Scalable Detection and Isolation of Phishing

Giovane C.M. Moura and Aiko Pras

University of Twente
Design and Analysis of Communication Systems (DACS)
Enschede, The Netherlands
{g.c.m.moura,a.pras}@utwente.nl

Abstract. This paper presents a proposal for scalable detection and isolation of phishing. The main ideas are to move the protection from end users towards the network provider and to employ the novel bad neighbourhood concept, in order to detect and isolate both phishing e-mail senders and phishing web servers. In addition, we propose to develop a self-management architecture that enables ISPs to protect their users against phishing attacks, and explain how this architecture could be evaluated. This proposal is the result of half a year of research work at the University of Twente (UT), and it is aimed at a Ph.D. thesis in 2012.

1 Introduction

With the rapid growth of services like e-business and electronic banking, *phishing* is quickly becoming one of the main security threats for the Internet. In recent years we have witnessed a shift from incidents, towards a well organized criminal phishing "industry". Despite of the fact that the real direct losses (money stolen from users) caused by phishing is still unknown – since usually banks and companies do not disclose this information – some estimates indicate that these losses could, in the United States, be between $ 61 million [1] and $ 3.2 billion [2] per year. In addition, phishing also inflicts indirect costs (*e.g.*, costs associated to the staff to deal with it, network and computer resources) and compromises the trust of Internet users in e-mail and the web [1].

In order to fight phishing, several approaches have been proposed in the last few years. These approaches can be categorized into two major areas: (i) the "human factor" and (ii) techniques and algorithms to detect phishing. The former comprises social and psychological studies, the design of user interfaces, and the education of users and providers [3,4]. The latter presents different approaches to protect the user against phishing. Since phishing attacks usually consist of two phases (distribution of malicious URLs via spam, and connecting to malicious URLs by victims), there are proposals that focus specifically on detecting phishing e-mail [5] and others that focus on blocking the access to spoofed URLs. In relation to these, most of the approaches are client-centric, i.e., they require software installed at the client side (*e.g.*, a browser extension module [6,7,8]) and often rely on Content Black Lists (CBLs) for finding suspicious words. Commercial solutions also exist; the Cyveillance Anti-phishing

R. Sadre and A. Pras (Eds.): AIMS 2009, LNCS 5637, pp. 195–198, 2009.

[9] has developed a proprietary Internet monitoring technology that claims to identify activities associated to phishing, such as suspicious domain registrations and spoofed web sites.

Despite all these efforts, a Gartner survey shows that phishing attacks numbers are, in fact, increasing [2]. We consider that this is mainly because current approaches present limitations such as: (i) Internet Service Providers (ISPs) are not directly involved (which means that anti-phishing approaches rely on end users no longer using outdated software), (ii) lack of scalability (detection of phishing messages puts a heavy load on mail servers), (iii) lack of automation (maintenance of blacklists is cumbersome), and (iv) non-persistent IP addresses (current approaches have problems dealing with the dynamics of botnets). In this paper we introduce an approach to address these limitations. The reminder of this paper is organized as follows: in Section 2 the approach is explained, in Section 3 the envisioned architecture is described – and how it can be evaluated – and conclusions are presented in Section 4.

2 Approach

To address the limitations of the current proposals, we present an approach that provides a scalable and automated framework to protect users against phishing attacks. The main ideas behind this approach are:

Move protection to the network: Instead of relying on end users, we propose to move protection towards the ISP, who should block malicious servers (by updating rules at firewalls and routers) and, in this way, offer transparent protection for *all* end users. ISPs could offer this as a service, in the same way as British Telecom's Managed Security Services are offered to enterprises [10].

Analysis of flows: To ensure scalability, we propose the use of flow analysis instead of deep packet inspection. Moreover, with flow based approaches it is possible to analyze flow *patterns* and compare the network behaviour of multiple sources.

Self-management: The high number of phishing attempts makes it impossible to manually cope with such attempts. Complete automation is desired to make detection and isolation fast and reliable. The idea is to develop an integrated architecture that is able to detect and block phishing. More details about the architecture are described in Section 3.

Bad neighbourhood concept: To cope with botnets and non-persistent IP addresses, we propose the novel bad neighbourhood concept. The idea behind bad neighbourhoods, is that the likelihood a certain IP address behaves badly, increases if neighbour IP addresses (systems within the same subnetwork) behave badly. The assumption is that we can distinguish between well managed subnetworks, in which the probability of misbehaviour is very small, and badly managed networks, in which the probability of misbehaviour is quite high. Bots will primarily be found in the later kind of networks. For example, Figure 1

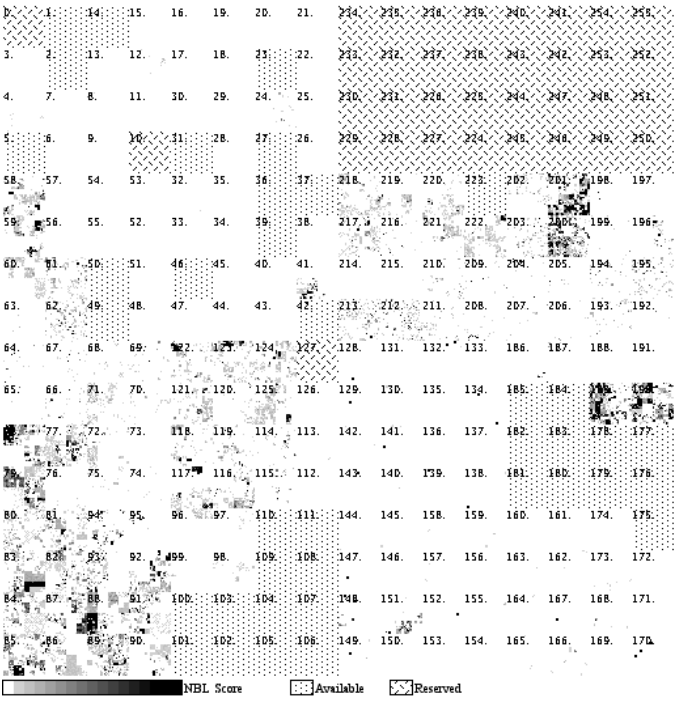

Fig. 1. Visualization of bad neighborhoods for spam senders (2008-11-01)

clearly shows that spam senders present this pattern, highlighting which subnetworks generate most spam. This picture was developed using different network blacklists. An algorithm took the blacklists data as input, and generated as output a Network Black List (NBL) score. Each point in the figure represents the NBL score for a subnet level of /24, using a Hilbert space filling curve [11]. We believe that many phishing e-mail senders and phishing web servers can also be found in bad neighbourhoods. This would provide us with statistical guidelines on the trustworthiness of neighbours (same subnetwork) IP addresses, which could be used as input to anomaly detection techniques in order to mitigate phishing attempts.

3 Architecture and Evaluation

To prove the concept and technical feasibility of our proposal, we will develop a self-management architecture for phishing detection and isolation. This architecture includes several elements, namely: e-mail servers, network routers, network firewalls, telescopes (specific kinds of honeypots, distributed all over the network, to capture and detect phishing messages), phishing e-mail senders, phishing web servers, and a phishing detector.

The architecture will be evaluated by building prototypes of key components. Data from real networks (like Géant, Surfnet and/or UT) will be used, such as

network flows and telescopes logs. Next, the idea is to analyze the data using self-learning techniques and the bad neighbourhood concept. This will result into two different lists. The first one contains IP addresses that distribute phishing messages, while the second contains names and IP addresses of phishing web servers. The first list could be used as input to mail servers for detection and removal of phishing messages. The second list could be used as input for network routers and firewalls, to block users from malicious web servers. In is important that both lists are maintained and employed in an automated way.

4 Conclusions

Phishing is a important problem that, despite all efforts, still causes significant monetary losses. In this research we propose to develop and evaluate an architecture to detect and isolate machines associated to phishing activities. A novel element in this architecture is the bad neighbourhood concept, of which the merits will be further investigated as part of this Ph.D. research. This paper was supported by the EC IST-EMANICS Network of Excellence (#26854). We would like to thank Ward van Wanrooij for his ideas on the bad neighbourhood concept and for providing Figure 1.

References

1. Herley, C., Florencio, D.: A profitless endeavor: Phishing as tragedy of the commons. In: Proc. of the ACM SIGSAC New Security Paradigms Workshop, Lake Tahoe, California, USA (September 2008)
2. McCall, T.: Gartner survey shows phishing attacks escalated in 2007 (2008)
3. Dhamija, R., Tygar, J.D., Hearst, M.: Why phishing works. In: CHI 2006: Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 581–590. ACM, New York (2006)
4. Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L.F., Hong, J.: Lessons from a real world evaluation of anti-phishing training. In: eCrime Researchers Summit, pp. 1–12 (2008)
5. Fette, I., Sadeh, N., Tomasic, A.: Learning to detect phishing emails. In: WWW 2007: Proceedings of the 16th international conference on World Wide Web, pp. 649–656. ACM, New York (2007)
6. Zhang, Y., Egelman, S., Cranor, L.F., Hong, J.: Phinding phish: Evaluating anti-phishing tools. In: Proceedings of the 14th Annual Network and Distributed System Security Symposium (NDSS 2007), San Diego, CA, USA (2007)
7. Dhamija, R., Tygar, J.D.: The battle against phishing: Dynamic security skins. In: SOUPS 2005: Proceedings of the 2005 symposium on Usable privacy and security, pp. 77–88. ACM, New York (2005)
8. Phishing Protection Design Documentation (2009), `https://wiki.mozilla.org/safe_browsing:_design_documentation`
9. Cyveillance Anti-Phishing (2009), `http://www.cyveillance.com/`
10. British Telecom Managed Security Services (2009), `http://bt.counterpane.com/managed-security-services.html`
11. Irwin, B., Pilkington, N.: High level internet scale traffic visualization using hilbert curve mapping. In: VizSEC 2007: Proceedings of the Workshop on Visualization for Computer Security, pp. 147–158. Springer, Heidelberg (2007)