The Compleat Story of Phish

Hilarie Orman • *Purple Streak*

It's an everyday occurrence: an email message requires your urgent attention. You must go to a website and update your personal information or lose an essential service — your online bank account, email, or perhaps a social networking page. It's easy enough — all you have to do is click on a link and enter your account name and password. Therein lies danger, but almost everyone at some point will be fooled into taking the bait. How can a security expert explain this danger to users and tell them how to protect themselves? And is there an effective way to eliminate the exploit altogether?

The Ins and Outs of Phishing

The phishing problem is thorny because these email messages look like thousands of others that are ordinary, run-of-the-mill corporate mailings. A typical message will sport corporate logos and a pleasing layout, almost like a brochure. The sender seems familiar with your account and archly, much like any autocratic organization, demands that you comply with new policies. The next moments are crucial; if life were a movie, the background music would have a throbbing bass and minor key treble. Take the offer and choose the lady or the tiger — keep your account valid or give your valuable data to a criminal gang? Or just ignore the message and perhaps lose access to your email or banking information and enter an administrative nightmare?

These email attacks are known collectively as *phishing*, an oddly apt name for a way of confusing a user. The email is the bait, the click-through and entering of sensitive information is the hook. Once a user takes the bait, the hook painlessly extracts the information for criminal use. Phishing works because at heart, just like its moniker, it's a pun — a pun on appearance and place, and it works because the Internet creates a plastic reality in which either appearance or place can be a complete deception.

"Phish" has been a growing phenomenon, comprising more than 50 percent of all reported Internet security reports. It hits ordinary users and sometimes fools even experts. The consequences can be devastating on a personal or organizational level.

To protect yourself and your organization from phish, you must know the underlying technology, the psychology of user trust, some simple tips to help users, and some techniques that can help you fight phishers and automate phish recognition.

A phishing attack has several elements:

- an email message that mimics the appearance of a trusted business or organization;
- a large email address list of intended victims;
- a website that mimics the trusted site;
- a name for the mimicking website;
- an IP address (or addresses) for the website's server;
- malware that collects information from users visiting the site; and
- optional malware that users might be tricked into installing on their computers.

The Internet wasn't constructed with a strong notion of either place or appearance. Place was simply an IP address, sometimes loosely associated with a text string name in a globally distributed file (`hosts.txt`). The place for appearance was in the obscurity of "layer 7," the presentation layer. Today, IP addresses and names are controlled by a complex bureaucracy, names often designate well-known organizations,

1089-7801/13/$31.00 © 2013 IEEE Published by the IEEE Computer Society

and the de facto presentation layer is HTML. Together, they form the Internet's look and feel, but it's a shaky foundation with no definite guarantees.

HTML allows the presentation of an Internet placename to be, like Humpty Dumpty's use of words, anything he wants it to be. HTML completely divorces the presentation of the "click here" data from the functional data, which is what phishing relies on.

Yet, the "click action" takes you somewhere, and the where must be a real Internet "place" — an IP address associated with the name in the HTML `href` element. The Internet provides some minimal semantics to addresses and names because they must have registered owners.

Domain name ownership implies the right to associate an IP address with the names in that domain. The owner of `example.com` has the right to say what IP address is associated with `myplace.example.com` or `ibm.example.com`. But what constitutes a "good neighborhood" on the Internet? Should your bank be located here? Is there really a military contractor selling ammunition at this address? Should your favorite social networking site be on this street? These questions are difficult to answer because the Internet is good at providing content but not so good on context.

The questions of place and provenance come down to asking, for the website mentioned in an email, "Who owns the domain name?" "Who authorized the domain name?" and "What do we know about the context of the website's IP address?" When we have these answers, are there any good reasons for these entities to be associated with the business described in the email? Usually, this information will make the underlying fraud abundantly obvious, but we must follow several steps to unwind the fantasy world that phishers create.

Recognizing Phish

An effective phishing message will closely mimic the appearance of a real message from a site that a user will likely trust. For the general user, large banks, ISPs, or social networking sites are the most likely places for imitation. But many phishing attacks are targeted at specific user classes or even particular individuals.

A phishing message for a social networking site might be a simple notification that friends have recent activity on the site; an apparent link to the website, or perhaps several links, will be in the message. One of these links will go to the phishing website, while the others might actually go to the social networking site's administrative functions or main page.

Phishing for a bank or email service account might contain a more urgent message: "Verify your account details immediately or your account will be locked," or something similarly dire. Increasingly, phishing attacks concentrate on a few large brands or services, knowing that their familiarity reduces the skepticism that experienced users display toward unsolicited email.

Messages with unexplained demands for account modifications are almost always phish and can be ignored, but some — maybe one or two per year — are real and do require attention. If the message comes at a time when the user is stressed, perhaps trying to make a credit-card payment near the due date before leaving on vacation, natural skepticism might be low. How can a prudent user quickly detect the fraud?

One obvious method is to simply call the help line for the institution. "Did you send an email requiring account verification this morning?" That's not a bad approach, but if you encounter "due to heavy call volume, delays are longer than usual," you might want a method just a tad speedier.

The best and simplest is to "use the hover." With most browser-based email readers, you can easily see the text representation of the actual URL, the one that the browser will navigate to, by putting the cursor over it. The browser will display the "real" URL in a margin of the browser window. Often, this simple action will show that the destination, far from being a familiar site, is something odd, like `dskfl31245.xy`. Anyone can hover, and it's a good habit to acquire. If hovering isn't possible, then "select and paste" will do the trick: copy the link from the email message to the browser address bar and reveal the true URL behind the formatting.

Although we can easily recognize nonsensical URLs or homomorphic ones that bear a clumsy resemblance to those of real companies, most phishing messages contain URLs that are at least mildly plausible to an ordinary user. An expert can comb the DNS information to form an educated opinion about domain provenance, but an ordinary user must fall back on something straightforward.

Another good habit is to ignore the URL in the email and contact the site through a URL you've used previously. This is usually easy to find through autocompletion in the browser address bar, but the history isn't always available if the user is traveling or using an alternative computer, handheld device, or new computer. In fact, many users delete their browsing history as a security measure.

My recommendation is to keep a small list of trusted URLs with you, on your computer or written down (*not* with all your passwords, yet another security problem). The information isn't sensitive; it can be a "bookmarks" list in the browser.

However, the DNSchanger malware was a reminder that the firmament of the Internet rests on the integrity of the device being used, and that's shaky ground indeed.

The DNSchanger reconfigured computers to use criminally controlled DNS data. This meant that any URL that a user tried to access was subject to a name-to-address mapping under third-party control. Millions of people were unknowingly affected. They tried to access `example.com` but ended up at `domainofevil.net`.

Another form of phishing can be more difficult to recognize because it appears to come from a trusted source that's well-known to the victim and might contain personal information, such as the user's correct full name. The message appears to be from the victim's employer or a contractor. This especially dangerous kind of directed attacked is called *spear phishing* (especially valuable targets are the subject of *whale phishing*). The victim might see a note from his or her employer's tech support group that says the company is using a new, secure portal, and that the victim should activate this account immediately. When the victim does this, he or she exposes the username and password used to access the company's internal network. If the victim has a trusted position within the company, his or her account might be used to harm the company through monetary transfers or to seize valuable information for ransom.

Consequently, phishing renders passwords dangerous. What about public-key certificates? This is another way that the Internet lets us down when it comes to names and places. Certificates might not be dangerous, but they're often confusing and useless in practice. A certificate is a binding between a name and a public key, and the technology is used in all major browsers today. But the user must make sure that the name is meaningful and that the certificate chain is sound. One company argues that "first-generation" certificates, the kind we're used to, are useless. If the name is "Bank of American" instead of "Bank of America," few people notice. You can have a perfectly authenticated connection to "Bank of American" and no technology will ask if you might have meant "Bank of America."

Besides all this, users regularly accept expired and new certificates and keys, despite the fact that this undermines the authentication altogether. The situation isn't helped by revelations that hackers were able to create a fake certificate for Microsoft (see www.computerworld.com/s/article/9227736/Researchers_reveal_how_Flame_fakes_Windows_Update).

Still, certificates are currently the best technology for ensuring that a user is talking to the entity he or she intended. It does mean that users must know how to read certificate information, must store the certificates of those who hold their high-value assets, must make sure those certificates are being used for trusted connections, and should probably consult an expert before accepting new certificates for known sites.

Automatic Phish Recognition

Some large organizations, such as Google, have the resources to locate the DNS names and websites used for phish. They provide blacklists for firewalls and browsers. If a user clicks on a blacklisted URL, the connection won't go through, and the user might receive a warning or an error message.

So, how are these lists accumulated? If a domain is reported for being present in a phishing email, and if that domain was recently registered, then it's likely to be a phish site. However, phishers overwhelmingly favor existing domain names, and they can use them because it's fairly easy to find compromised sites. Instead of damaging the site, phishers simply use part of it for their own purposes. Site owners might have no idea that they're hosting `example.com/data451/bankofamerica`, a directory containing a phisher's copy of the real bank site. That copy will have a special code that prompts for usernames and passwords and sends them to the criminal phishers.

Because roughly half of all phishing sites are constructed using one malware toolkit, Avalanche, defenders can scan the site webpages for evidence of phishing code. Matching sites will go onto a blacklist.

Nonetheless, Google claims to deliver more than 10 million warnings per day to users who are visiting sites on their phish list, and they add about 10,000 new sites each day. Unfortunately, this technique isn't perfect, and a few legitimate sites will end up on the list.

Those sites that are set up solely for phishing (as opposed to those that are subverted) might use "agility" to try to hide from automatic detectors. These sites might register new names daily. Such names might be meaningless, but because they last for only a day or two, the blacklists are ineffective.

Another evasive maneuver phishers use is to install code that checks incoming connections and compares them against an "antiphisher" list. If the connection is from an antiphisher site, such as Google's Safe Browsing project,[1] then the phishing site will return benign results instead of the credential stealer that most users will see. This kind of evasion is called *IP cloaking*.

Agile Sites

Malicious websites must move to new DNS names often to avoid blacklisting, which can occur in as little as a couple of hours. Phishers must change their email messages to point to a new name quickly, but it can take some time to register a new name with a DNS registrar and have the DNS record widely circulated. In one study, researchers found that they could detect large groups of

names that were registered by single entities all at roughly the same time. They hypothesized that these were phishing names waiting to be used.[2]

A measure of domain name to IP addresses used in phishing shows that there are generally 10 names to one address. This high degree of reuse shows that phishers can use a website for a fairly long time before they must switch. Compromised websites are especially valuable, and if administrators are slow to respond to the problem, the site could be used for quite some time.

Compromised Sites

Sometimes, phishers are lucky enough to have illicit access to a website server that's part of a legitimate business. An SQL injection attack can be an entry point for an attack that compromises the server. Instead of pillaging the hapless victim, the attackers might opt to use the server as a base of operations for phishing. Thus, a user might find that an email recommends that he or she update bank account information at an auto parts store's website. Although this seems to be an obvious clue to malfeasance, the situation can be complicated. Some ISPs, for example, contract out their email service to a third-party provider, using a complicated series of URL redirects in and out of third-party sites. Evil, or not? Sometimes the only answer is to endure that phone queue.

Specialty Phishing

Phishers normally cast a wide net, looking for as many gullible users as they can entrap with short emails. However, some phishers find more lucrative hauls through high-value targets. A business that fails to protect its customers' contact information might open the door to having those customers spear phished.

Sometimes, phishers look for "the keys to the kingdom" by going after trusted insiders in a company. A small amount of knowledge about a company's staff and internal email configuration is valuable for a whale phisher, who will craft a message to target a CFO or IT manager. Particularly clever phishers will even look up company officials' names so as to mimic the division director or even the CEO as the sender.

Unfortunately, most natural caution disappears when the sender is well-known. One IT director told a story about finding an insidious phishing email in his inbox. He forwarded it to executive staff with the comment: "This is an example of the kind of email we would never act on." One of the VPs clicked through to the website in the email and downloaded what turned out to be malware. His reasoning was that he expected that anything sent by an IT director would be safe.

This brings us to the peculiar psychology of the chronic clicker, the incautious user whom phishers feast on. Some researchers asked "Who falls for phish?" and came up with some interesting results that can help security advisors.[3] One way to interpret the study is to say that people who value trust in their noncomputer lives are more likely to click through on phish than others, even after being trained to recognize the danger signs.

Finally, phishing sites can be more dangerous than a simple grab of credentials; they can be vectors for malware transmission. If a site claims users must upgrade their software to use what seems to be a trusted site, the result might be the installation of a keystroke logger while the user sees ordinary, unsuspicious website interactions. This combination of phishing and malware is especially damaging, but often difficult to avoid because, all too frequently, sites do require that users upgrade software.

The Future, Zero Day Phish

So, where is phishing technology going, good or bad? The number of phishing attempts has been increasing for years, the techniques for evading detection keep improving, and the problem remains a serious one for the billions of Internet users who see an endless series of attacks designed to separate them from their assets.

The information that we have about trends comes largely from an organization devoted exclusively to fighting phish, the Anti-Phishing Working Group (APWG; http://apwg.com) and its regular, detailed reports on the state of phishing.

Their most recent report notes that the number of attacks continues to increase, while the number of targeted institutions declines.[4] In an encouraging sign for the white hats, phishing sites get shut down faster than ever, showing that despite the cat-and-mouse games, it's possible to find malicious sites with some accuracy.

Last year, the number of domains phishers used took a steep drop, perhaps because registrars became more diligent about looking into would-be registrants' credentials. Although some of this improvement has reversed, the number remains lower than at its height of nearly 80,000 domains used for phishing in early 2011.

Worldwide, users in China are the most likely to be phish recipients. The perpetrators, however, don't use domain names registered to Chinese registrars for .cn, instead opting for one of the other 202 top-level domains used in attacks the APWG has detected.

These numbers show that phishing remains an ever-present threat, one almost every Internet user will see from time to time, and worth some effort on the part of the bad guys. However, their gains are probably declining, and new efforts are targeting specific high-value brands. Moreover, like many other kinds of crime, the most vulnerable targets

continue to be the frontiers where the newest users live.

The concentration on high-value targets is likely to continue, if the current trends are indeed based on economic principles, and spear phishing will become more prevalent and more likely to present a trustworthy facade. Every compromise of user email addresses and personal information sets the stage for a later opening of the phish floodgates.

In this evolution, keep in mind that zero-day phishing attacks can still emerge. A network of a thousand hacked servers yet unexploited, flaws in DNS servers, silent malware launched from "required" downloads — this could all be lying in wait as hackers clear the path for a clever new attack. We'll continue to search for ways to merge common sense, real-world trust with the artifacts of the virtual, untrustworthy world of Internet reality.

References

1. N. Provos, "Safe Browsing – Protecting Web Users for 5 Years and Counting," Google Online Security blog, 19 June 2012; http://googleonlinesecurity.blogspot.com/2012/06/safe-browsing-protecting-web-users-for.html.
2. M. Felegyhazi, C. Kreibich, and V. Paxson, "On the Potential of Proactive Domain Blacklisting," *Proc. 3rd Usenix Conf. Large-Scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and More*, Usenix Assoc., 2010, pp. 6–6.
3. S. Sheng et al., "Who Falls for Phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems*, ACM, 2010, pp. 373–382.
4. *Global Phishing Survey: Domain Name Use and Trends in 1H2012*, tech. report, Anti-Phishing Working Group, 23 Oct. 2012.

Hilarie Orman is a security consultant and president of Purple Streak. Her research interests include applied cryptography, secure operating systems, malware identification, security through semantic computing, and personal data mining. Orman has a BS in mathematics from the Massachusetts Institute of Technology. She's a former chair of the IEEE Computer Society's Technical Committee on Security and Privacy. Contact her at hilarie@purplestreak.com.