Combining Authentication, Reputation and Classification to make Phishing Unprofitable

Amir Herzberg

Abstract We present and analyze a design of an filtering system to block email phishing messages, combining reputation, authentication and classification mechanisms. We present simple economical model and analysis, showing sufficient conditions on the precision of the content-classifier, to make phishing unprofitable.

1 Introduction

Phishing is a common social-engineering attack on computer users, causing significant losses to individuals and society. In a phishing attack, Phil, the 'phisherman', sends email (or other message) to Vic, the victim (user). The email lures Vic into exposing herself to further attacks. Phishing is based on deception; Vic is led to believe that the email is from a trustworthy source, such as her bank, e.g. VIC-Bank.com. In a typical attack, Vic follows a hyperlink in the message, which causes her browser to open a *spoofed website*, e.g. a clone of the login page of VIC-Bank.com. If Vic does not detect that the site is spoofed, she may enter her credential, thereby allowing Phil control over Vic's account.

Phishing emails are one of the most harmful categories of spam. There are many products, services and proposals to allow mail servers and readers to block phishing (and spam) emails. Many of these mechanisms fall into the following three classes:

Reputation mechanisms, e.g. blacklists: these systems map the identity of the sender, to some measure of his reputation as a mail sender. The simplest reputation systems, which are also most common, are *blacklists* (and *whitelists*), which simply list known phisherman/spammers (or, respectively, trustworthy senders known not be phishermen/spammers). More elaborate reputation systems may

Amir Herzberg
Bar Ilan University, Computer Science Department, Ramat Gan, 52900, Israel, e-mail: herzbea@cs.biu.ac.il

return a measure of the reputation of the sender. Notice that many blacklists are not sufficiently reliable, and may suffer from many false positives. It is often advisable for organizations to use two blacklists, a 'short' blacklist (often maintained locally), where false positives are very rare,and a 'long' blacklist, which contains many more suspected senders (and more false positives). Most blacklists use the IP address of the sending mail server as the identifier, allowing for highly efficient lookups (using DNS).

Authentication mechanisms: these mechanisms authenticate the identity of the sender, or of the sending domain. There are several authentication mechanisms for email, mostly based on the security of routing and DNS, and/or on cryptographic authentication such as using digital signatures. We discuss the predominant mechanisms: SPF [14] and SenderID (SIDF) [11], based on security of routing and DNS, and DKIM [2, 10], based on security of digital signatures.

Content classifiers: These mechanisms classify emails based on their contents, typically to suspect email (spam or phishing) vs. 'good' email (sometimes referred to as 'ham').

Many email systems, employ some combination of reputation, authentication and classification mechanisms. A high-level design of an email filtering system is shown in Figure 1. In this design, we use four steps: step two is sender authentication, step four is classification, and steps one and three (either 3a or 3b) use reputation (a 'short' blacklist in step one, a domain-name sender reputation lookup in step 3a, or a 'long' blacklist in step 3b). We next give a brief description of these steps; for more details on this design, see Section 2.

In the first step, we confirm that the sending mail server is not listed in a blacklist of servers suspected of frequently sending spam/phishing emails. This step is very efficient, esp. since blacklists are usually kept as DNS records, and hence retrieved and cached efficiently. Unfortunately, most phishing messages are sent from legitimate domains, see e.g. [5]. Hence, often the sender of the phishing email will not have bad reputation (e.g. not be in the blacklist), and will only be detected by the following steps.

In the second step, we authenticate the sender identity (name), if an appropriate authentication mechanism is available. Such authentication mechanisms include validating a digital signature (e.g. using DKIM) and/or checking that the sending server is listed in a 'email sending policy' DNS record controlled by the sender (e.g. using SPF or SIDF). If no authentication data is available, we cannot identify the sender (by name), and proceed using only the IP address of the sending server (in step 3b). If authentication data exists but the validation fails, then we reject the email, and optionally add the sending server's IP address to the 'short blacklist' (so future emails from this server are blocked immediately and efficiently by the IP-based 'short' blacklist, in step 1).

If the authentication validates correctly the identity of the sender and/or of the sending mail server, then we use this identity (or identities) in step 3a, to check the reputation of the sender. In this case, we can block the email if the sender is a known spammer/phishermen, or display it if the sender is known to be trustworthy.

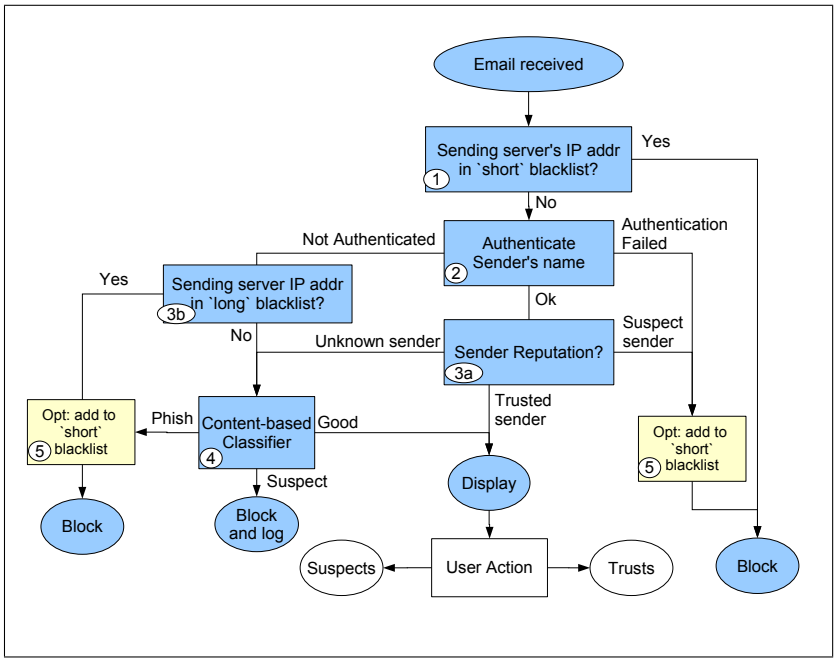

Fig. 1 High level design of an email filtering system, using four steps. Steps 1 and 3 use reputation (by blacklist of IP addresses, and by a reputation database for senders). Step 2 authenticates the sender, and step 4 classify the email (based on its contents).

If the email is not authenticated at all, then we may check if the sender (or sending mail server) is listed in the 'long' blacklist (step 3b). The 'long' blacklist is applied only for unauthenticated senders, since it is less reliable than other mechanisms (contains many false positives). In addition, the 'long' blacklist is often located as a remote server, therefore querying it involves delay and overhead. If the sender appears in the 'long' blacklist then the email is blocked; optionally, the sender is also added to the 'short' blacklist, for improved efficiency of additional messages from that sender.

If the sender is not authenticated, yet not blacklisted, or authenticated by without sufficient (positive or negative) reputation, then we must invoke the last and most computationally-consuming step (4): content-based classification. The content classification system determines whether the email is good (to display) or bad (to block, and possibly to log). Notice that content classification systems are both computationally expensive and never fully reliable, therefore it makes sense to apply them only if the more efficient authentication and reputation mechanisms failed to produce a conclusive determination. The classification system may also use results from the previous mechanisms; in particular, often it may consider the (inconclusive) reputation as part of its input.

Finally, once the mail reader displays the email to the user, then the user makes the final determination: to trust the email or to suspect it (in which case, the user may simply discard it, or may report this, e.g. to a blacklist. The user's decision may be (partially) based on sender identifiers presented by the mail reader, e.g. the sender's address, usually from the FROM email message header [12]; these email identifiers may be authenticated, e.g. with Sender-ID and/or DKIM. However, practical experience, as well as usability experiments [4], show that users often trust phishing email based on its contents, even when it has the wrong 'from' address, especially when using similar characters, e.g. accts@VIC-Bank.com vs. accts@VlC-Bank.com (it may indeed be difficult to see, that the second address uses (lower case) *l* instead of (upper case) *I*).

Combinations of reputation, authentication and classification mechanisms, similar to the design outlined above, are often deployed by email systems, to block phishing and other spam emails; see also [9]. In this paper, we describe and analyze the details of this design. We also explain the relevant adversary models, with brief background on relevant Internet protocols (IP, DNS and SMTP).

Furthermore, we present a simple modeling of the economics of phishing. Our analysis shows sufficient conditions under which a phishing-defense system following the design in Figure 1, can ensure that phishing is not profitable. These conditions are derived under reasonable simplifying assumptions.

Our analysis is especially meaningful for the design of the content classification mechanisms. First, the conditions we identify for making phishing unprofitable, imply required level of precision for content classification. Second, the analysis shows that it may suffice to ensure that phishing messages are either classified (as 'suspect'), or simply suspected (or ignored) by the user.

This motivates us to recommend that sensitive senders, e.g. banks, use email authentication mechanisms (e.g. DKIM and SPF), and in addition adopt a standard form for their emails, allowing easy classification of emails with similar form and ensuring that users will suspect (and ignore) emails which claim to be from the bank but have different form. When senders use this combination of authentication and easy-to-classify form, the content classifier can identify emails which use the bank form; any such email which is not properly authenticated, is probably phishing email. Such high-precision detection of phishing emails allows the use of automated means to detect and punish the phishermen, making phishing less lucrative or unprofitable. Details within.

Email authentication is a central element in our phishing-detection design, as shown in Figure 1. Currently, there are several proposals for email authentication. We describe and evaluate the three predominant proposals: the Sender Policy Framework (SPF) [14], the Domain-Keys Identified Mail (DKIM) design [2, 10] and the Sender-ID Framework (SDIF) [11]. We make several recommendation as to best method to use (and combine) these mechanisms, and explain their security properties, clearing up some possible misconceptions and unjustified expectations.

To summarize, we believe this paper has the following contributions. First, we present a detailed design combining authentication, reputation and classification mechanisms, to filter phishing and spam messages; our design includes some new

insights, such as improving the classification by identifying emails which may appear to the user to come from specific senders. Second, we present economic analysis, showing sufficient conditions for phishing to be unprofitable. Third, we present and compare the three predominant email authentication mechanisms (SPF, Sender-ID and DKIM), describing their correct usage and limitations, and map them to the corresponding adversary models.

2 Design of an Integrated Email Filtering System

In this section, we present and discuss a high-level design for an email filtering system, incorporating reputation, authentication and classification mechanisms. As illustrated in Figure 1, the design incorporates four steps; these steps are denoted by the rectangles with gray background, numbered 1-4. Notice that not all recipients will use exactly these steps or exactly this order of steps.

In the first step, the filter looks up a blacklist containing IP addresses suspected to be in use, or to be available for use, by phishermen and spammers. Such lookup is very efficient, in particular since it is usually done by an DNS query, and the results are cached. The IP address to be used here should be of the last untrusted MTA which sent (relayed) the message; if this IP address appears in the blacklist, the message is blocked. This step can be skipped if it was already performed by some trusted MTA along the route (after which the mail passed only via trusted agents), e.g. by the incoming border MTA of the recipient's organization or ISP. Some recipients may also block email when the IP address used by sending MTAs has not been used in the (reasonably recent but not immediate) past to send email. This can block many 'bad' servers (albeit also few legitimate but new servers), since these newly-used addresses may not yet appear in blacklists, yet much of the spam and phishing email arrive from such 'new' IP addresses [5].

In the second step, the filter tries to authenticate the sender, using IP-based authentication (e.g. SPF) and/or cryptographic authentication (e.g. DKIM). If the authentication fails, i.e. the email is signed but the signature is invalid (for DKIM) or the SPF record last untrusted MTA which sent (relayed) the message, then the email is blocked. If authentication is successful, namely the email sender or sending domain is authenticated, then this identity is passed to the next step, to check the reputation of the sender (or sending domain). If there is no authentication data, then we skip the next step (cannot check reputation for unidentified senders) and move to the following step (content classification).

The third step is reached only if the sender of the email, or the sending domain, was successfully authenticated in the previous step. In this case, we can now consult reputation database, using the identity of the sender (or sending domain) as keys. If the reputation data for this sender is conclusive, we block the email (for a suspected sender) or display it to the user (for a trusted sender). If there is no conclusive reputation data for this sender, we pass whatever reputation data we obtained to the next and final step of content classification.

The fourth (and last) step is content-based classification, based on heuristic rules, machine learning and/or other approaches. Unfortunately, all content classification mechanisms are both computationally intensive, as well as not fully reliable. Therefore, we execute this step only when all previous steps failed to provide an conclusive decision; furthermore, at this step, we may use the outputs of the previous steps, such sender identity (if identified) and reputation (if some, non-conclusive, reputation data was found). We make additional recommendations about this step below.

Identification of phishing email is challenging, since phishing messages are designed to mimic legitimate messages from a specific, trusted sender (e.g. VIC-Bank.com), in order to trick Vic into believing the message came from VIC-Bank.com. This may make classification of messages to phishing vs. non-phishing more challenging, compared to classification of 'regular' spam messages.

In spite of this challenge, classifiers have been shown to achieve good precision in identifying phishing messages, over collections containing typical phishing messages [6, 3, 1], using features which are often unnoticed by (human) victims, e.g. hyperlinks to suspect websites in the email. In existing email filtering systems, there is usually a 'classification engine' which applies heuristic or machine learning rules, to classify directly to undesirable (spam/phishing) vs. legitimate ('ham'). Alternatively, the classification engine may output a 'grade', which is combined with 'grades' from the reputation steps, to determine if to accept or block the email.

However, it is hard to predict whether automated classifiers would be able to maintain such good precision in the long run, after being widely adopted, since at that point phishermen are likely to try to adapt their messages to try to avoid detection (via phishing-related features). This motivates our different, possibly complementing, approach, namely to use classifiers to identify *PhishOrReal* emails, i.e. messages which *appear* to come from VIC-Bank.com (regardless of whether they really come from VIC-Bank.com, or are phishing). Since phishermen try to mislead Vic into believing their phishing email is really from VIC-Bank.com, the identification of *PhishOrReal* emails should be easier, than classifying emails as phishing. Furthermore, it should not be too difficult to generate and collect a large corpus of *PhishOrReal* messages, to use to train, test and/or fine-tune the classifying engine.

Therefore, we suggest to use a 'classification engine' (using heuristics, machine learning, etc.), to classify incoming emails to *three* groups: messages directly identified as spam or phishing; *PhishOrReal* messages; and other messages. Since our design invokes the classification engine only at step 4, and, assuming VIC-Bank.com emails are properly authenticated, then they were already been identified and displayed (in step 3). Therefore, email classified as *PhishOrReal* at step four, is almost certain to be phishing email, and can be blocked. Furthermore, since this identification is automated and with high confidence, the system can respond to it in ways that will penalize the phishermen, e.g. alert blacklists and other reputation mechanism, or traceback and punish the phisherman; we later model this by a relatively high cost c_f to the phishermen from such 'step 4 blocking'. This simple design for the classification phase is illustrated in Figure 2.

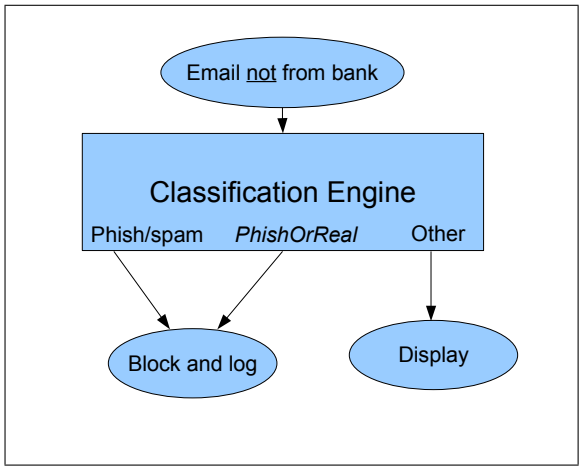

Fig. 2 Design of the classifier phase, using arbitrary classifying engine.

Notice that trustworthy senders, e.g. VIC-Bank.com, often use distinctive visual identification such as company name, trademarks, and logos; we refer to such visual identifiers as the bank's *letterhead*. We believe that users can be educated to look for the bank's letterhead and to suspect emails containing variants of it; when users suspect the email, we can try to detect this (possibly by user signaling, e.g. 'spam' button), and then penalize the phisherman; however we expect the penalty (cost) c_u to the attacker due to a user suspecting the email, to be much smaller than the cost c_f when the email is filtered by the classifier (step 4), i.e. $c_u << c_f$.

To avoid detection by the user, thereby losing c_u as well as any potential gain from phishing message, the phishermen will have to try to clone VIC-Bank.com's letterhead in phishing messages, which will make it easier to classify these messages as *PhishOrReal* emails. This places the phishermen in a dilemma: if he sends messages that are more likely to mislead the user, then these messages are also more likely to be *PhishOrReal*-classified; and on the other hand, messages that are less likely to be *PhishOrReal*-classified, are also less likely to mislead the user. In addition, the phisherman will be wary of using 'evasion techniques' designed to avoid classification as *PhishOrReal*, since the classifier may detect these technique and directly classify the email as 'phishing'.

We model this trade-off by assuming that the attacker can select the probability of the message being *PhishOrReal*-classified, p_f, and the probability of the user ignoring the message, p_u, but only *as long as their sum* is below some threshold x, i.e. $p_u + p_f \leq x$. Notice that this dilemma holds only if the phishermen are not able to send messages that pass the authentication (otherwise, these messages will be delivered even if they are *PhishOrReal*-classified).

It is desirable to evaluate the ability of users to detect phishing emails when a company uses (different types of) letterheads. Furthermore, it would be interesting

to evaluate the ability of phishermen to create letterheads, that users will consider as legitimate email from VIC-Bank.com, yet would not be classified as *PhishOrReal* (or as spam/phishing) by the content classifier. Such evaluation is challenging and requires careful, long-term usability studies, to ensure reliable results and to maintain ethical standards, and is therefore beyond the scope of this paper; see e.g. [8, 13, 7]. Notice that there may be significant impact to the design and consistency of using the letterhead, on the ability of the classifiers and the users to detect *PhishOrReal* and suspect emails, and on the ability of the phishermen to trick both classifier (to consider message as 'other' - neither phish not *PhishOrReal*) and user (to consider the message as valid message from bank). For example, intuitively, we may expect an advantage to simple textual letterheads, compared to more elaborate letterheads involving graphics and (dynamic) HTML; of course, this intuition should be validated experimentally.

3 Analysis of Effectiveness

In this section we present a simple economical model, and use it to analyze the effectiveness of an email anti-phishing filtering system. Our analysis focuses on the design we presented in the previous section (and in Figure 1), but it is applicable to many practical email filtering systems. The goal of our analysis is to identify sufficient conditions, under which the phishermen is likely to lose more, in average, per phishing message (due to costs due to detection), than the average profit he hopes to make from the message (due to profits when it succeeds in reaching and misleading the user). Our analysis is focused on the utility for the phishermen; we do not consider the expected utility to the user, which is mostly impacted by the false positives and false negative ratios, and the costs associated with the filtering mechanism.

Figure 3 illustrates the processing upon receipt of a phishing message by the filtering system. The filter first applies the authentication and repudiation mechanisms, which filter out messages from reputable, known senders such as VIC-Bank.com, as well as messages from known spammers and phishermen. The probability of filtering in these steps appear unrelated to the probability of filtering by the classifier and of trust by the user, and related to expenses for the phisherman (e.g. to use many IP addresses). Therefore, for simplicity, we ignore this probability, i.e. our analysis is for a phishing email that is *not* filtered by the authentication and reputation mechanisms (steps 1-3).

Phishing email is often classified as 'phishing' or *PhishOrReal* in both cases, it is 'suspected' and therefore blocked, and since this is automated, high-confidence detection, this result in significant penalty (cost) c_f to the adversary. We assume that the phisherman can determine the probability of classification as 'phishing' or *PhishOrReal* by the classifier, by appropriate selection of the contents of the email. Namely, we assume that the classifier suspects the email with probability p_f, controlled by the phishermen.

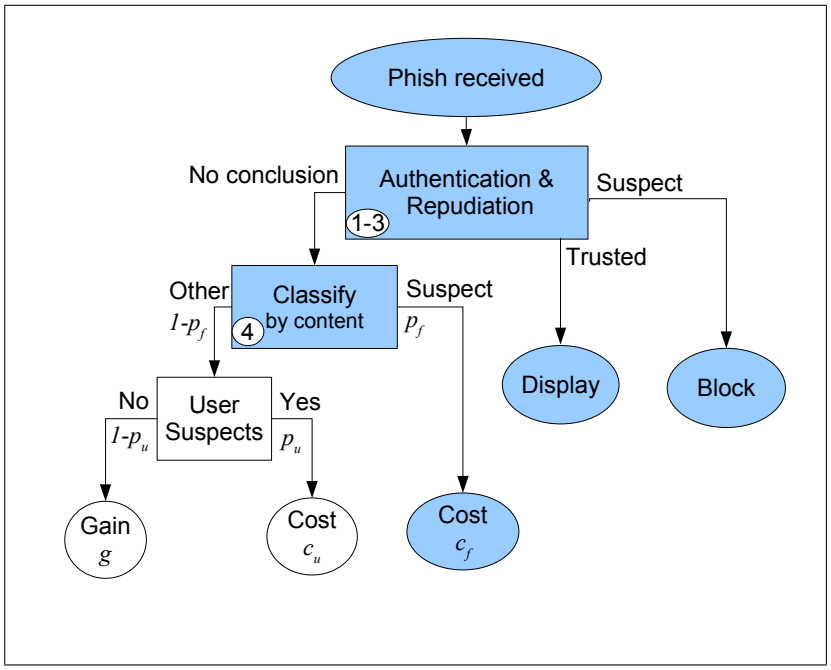

Fig. 3 Processing an incoming phishing email. We assume phishing email never authenticates successfully (as legitimate bank email). With probability p_f, it is detected as 'Phish' by the classifier, with cost c_f to the phisherman. Otherwise, it is displayed to Vic. With probability p_u, Vic suspects the email (and either ignores it or reports it), with average cost c_u to the phisherman. With probability $1-p_u$, Vic trusts the email, with average gain g to the phisherman.

Email which is *not* suspected by the filter, is displayed to the user Vic. With probability p_u, the user will suspect the phishing email. In this case, the user may report this phishing email, or the system may detect that the email is phishing by user's reaction to it; this may impact the phishermen, e.g. by reduction of reputation (or entering the phisherman's IP address to a blacklist). We denote the amortized cost to the phisherman due to each time the user suspects the email, by c_u. We expect c_u to be non-negligible, yet much smaller than the penalty due to the (higher-confidence, automated) detection by the classifier, i.e. $c_u << c_f$.

The phisherman may try to find and send messages that minimize p_u and p_f, and in fact, finding a message that minimizes only one of the two is usually easy. However, the challenge to the phisherman is to minimize both p_u and p_f. We model this constraint of the adversary, by assuming that the phisherman can only find messages s.t. $p_u + p_f \geq x$, where x is some bound on the ability to minimize both probabilities. Clearly $0 < x < 2$; and based on typical detection rates in usability testing and on typical precision of classifiers, it seems reasonable to expect that typically $0.5 < x < 1.5$. Since we assumed that x is fixed, the attacker can only select p_u (and then use $p_f = x - p_u$).

The attacker gains only if the email is displayed to the user, which then does not suspect it. This happens with probability $(1-p_f)(1-p_u)$. Let g denote the amortized gain to the phisherman from each such successfully displayed phishing email.

While a more detailed analysis can be done, we will show that two simple conditions, s.t. if one of them holds, phishing is not profitable. Specifically, the two sufficient conditions are:

$$\left(g \leq c_f - 2c_u\right) \wedge \left(g < \frac{c_u \cdot x}{1-x}\right) \wedge (x < 1) \tag{1}$$

$$\left(g \leq c_f - 2c_u\right) \wedge (x \geq 1) \tag{2}$$

We believe that these conditions are reasonable. In particular, the common condition of $g \leq c_f - 2c_u$ should hold, provided there is rapid, decisive response to confirmed detection of phishing emails (increasing c_f), together with the use of web-based phishing and other defenses, which can reduce significantly the amortized gain g to the phisherman from a displayed phishing message. In particular, $x \geq 1$ seems a reasonable goal for content classification systems.

Theorem 1 (Sufficient conditions for phishing to be unprofitable). *The maximal amortized utility of the phisherman $U^*_{g,c_u,c_f,x}$ for a phishing message received by the process in Figure 3 is non-positive, if $c_f > c_u$, and at least one of the two conditions 1, 2 above hold.*

Proof: The phisherman's amortized utility for a message received, U, is the following function of the 'fixed' parameters g, c_u, c_f, x and of the user detection probability p_u ($0 \leq p_u \leq \min(1,x)$):

$$\begin{aligned} U_{g,c_u,c_f,x}(p_u) &= -p_f \cdot c_f + (1-p_f)(-p_u \cdot c_u + (1-p_u)g) \\ &= (p_u - x) \cdot c_f + (1 - x + p_u)(-p_u \cdot c_u + (1-p_u)g) \end{aligned}$$

Which gives:

$$U_{g,c_u,c_f,x}(p_u) = -(g+c_u)p_u^2 + \left(c_f + (x-1)c_u + x \cdot g\right) \cdot p_u + (1-x)g - x \cdot c_f \tag{3}$$

Let $\hat{p}_u = \arg\max_{p_u}\left(U_{g,c_u,c_f,x}(p_u)\right)$, i.e. the value of p_u bringing phisherman's utility U to maximum, ignoring the restriction $0 \leq p_u \leq \min(1,x)$. Since the utility in Eq. 3 is simply a parabola, $\hat{p}_u$ is given easily as:

$$\hat{p}_u = \frac{c_f - c_u}{2(g+c_u)} + \frac{x}{2} \tag{4}$$

Since in both conditions 1 and 2 holds $g \leq c_f - 2c_u$, we have:

$$\hat{p}_u \geq \frac{1}{2} + \frac{x}{2} \geq \min\{x, 1\} \tag{5}$$

Let $p_u^* = \arg\max_{0 \le p_u \le \min(1,x)} \left(U_{g,c_u,c_f,x}(p_u) \right)$, i.e. the value of p_u bringing phisherman's utility U to maximum, *considering* the restriction $0 \le p_u \le \min(1,x)$. The maximal utility for the phisherman is $U^*_{g,c_u,c_f,x} = \max_{0 \le p_u \le \min(1,x)} \left(U_{g,c_u,c_f,x}(p_u) \right)$, i.e. $U^*_{g,c_u,c_f,x} = U_{g,c_u,c_f,x}(p_u^*)$. We next analyze the following cases:

1. $1 \le x$ and $1 \le \hat{p}_u$, i.e. $g \le \frac{c_f - c_u}{2-x} - c_u$. In this case, $p_u^* = 1$, hence trivially phisherman's utility for message received is negative.
2. $x \le 1$ and $x \le \hat{p}_u$. Since $x \le 1$, condition 2 definitely does not hold; hence we can assume that condition 1 holds, and in particular that $g \le \frac{c_u \cdot x}{1-x}$.
3. $\hat{p}_u < 0$. In this case, $p_u^* = 0$. This happens if and only if $c_f \le c_u(1-x) - xg$. However, this contradicts our assumption that $c_f > c_u$. Therefore, this case never holds.
4. Otherwise, i.e. $0 \le \hat{p}_u \le \min\{1,x\}$. In this case, $p_u^* = \hat{p}_u$. However, from Eq. 5, it follows that this case cannot hold (if either condition 1 or condition 2 hold).

It remains to analyze case 2, i.e. $x \le 1$ and $x \le \hat{p}_u$. Since $\hat{p}_u = \frac{c_f - c_u}{2(g+c_u)} + \frac{x}{2} \ge x$, we have $c_f \ge c_u(1+x) + g \cdot x$, or equivalently $g \le \frac{c_f - c_u}{x} - c_u$.

Since the parabola is monotonously increasing, the phisherman uses $p_u^* = x$, and his utility is at most:

$$\begin{aligned} U^*_{g,c_u,c_f,x} &= U_{g,c_u,c_f,x}(x) \\ &= -(g+c_u)x^2 + \left(c_f + (x-1)c_u + x \cdot g\right) \cdot x + (1-x)g - x \cdot c_f \\ &= (1-x)g - x \cdot c_u \end{aligned}$$

However, since we know that condition 1 holds here, and in particular that $g \le \frac{c_u \cdot x}{1-x}$, we see that the utility cannot be positive, i.e. also in this case phishing is not profitable. □

Acknowledgements Many thanks to Jim Fenton and Nathaniel (Nathan) Borenstein for their extremely detailed and helpful feedback, and to Ahmad Jbara, who participated in early discussions about this research. Thanks also to Haya Shulman for her assistance.

This work was supported by Israeli Science Foundation grant ISF 1014/07.

References

1. Abu-Nimeh, S., Nappa, D., Wang, X., Nair, S.: A comparison of machine learning techniques for phishing detection. In: L.F. Cranor (ed.) Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit 2007, Pittsburgh, Pennsylvania, USA, October 4-5, 2007, *ACM International Conference Proceeding Series*, vol. 269, pp. 60–69. ACM (2007). URL `http://doi.acm.org/10.1145/1299015.1299021`
2. Allman, E., Callas, J., Delany, M., Libbey, M., Fenton, J., Thomas, M.: DomainKeys Identified Mail (DKIM) signatures. Internet Request for Comment RFC 4871, Internet Engineering Task Force (2007). URL `http://tools.ietf.org/html/4871`

3. del Castillo, M.D., Iglesias, Á., Serrano, J.I.: An integrated approach to filtering phishing E-mails. In: R. Moreno-Díaz, F. Pichler, A. Quesada-Arencibia (eds.) Computer Aided Systems Theory - EUROCAST 2007, 11th International Conference on Computer Aided Systems Theory, Las Palmas de Gran Canaria, Spain, February 12-16, 2007, Revised Selected Papers, *Lecture Notes in Computer Science*, vol. 4739, pp. 321–328. Springer (2007). URL http://dx.doi.org/10.1007/978-3-540-75867-9_41
4. Dhamija, R., Tygar, D., Hearst, M.: Why phishing works. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI2006), pp. 581–590. Montreal, Quebec, Canada (2006)
5. Duan, Z., Gopalan, K., Yuan, X.: Behavioral characteristics of spammers and their network reachability properties. In: Proc. of the International Conference on Communications (ICC), Glasgow, UK (June 2007)
6. Fette, I., Sadeh, N.M., Tomasic, A.: Learning to detect phishing emails. In: C.L. Williamson, M.E. Zurko, P.F. Patel-Schneider, P.J. Shenoy (eds.) Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007, pp. 649–656. ACM (2007). URL http://doi.acm.org/10.1145/1242572.1242660
7. Herzberg, A., Jbara, A.: Security and identification indicators for browsers against spoofing and phishing attacks. IEEE Transactions on Internet Technology (2008)
8. Jakobsson, M., Ratkiewicz, J.: Designing ethical phishing experiments: a study of (rot13) ronl query features. In: WWW '06: Proceedings of the 15th international conference on World Wide Web, pp. 513–522. ACM Press, New York, NY, USA (2006). DOI http://doi.acm.org/10.1145/1135777.1135853
9. Leiba, B., Borenstein, N.S.: A multifaceted approach to spam reduction. In: CEAS 2004 - First Conference on Email and Anti-Spam (2004)
10. Lieba, B., Fenton, J.: DomainKeys Identified Mail (DKIM): Using digital signatures for domain verification. In: CEAS 2007: The Third Conference on Email and Anti-Spam (2007)
11. Lyon, J., Wong, M.W.: Sender ID: Authenticating E-mail. Internet Request for Comment RFC 4406, Internet Engineering Task Force (2006)
12. Resnick, P.: Internet message format. Request for comments 2822 (2001)
13. Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J.I., Nunge, E.: Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In: L.F. Cranor (ed.) Proceedings of the 3rd Symposium on Usable Privacy and Security, SOUPS 2007, Pittsburgh, Pennsylvania, USA, July 18-20, 2007, *ACM International Conference Proceeding Series*, vol. 229, pp. 88–99. ACM (2007). URL http://doi.acm.org/10.1145/1280680.1280692
14. Wong, M., Schlitt, W.: Sender Policy Framework (SPF) for authorizing use of domains in E-mail, version 1. Internet Request for Comment RFC 4871, Internet Engineering Task Force (2006). URL http://tools.ietf.org/html/4408