a piece of email lying in their inbox is safe or malicious. Therefore, user education is another major approach to protect users against phishing. To better engage learners and to change user behaviour, several anti-phishing games have already been proposed.

Although all of these games' evaluations demonstrated that they had improved users' ability to identify phishing emails and websites, the existing games leave out email context that hackers often leverage to demand immediate attention and encourage rash decision making. Moreover, these game designs are not capable of teaching players how to detect combined phishing techniques. After finishing these games, players who are reading a real phishing email might not fall for malicious URLs but they might click on the malware attachment enclosed in the same email. Incorporating these combinded phishing techniques into our game's design can lead to more engaging challenges.

To develop a comprehensive anti-phishing game, we designed *What.Hack* (pronounced what dot hack), an online simulation game that features an engaging sequence of puzzles. Each puzzle requires the player to study a list of ever-increasing rules in a rulebook that states which emails are safe or unsafe. Unsafe phishing emails in the game are generated by templates collected from real phishing emails. The player needs to carefully identify phishing emails or they will end up with a bad ending (e.g. losing their job).

Related Work

Control-Alt-Hack [3] is a board game that teaches players high-level security concepts such as phishing, social engineering, etc. By interacting with the cards, the game lets players become a little more aware of the bucket of tricks that hackers use. However, it is not meant to teach hands-on security skills like how to identify a phishing attack, which is our game's main design goal. With respect to teaching hands-on security skills, *Anti-Phishing Phil* [8] is the most popular interactive game that teaches players how to identify phishing URLs. In each round, the player uses a mouse to move a fish to "eat" the worm that shows safe URLs and "reject" the bait that shows phishing URLs. Visualising URLs as worms makes the game fun but it takes the URLs out of context. This design can not give players the real experience of detecting phishing emails.

We observed that our scenario shared some similarities with a very popular game, *Papers, Please* [6], which puts

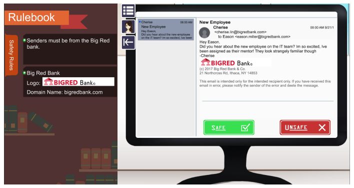

Figure 1: Game UI Overview

Figure 2: Rulebook at Round 1

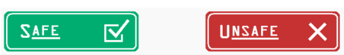

Figure 3: Player only makes yes or no decision.

Figure 4: Violation note for wrong decision.

the player in the role of a border officer. As the officer, the player must review passports and other supporting documents against rules of border control. In principle, the player only accepts passports that come with valid paperworks and rejects those with improper forms. We could make learning email safety engaging by borrowing the Papers Please game mechanics of (1) document inspection, (2) acceptance/rejection, and (3) rulesets that gradually grow more complex.

Gameplay Design

The educational purposes of *What.Hack* are to: (1) teach players how to safely handle URLs, attachments and social media and (2) provide an engaging narrative to show how social engineering attacks can lead to security breaches (e.g. insider trading in banking).

Reapplies the Mechanics of "Papers,Please!"
What.Hack provides an email client software (on the right hand side in Figure 1) that presents business emails to the player. Many of these emails are exchanged to close a deal. Some are emails from hackers who want to break the deal. Still others might be sent from insiders who want to illegally make a profit from the deal. The player needs to pick up ever-increasing rules for email safety to help their bank, Big Red Bank, seal more deals and prevent hackers and bad insiders from undermining the business. To achieve this goal, the player simply needs to make a yes or no decision on whether the email displayed on the software is a safe or an unsafe phishing email (Figure 3), which resembles a real situation almost everyone faces when they are asked for authentication or personal information via emails.

At the beginning of the game, the player starts with a single rule (see Figure 2). The game become harder over time because each new round the rulebook adds new rules on top of previous ones (see Table 1 and 2). In the first round, hackers only send 'Nigerian prince' scams [4] that can only succeed if the recipient replies. These scams can be easily separated from the bank's internal emails. Then in the higher level challenges, hackers start to use malicious URL links in their phishing emails. By introducing more rules to the player, phishing emails get more and more realistic like the ones that people constantly receive. In the last two challenges of our latest version, we also added social engineering attacks like LinkedIn friend requests from fake accounts. Social engineering attacks are now a big headache for many large firms and institutions, yet haven't been thoroughly explored in game-based security education. If the player does not read the rulebook when deciding whether an email is safe or unsafe, they will easily make the same mistakes they do in real life.

Provides Immediate Feedback & Engaging Narrative
To cut down on player's time spent in error states, this game provides immediate feedback on their decisions. Figure 4 shows the violation note that will appear if the player makes a wrong decision. If the player labels an unsafe phishing email as safe, the note will state which specifics rules the email violates. Similarly, if the player thinks a safe email is a phishing email then the note will remind the player that this email is safe. Hence, the player will have the opportunity to stop and think about why they made a mistake. This approach helps players reflect on the anti-phishing knowledge with which they are unfamiliar, which helps increase learning [2].

If the recipient got a key phishing email, *What.Hack* will play a phishing animation (screenshot at Figure 5). This animation is tailored according to the phishing email's content. It shows every word the hacker says and every key step a victim follows. Presented with a concrete context, the

Figure 5: Phishing Animation Screenshot

Figure 6: Interactive Non-Player Assistant

Table 2 Rules in effect since Round X (Cont'd) :

Round 7

j. Revoke i.

k. Attachments must end with safe file extensions: pdf, docx, xlsx, and pptx.

Round 8

(New: Social Media Section)

l. Social media should not send friend requests via emails.

Round 9

m. Revoke l.

n. Social media friend requests must match the information on the contact list.

player learns a vivid and straightforward phishing process that deceives the poor email recipient into losing money, unintentionally granting access to restricted information or devices, or both.

To gamify the process of teaching new conceptual knowledge at the beginning of each challenge, we designed an interactive non-player assistant, Cherise, in *What.Hack*. She shows up before each challenge starts and tells the player what new features make an email malicious (Figure 6). The player can also ask Cherise for help at any time by clicking on Cherise's profile image. Sometimes, Cherise also shows up at the end of the round to ask the player to find out who is the target of spear phishing or who is the villain within the company guilty of insider trading (which players may deduce from the incoming emails).

Conclusion and Future Work

In conclusion, *What.Hack* challenges players to identify real life phishing emails in a story-based game context. The player is motivated to use an invaluable reference tool, the rulebook, to determine whether or not emails are safe. Our game is now available online: www.cs.cornell.edu/~zkwen/whatdothack/. Our future work is to measure the game's effectiveness and explore new ways to motivate and teach.

We will deploy *What.Hack* to security classrooms as well as game websites. We are going to use pretests and posttests and Signal Detection Theory [7] to quantify the player's ability to distinguish between phishing emails (signal) and safe emails (noise) before and after the game. During the game, we will also collect data such as mouse events to study their behavioural changes over time. For instance, We will analyse how often users hover their mouse on URL links before they press the button. We will implement the investigation mode that allows players to match a rule with a piece of information in the email content and get feedback from the game whether they have found a discrepancy. In that case, we could study which safety rules general players are not good at following.

Acknowledgement

We gratefully acknowledge our adviser, Prof. Erik Andersen, for helpful discussions and valuable comments.

References

[1] Gupta BB Atawneh S. Meulenberg A. & Almomani E. Almomani, A. 2013. A survey of phishing email filtering techniques. In *IEEE communications surveys & tutorials*, Vol. 15.

[2] Brown A. L & Cocking R. R Bransford, J. D. 1999. *How people learn: Brain, mind, experience, and school.* National Academy Press.

[3] Lerner A. Shostack A. & Kohno T. Denning, T. 2013. Control-Alt-Hack: the design and evaluation of a card game for computer security awareness and education. In *ACM CCS*.

[4] C. Herley. 2012. Why do nigerian scammers say they are from nigeria?. In *WEIS*.

[5] J. Hong. 2012. The state of phishing attacks. In *Communications of the ACM*, Vol. 55.

[6] 3909 LLC Lucas P. 2013. Papers, Please: a dystopian document thriller. http://store.steampowered.com/app/239030/

[7] N. A. Macmillan. 2002. Signal detection theory. In *Stevens' handbook of experimental psychology*.

[8] Magnien B. Kumaraguru P. Acquisti A. Cranor L. F. Hong J. & Nunge E. Sheng, S. 2007. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *ACM SOUPS*.

[9] Wikipedia. 2017. Podesta emails. http://en.wikipedia.org/w/index.php?title=Podesta%20emails&oldid=759435543.