A DESIGN FOR AN ANTI-SPEAR-PHISHING SYSTEM

John Aycock
Department of Computer Science, University of Calgary, Calgary, Alberta, Canada

Email aycock@cpsc.ucalgary.ca

ABSTRACT

Phishing is a widespread and effective computer-mediated social attack. Phishers have proven highly adaptable in terms of exploiting new communications channels – witness 'vishing' and 'SMiShing' – and are becoming increasingly sophisticated. At the same time, research has shown that current anti-phishing measures are less than adequate.

One concern in terms of malicious software is targeted attacks; the phishing equivalent is 'spear phishing', where a phishing attack is directed at a specific organization or even individuals. Spear phishing may present users with some difficult decisions regarding the authenticity of messages. We propose a design for an anti-spear-phishing system to help users in this regard, which will take advantage of the characteristics of spear phishing to detect such targeted attacks.

The system we propose would work at two levels: a global level and an institutional level. We conjecture that taking these indicators together will yield an effective defence against spear phishing.

INTRODUCTION

Phishing is a social attack where targets are tricked into revealing confidential information, like a password or credit card number. While this has occurred as a social engineering attack for a long time (e.g. [1, pp.86–87]), the social engineering variant is very labour-intensive in terms of the attacker's time; consequently, the attacker must be highly selective when choosing targets. Phishing on the Internet, on the other hand, involves the computer and can be done on a massive scale, for example by bombarding millions of randomly chosen people with phishing emails, in the hopes that some recipients will a) bank with *Citibank*, and b) be tricked into revealing their banking information.

Estimates vary as to how many people fall victim to phishing, and in general it is probably impossible to know. However, the numbers that have been put forward are uniformly low – studying specific populations, one case claims 0.4% [2], another [3] cites a *Gartner* report claiming 3%. This suggests that phishing *must* be done on a massive scale in order for the attacker to have even a modest success rate. Or does it?

Lack of phishing success could be attributed to a number of factors. First, anti-phishing defences may simply be good enough that they thwart the majority of phishing attacks. Unfortunately, this does not seem likely. An increasingly large body of research suggests that users tend to ignore warnings from anti-phishing tools [4–6]. Second, the phishing message may not reach the correct target audience. This may be because the message is sent to the wrong audience to begin with, or because the message is caught by anti-spam defences before being deposited in mailboxes. Both situations occur; in the latter case, the same message sent all over in large quantities is a fairly obvious indicator to anti-spam software. Third, the phishing message may be received, but the recipient is not convinced by the message – user education may play a part – or the message (because it is sent to many people) is not sufficiently customized to be convincing.

To that end, Jakobsson argued that phishing success rates could increase if phishing messages took more context into account, making the messages more convincing [3]. More context can be said to be used by the phishing attack that is the topic of this paper: spear phishing.

Spear phishing is a phishing attack that targets a single organization. It allows a phisher opportunities to create a more customized phishing message, because more context is known. Also, if the message is only sent to the targeted organization, it stands less of a chance of being detected by some anti-spam defences, because the message is novel and not seen outside the organization. Spear phishing can be seen as the phishing equivalent of targeted malware attacks [7], which are a growing concern within the anti-virus community.

We think that spear-phishing attacks may exhibit characteristics that make them detectable, independent of the exact content of the phishing messages. In the remainder of this paper, we examine likely spear-phishing scenarios as a lead-in to our proposed anti-spear-phishing system, before finishing with our conclusions.

SPEAR-PHISHING SCENARIOS

We begin by bounding the type of spear-phishing attacks we consider, in three ways:

1. While phishers have made fledgling attempts to use other communication channels like SMS [8], the typical delivery method for phishing messages at present is email. We thus only consider email-based spear phishing.
2. We are only concerned with the behaviour of incoming email traffic to try and catch spear-phishing messages. Other anti-phishing systems watch outbound traffic, looking for sensitive information being sent [9, 10]. Such outbound-traffic systems are complementary to the inbound-traffic systems we study.
3. We restrict the attacks that we examine to ones which target an organization's employees, but not its clients. For example, a spear-phishing attack against a bank's clients would not be within our scope, but a spear-phishing attack targeted at the bank's employees would be. This restriction still means that enterprise networks are considered.

Given these bounds, the distinguishing factor between different spear-phishing scenarios is whether the email originates from outside an organization's network, or from inside it. We call these external and internal attacks, respectively.

External attacks

An external attack is one where the spear phisher emails the phishing message into a targeted organization from outside. There are no distinguishing characteristics beyond that: for example, the emails may come from a single machine or many machines (i.e. sent using a botnet), and may appear legitimate. Hemmingsen *et al.* [11] survey a wide variety of techniques

that botnets can use to look legitimate and skirt anti-spam defences.

One of these techniques is where each zombie computer in a botnet modifies the message to customize it. We will, however, assume that the content of the spear-phishing messages is relatively similar. The message would have to be hand-tailored by the phisher to the targeted organization, and would have to be convincing-looking. Apart from adding in some context-appropriate hash busters, there is probably little the phisher could do automatically (or have zombies do automatically) without making the email look obviously spammy to anti-spam defences. For example, a faked invoice number could be added that changes for each message, or an image of the organization's logo which is altered slightly for each message, but a block of white-on-white word salad would look suspicious. We will therefore assume that the signature of the message body will remain fairly constant for a convincing spear-phishing email – as we argue later, this is important for detection of external attacks.

Internal attacks

Here, the spear-phishing message is sent from within the targeted organization. There are a number of ways this can happen:

- Insider attack. There is always the possibility that someone who legitimately works inside the targeted organization is the spear phisher, in which case they can easily send the phishing message from within. Note that, in the other cases below, the spear phisher is assumed to be a person based outside the targeted organization.
- Phishing message sent from phisher's computer. The spear phisher could accomplish this in several ways. Although entailing personal risk, the phisher could use social engineering to physically enter the organization's premises, plug in their own computer, and send the phishing message. Much less dramatic is the prospect of the phisher connecting to an insecure wireless network, which permits the phisher to send mail from behind the organization's firewall.
- Phishing message sent from organization's computer. This obviously means that the phisher must gain access to at least one of the organization's computers. Again, there are several ways this could be achieved. A manual approach might involve the phisher breaking into one of the organization's computers, for instance.

  Malware could be employed to compromise an internal machine too. A worm or virus could be written by the phisher that would spread, but only activate its payload when it runs on the target organization's computers. Alternatively, a Trojan horse could be installed on USB keys and left in the organization's parking lot [12]. Any of this malware could check to ensure that it was running inside the organization, such as by checking the domain name. It could also obfuscate its payload using strong encryption whose decryption key is the organization's domain name; in this case, the payload would decrypt correctly only when running inside [13].

Regardless of how the spear-phishing email is sent from inside the organization, or how legitimate email is sent within the organization (e.g. webmail, a desktop client), there is one thing in common. Mail delivery inside the organizations we study needs to have a centralized machine through which email is sent. Because of this, it does not matter how the email originates internally; it must all be sent through the centralized mail machine.

This means that our anti-spear-phishing system, as we asserted before, need only consider two scenarios: external attacks and internal attacks. Furthermore, the details of the scenarios, like how exactly the internal attack occurred, may be abstracted away without loss of generality.

A PROPOSAL

The two types of attack require two types of detection because of their different properties. In this section, we present our proposal for an anti-spear-phishing system, which we break into two parts, one for external attacks and one for internal attacks. We follow this with a discussion about the system.

Defence against external attacks

Detection of an external spear-phishing attack requires that a system detect three properties:

1. The mail originated from outside the organization.
2. The same message is being sent in bulk to the organization.
3. The message is not being sent in bulk to other organizations.

The first property is trivial to detect, of course – the mail will have arrived at the outward-facing SMTP server, or (if there is only one SMTP server) the TCP connection to the SMTP server will have come from outside the organization.

The second and third properties are different applications of the same defence: the Distributed Checksum Clearinghouse (DCC) [14]. DCC is actually intended for anti-spam. Mail systems compute fuzzy checksums of incoming messages, checksums that are robust in the face of hash busting, and use the checksums to query a DCC server. In response to a checksum, the DCC server returns a count of how many times that checksum has been seen; the DCC server also increments its own count for that checksum, and periodically shares its counts with other DCC servers. The DCC server thus approaches a global viewpoint of what email checksums are frequently seen, and can share that data with mail systems to help them detect spam.

In our case, a targeted organization would need to run a DCC server local to the organization; this would be able to detect the second property above, specifically the same message being sent in bulk. The anti-spear-phishing system would also query a global DCC server to find out if the message was *not* seen elsewhere, which was the third property.

Defence against internal attacks

Internal spear-phishing attacks are trickier to detect because email, even sent in bulk, can be perfectly legitimate inside an organization. We suggest that the defence against internal attacks rely upon something which is often the target of scorn in security: typical user behaviour.

Research has been conducted suggesting that individual users can be profiled in terms of their email-sending characteristics [15]. For example, users tend to email people within their social network, and make even finer distinctions; Stolfo *et al.* point out that people would not typically send the same

message to 'a spouse, a boss, "drinking buddies," and church elders' [15, p.196].

We suggest that broadcasting a spear-phishing message to all members of an organization would not be typical email behaviour for most users. A phisher who attempted to do this would violate the user profile, which could be tracked by the organization's mail server. Suspect emails would be quarantined pending examination by a human, which would mitigate the effect of false positives. So, even though Stolfo *et al.*'s system focused on viral propagation – they only mentioned fraud in passing – we think it would be an effective part of an anti-spear-phishing system.

We would go a step further, however. The 'e' in 'email' doesn't stand for 'egalitarian'. Within an organization, everyone does not need to email everyone else, and in fact there are usually social conventions discouraging it. For example, an employee may feel quite comfortable emailing their boss, but be much more hesitant to email the president of the company. There is no reason social norms cannot be reflected in an organization's email system and be enforced by the mail server. For instance, the mail server could be equipped with a representation of the organizational chart, along with some rules:

1. A user can email their boss.
2. A user can email their peers (i.e. the 'children' of their boss in the organizational chart).
3. A user can email people below them in the organizational chart.
4. A user can reply to mail sent to them by others in the organization.

Obviously this would need adjustment according to the organizational norms, but particularly in larger organizations this could help compartmentalize damage due to spear phishing and email-borne malware.

DISCUSSION

There are a number of advantages to this design. First, it is not based on the content of the spear-phishing messages. Second, as a result of it being content-independent, it is a proactive defence that does not rely on signature updates.

As with any defence, there are also disadvantages. For example, employees of an organization need to be made aware that their email behaviour is being profiled, and that suspect messages they send may have their content examined by a human. There are also situations in which the system would not be effective or might be circumvented:

- Messages which are highly targeted by the phisher to a carefully selected group of recipients within an organization would not be detected. The recent 'Better Business Bureau' phish [16], sent only to high-ranking company managers, is an example that had to involve manual targeting by the phisher. Our proposed system would not catch this, and more generally, any external spear-phishing attack with a small enough number of messages might slip by. This is because the heuristic threshold distinguishing bulk email from non-bulk would not be crossed.
- Stolfo *et al.* note that their system had difficulty detecting slow email virus propagation [15]. This same observation would also apply to internal spear-phishing emails sent slowly.
- This system would likely not be useful for small organizations, where everyone in the organization normally mails everyone else. It is debatable, though, if spear-phishing attacks would be effective in that setting anyway.
- Malware could conceivably mimic a user's email-sending profile. There is nothing preventing malware from mining a user's social network and other email-sending characteristics from saved email [17, 11].
- Insider attacks are problematic in terms of email profiling, because the insider may not stray from their email profile in order to send the spear-phishing email. Also, insiders may know of, and have access to, non-direct communication channels. For instance, a message posted to an internal-only mailing list may be effective in spreading a spear-phishing message. This is a topic that would benefit from future work.

We also considered extending the defensive system onto desktop computers within the organization, which would be able to send indications that sent email corresponded to a period of user activity on the machine. This could easily be rendered ineffective if malware were to infect the machine, though. Even in an ideal situation where machines were never infected, there would be logistical problems: supporting a variety of platforms, handheld and laptop devices, machines not under an organization's administrative control (e.g. visitors). For these reasons, we believe that spear-phishing defences are best kept centralized on an organization's email server.

CONCLUSION

Computer security is a matter of establishing multiple layers of defence. The approach we suggest here, while not perfect, would provide one extra layer of defence against phishing, and to the best of our knowledge no defences specifically for spear phishing have ever been proposed. Such a system should still be used in conjunction with existing defences, such as anti-spam and anti-virus software. It will be interesting to see how well an implementation of the proposed system operates, and its performance in terms of classifying both legitimate traffic and spear-phishing attacks.

ACKNOWLEDGEMENTS

The author's research is supported in part by grants from the Natural Sciences and Engineering Research Council of Canada. José Fernandez participated in initial discussions about this system, and Ryan Vogt made a number of helpful comments on a draft on the paper.

REFERENCES

[1] Landreth, B. Out of the Inner Circle, Microsoft Press, 1985.

[2] Florêncio, D.; Herley, C. A large-scale study of web password habits. 16th International World Wide Web Conference (WWW 2007), pp. 657–665.

[3] Jakobsson, M. Modeling and preventing phishing attacks. Financial Cryptography '05, phishing panel.

[4] Dhamija, R.; Tygar, J. D.; Hearst, M. Why phishing works. CHI 2006, pp. 581–590.

[5] Schechter, S. E.; Dhamija, R.; Ozment, A.; Fischer, I. The Emperor's new security indicators. IEEE Symposium on Security and Privacy, 2007.

[6] Wu, M.; Miller, R. C.; Garfinkel, S. L. Do security toolbars actually prevent phishing attacks? CHI 2006, pp. 601–610.

[7] Shipp, A. Targeted Trojan attacks and industrial espionage. 16th Virus Bulletin Conference, pp. 56–60, 2006.

[8] Rayhawk, D. SMiShing – an emerging threat vector. McAfee Avert Labs Blog, August 2006.

[9] Chou, N.; Ledesma, R.; Teraguchi, Y.; Mitchell, J. C. Client-side defense against web-based identity theft. 11th Annual Network and Distributed System Security Symposium (NDSS '04), 2004.

[10] Kirda, E.; Kruegel, C. Protecting users against phishing attacks. Computer Journal, 49(5):554–561, 2006.

[11] Hemmingsen, R. H., Aycock, J.; Jacobson, M., Jr. Spam, phishing, and the looming challenge of big botnets. EU Spam Symposium, 2007.

[12] Stasiukonis, S. Social engineering, the USB way. Dark Reading, 2006.

[13] Riordan, J.; Schneier, B. Environmental key generation towards clueless agents. Mobile Agents and Security (LNCS 1419), pp. 15–24, 1998.

[14] Distributed Checksum Clearinghouse. 2007. http://phyolite.com/anti-spam/dcc.

[15] Stolfo, S. J.; Hershkop, S.; Hu, C.-W.; Li, W.-J.; Nimeskern, O.; Wang, K. Behavior-based modeling and its application to email analysis. ACM Transactions on Internet Technology, 6(2):187–221, 2006.

[16] Stewart, J. BBB phishing Trojan. May 2007. http://www.secureworks.com/research/threats/bbbphish.

[17] Aycock, J.; Friess, N. Spam zombies from outer space. 15th Annual EICAR Conference, pp. 164–179, 2006.