网络钓鱼攻击的防御技术及防御框架设计

赵跃华，胡向涛†
（江苏大学 计算机科学与通信工程学院，江苏 镇江 212013）

摘　要：现有的反钓鱼技术多是通过检测可疑网站与真实网站的 URL 和网页相似度来检测钓鱼攻击，而对于新出现的在网页中嵌入恶意代码的钓鱼攻击防御效果不佳。在分析当前的网络钓鱼攻击防御技术的基础上，针对传统方法不能防御的新型钓鱼攻击给出了解决方法，并融合传统的防御技术提出了一个防御钓鱼攻击的整体框架，弥补现有防御方法的不足，从而提高了钓鱼攻击的检测率，降低了漏报率。实验结果表明，提出的方法是有效的。

关键词：网络钓鱼；跨站脚本；木马；键盘记录
中图分类号：TP393.08　　**文献标志码**：A　　**文章编号**：1001-3695(2013)06-1863-04
doi:10.3969/j.issn.1001-3695.2013.06.069

Defense technology of phishing attack and design of defense framework

ZHAO Yue-hua, HU Xiang-tao†
(*School of Computer Science & Telecommunication Engineering, Jiangsu University, Zhenjiang Jiangsu* 212013, *China*)

Abstract: The current anti-phishing technologies detect the phishing attacks through comparing similarity of URL or Web pages between suspect Website and real Website. However, it cannot work effectively on those new attacks by the insertion of malicious code to Web pages. Based on analysis of existing defense technologies of phishing attacks, this paper put forward solutions to those new attacks which couldn't be prevented by traditional methods. It also proposed a further new overall defensive framework which incorporated traditional ones, to make up defects of the original methods, increased the detection rate of phishing attacks, and decreased the false negative rate. Experimental results show that the proposed defense methods are effective.
Key words: phishing; XSS; Trojan; key logger

1　网络钓鱼攻击及其现有防御方法的分析

互联网为人们的生活带来了方便和快捷的同时，也带来了威胁，网络欺诈事件频频发生，而网络钓鱼攻击是网络欺诈的典型代表。网络钓鱼对用户的隐私和个人财产构成了严重威胁，对用户的信任感造成了恶劣的影响，并且严重阻碍了电子商务的发展。金山发布的《2010 年中国网络购物安全报告》指出，2010 年网购用户的人均经济损失已由 2009 年的 80 元上升至 150 元左右[1]。因此，加强网络钓鱼攻击防范对策的研究具有很大的实用价值。网络钓鱼(phishing)一词融合了 phone 和 fishing，起源于 1996 年左右。国际反钓鱼工作小组(APWG)指出，网络钓鱼是一种在线身份伪造的欺诈方式，它使用社交工程和技术伎俩等手段来达到窃取客户个人身份数据和金融账号等敏感信息的目的。

传统的网络钓鱼攻击利用域名欺骗、URL 隐藏、IP 地址欺骗、欺骗性的超链接、Unicode 编码等欺骗技术进行伪装，通过欺骗性的电子邮件、假冒网上银行等形式将受害者引诱到一个与真实网站极其相似的仿冒网站，从而使受害者泄露自己的银行卡账号密码等隐私信息。

随着网络钓鱼攻击技术的不断发展，网络钓鱼又出现了新的攻击形式，即通过在真实的网站中嵌入恶意代码来进行钓鱼攻击。这种新型网络钓鱼攻击主要有两种表现形式，一种是利用 XSS 漏洞以合法的身份将一段恶意的 XSS 脚本上传到 Web 服务器，或者把一个同样含有恶意脚本代码的 URL 链接发送给目标用户，这些脚本将被攻击者引向一个精心构造的钓鱼网站，甚至直接利用脚本构造一个收集用户信息的对话框；另一种则是在网页中嵌入木马程序，当用户中了木马后利用其键盘记录功能监控用户的输入从而窃取用户的账号密码等信息。

目前网络钓鱼攻击防御方案可分为服务器端防御机制、客户端防御机制和预防性防御机制三个类别。在服务器端防御网络钓鱼攻击会加大服务器的负担，降低服务器的性能；预防性机制会造成大量的域名闲置；在客户端防御机制主要是在客户端浏览器上嵌入网络钓鱼防御插件，这种方法要求每个用户都添加防御措施。为了不降低服务器性能，从保护潜在受害者本身出发，大部分的防御方法将防御放在了客户端。

客户端防御的方法有：

a)通过对 URL 地址进行识别，检测是否是钓鱼网站，包括基于 URL 黑名单技术和基于机器学习的 URL 检测技术。Ma 等人[2]提出通过学习检测一个可疑的 URL 是否为恶意网站。黄华军等人[3]提出的基于异常特征的钓鱼网站 URL 检测技术通过分析钓鱼网站 URL 地址的结构特征和词汇特征，抽取 URL 中的 12 个特征向量，用 SVM 进行训练和分类。郑礼雄等人[4]提出的基于域名信息的钓鱼 URL 探测使用编辑距离寻找与已知正常域名相似的域名，根据域名信息提取域名单次最大

收稿日期：2012-10-09；**修回日期**：2012-11-30
作者简介：赵跃华(1958-)，男，江苏苏州人，教授，博士，主要研究方向为信息安全；胡向涛(1987-)，男(通信作者)，河南邓州人，硕士研究生，主要研究方向为信息安全(huxiangtao1000@163.com).

匹配特征、域名分割特征和 URL 分割特征,利用上述特征训练分类器,由此判断其他 URL 是否是钓鱼 URL。

b)通过对 Web 页面的识别进行网络钓鱼检测,包括对页面的异常特征分析判定和 EMD 视觉相似识别技术进行检测。郭敏哲等人[5]提出了基于页面文档对象模型分析的网络钓鱼页面异常检测算法,将其与改进后的黑名单技术和已知攻击特征检测进行结合构建出一个较为完整的网络钓鱼检测机制。Pan 等人[6]提出基于文档对象模型的 Web 页面异常检测的方法,该方案不需要用户具有专业的反钓鱼知识,且具有较低的误报率和漏报率。曹玖新等人[7]提出了一个基于嵌套 EMD 的网页相似度判定算法,对 Web 图像进行分割,抽取子图特征并构建网页的 ARG,在计算不同 ARG 属性距离的基础上采用嵌套 EMD 方法计算网页的相似度。

c)对 URL 和 Web 页面内容综合特征进行识别、分类,判断钓鱼攻击。何高辉等人[8]提出了一个基于 SVM 主动学习的网络钓鱼检测方法,该方案对 URL 进行黑/白名单库过滤后,对既不在黑名单库中也不在白名单库中的可疑 URL 进行检测分类,采用 URL 和 Web 页面的综合特征,确保网络钓鱼检测的全面性,从 URL 中提取出七个敏感特征,从 Web 页面中提取出六个敏感特征,然后将这些特征数据转换为 SVM 所需的向量形式,在分类器中对这些敏感特征进行学习,然后对得到的 SVM 分类器进行评估。当 SVM 分类器达到一定的性能时可以用该分类检测模型对可疑的 URL 进行检测。

基于黑名单的检测技术具有很高的效率,但是漏报率很高,并且具有很大的滞后性;基于机器学习的 URL 检测技术可以检测未知的钓鱼网站而且具有很高的效率,但是这种 URL 的检测技术具有较高的误判率和漏判率,因为 URL 中并不具有钓鱼攻击网站的决定性特征,即窃取用户信息的手段。基于 Web 页面识别的技术可以提高检测的准确率,降低漏误报率,但是检测速度和效率都不是很高[8]。基于对 URL 和 Web 页面综合特征进行识别分类的方法,利用 SVM 主动学习算法进行分类,提高了检测的速度和效率而且提高了检测率,弥补了各自的缺点。这些防御方法可以防御传统的网络钓鱼攻击,但它们共同的缺陷是对于在网页中嵌入恶意代码的新型网络钓鱼攻击无能为力,因为新型钓鱼攻击并不是对真实的网页进行模仿,网页的 URL 和页面并没有异常。

2 新型网络钓鱼攻击的防御

2.1 XSS 型网络钓鱼攻击的防御

2.1.1 防御方法的分析

XSS 型网络钓鱼攻击主要是利用反射型 XSS 漏洞和保存性 XSS 漏洞将一段恶意的脚本代码上传到 Web 服务器,或者把一个同样含有恶意脚本代码的 Web 站点的 URL 链接发送给目标个人用户,而这些恶意脚本将用户引诱到一个精心设计的钓鱼网站,或者将直接构造一个收集用户信息的表单。当个人用户访问了含有恶意脚本代码的页面或者打开收到的 URL 连接时,恶意脚本就会执行,从而遭受钓鱼攻击。

XSS 漏洞的种类和形式繁多,通过寻找一个网页中是否存在跨站漏洞进而判断是否是钓鱼攻击的工作量巨大,同时一个网站中存在跨站漏洞但不一定被攻击者所发现和利用,但是这类钓鱼攻击的发生必须将 XSS 脚本通过注入点注入到网页中。从攻击者可能注入恶意脚本的注入点出发,判断这些注入点注入的内容是否用于钓鱼攻击的恶意脚本。而所有的注入脚本既会出现在 HTTP 请求中,又会出现在 HTTP 响应中,因此只需在一个恰当的近似度量下匹配含有特殊关键词的用户输入数据(即请求参数与服务器响应中的可疑字符串),当相似度超过一定的阈值就认定是 XSS 攻击,然后再检测这些脚本是否用于钓鱼攻击。这样可以检测出几乎所有的反射型和部分存储型 XSS 钓鱼攻击。对于部分存储型攻击只需检测截获响应中的显示内容的标签是否有脚本,然后检测这些脚本是否用于钓鱼攻击,最后将这些脚本进行编码以文本的形式显示。本方法的核心是用户请求参数的提取、可疑脚本的提取及其匹配。

对于用户请求参数的提取可以通过截获 HTTP 请求来获得。HTTP 协议若以 Get 方式提交,则请求参数位于 URL 的“?”之后;若以 Post 方式提交,则请求参数位于 HTTP 请求头的空行之后,参数之间通过“&”分割。由于任何合理且复杂的脚本都不会短于 15 个字符,并且数字、纯字母字符等参数字符串不可能构成恶意字符串,因此简单、短小的参数不提取。

HTML 源代码中的 URL 链接、事件句柄、脚本环境的值通常是引进动态脚本来执行特定的行为操作,恶意用户主要在这三处嵌入代码来执行超出自己权限的操作,这三处出现用户输入数据通常是一种异常行为,因此可将这三处提取出的含有用户提交内容的字符串认定为可疑脚本。可疑脚本的提取从注入脚本的注入点出发,这样可以减少匹配文本的数量,提高检测效率。然后将之与可疑脚本进行匹配,求出两者的最长公共序列,若最长公共序列的长度达到阈值则进一步检测该公共序列是否用于钓鱼攻击。由于一次有效的反射型攻击字符串的长度不小于 15,因此阈值选取 15。

目前一些攻击者为绕过过滤通常将请求参数进行编码(URL 编码、JavaScript 编码或 HTML 编码)处理,而可疑脚本也可能进行编码。所以为防止漏报,本文方法可将提取的请求参数和可疑脚本进行解码处理。

具体流程如图 1 所示。

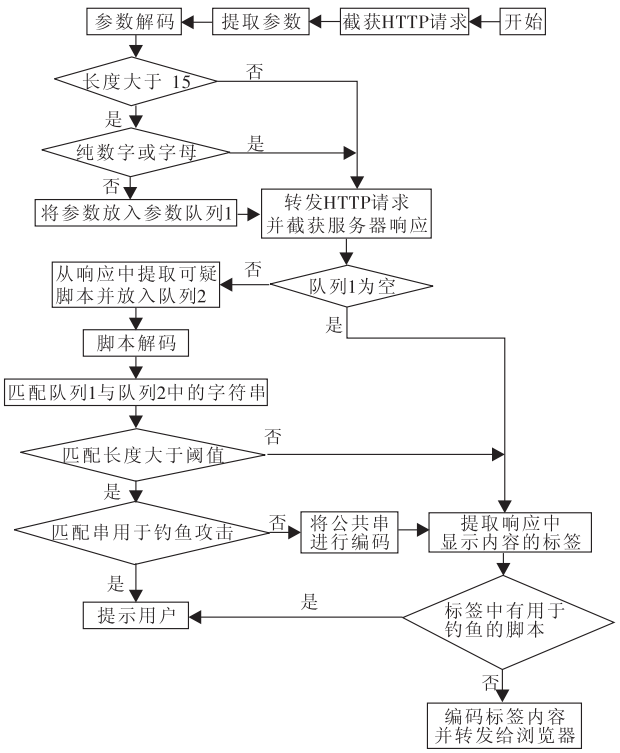

图1 XSS型钓鱼攻击防御流程

2.1.2　防御方法的测试

通过样本测试对本方法的正确性进行测试。样本为一个利用 XSS 构造的钓鱼 URL,即 http://localhost:8090/login.aspx?param = %3C%69%66%72%61%6D%65%20%73%72%63%3D%22%68%74%74%70%3A%2F%2%77%77%77%2E%68%78%74%2E%63%6F%6D%20%22%65%72%66%6C%6F%77%3D%68%69%64%64%65%6E%20%3E,当用户提交了该 URL 后,就会向服务器发送 HTTP 请求,通过截获该 HTTP 请求获取输入参数,截获的参数如图 2 中阴影部分所示。该参数是经过编码的,通过解码后为 http://localhost:8090/login.aspx?param = 〈iframe src = " http://ww.hxt.com" frameborder = 0〉。当浏览者点击该 URL 后会被引向一个收集用户信息的页面,通过截获 HTTP 响应并解码提取可疑字符串,提取的字符串如图 3 中阴影部分所示。通过匹配,其公共字符串为〈iframe src = " http://ww.hxt.com" frameborder = 0〉,超过阈值并认定为 XSS 攻击,又利用 iframe 标签,进一步认定为 XSS 型钓鱼攻击,提示用户。由于利用 iframe 标签,恶意 URL 地址和真实地址一样,所以传统的检测方法无法检测到这种钓鱼攻击,但本方法可以检测到这种攻击。

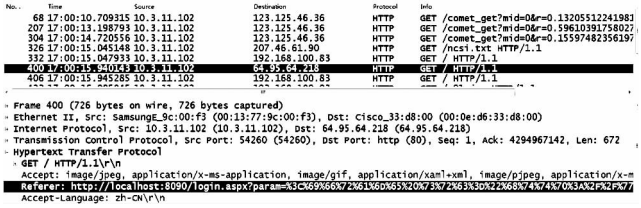

图 2　截获 HTTP 请求参数

```
<body>
    <form name="form1" method="post" action="login.aspx?param=%3ciframe+src%3d%
 22http%3a%2f%2fwww.hxt.com%22+frameborder%3d0%3e" id="form1">
<div>
<input type="hidden" name="__VIEWSTATE" id="__VIEWSTATE"
 value="/wEPDwUJMTQzNjAzMjMxZGQv9HB/qvsYSiPiiqX18KtYduC2Qw==" />
</div>

    <div>
    <iframe src="http://www.hxt.com" frameborder=0>
    </div>
    </form>
</body>
</html>
```

图 3　HTTP 响应中提取的可疑字符串

2.2　利用木马进行网络钓鱼攻击的防御方法

利用木马进行钓鱼攻击是指恶意攻击者利用网站漏洞在正常的网页中嵌入恶意代码进行网页挂马,当用户访问了这样的网站之后,就会在后台将木马服务器端下载到本地并自动运行,而木马的键盘监控功能对用户输入的信息进行记录,从而窃取用户的隐私信息。根据键盘记录位置的不同将键盘记录分为用户层键盘记录和内核层键盘记录。

2.2.1　用户层键盘记录及其防御方法的分析

用户层键盘记录的方法有很多种,归纳起来有SetWindowsHookEx、GetKeyboardState、GetKeyState、GetAsyncKeyState、GetRawInputData。

通过调用这些函数可以直接或间接地进行键盘记录,这些函数是由 user32.dll 提供的,它是 Win32 子系统的动态链接库。Win32k.sys 是 Win32 子系统的内核模式。在应用层调用这些函数最终会通过调用内核模式 Win32k.sys 中实现的系统服务调用来完成函数的操作,如 SetWindowsHookEx 在内核中与之对应的是 NtUserSetWindowsHookEx,而这个对应关系是通过表 KeServiceDescriptorTableShadow 指向确定。为此,可以通过修改这个表 hook 挂钩这些 NtUserXXXXXX 函数。当可疑进程调用这些函数进行键盘记录时首先调用的是已经替换成自己的函数,在自己的函数中判断是否允许执行。

详细方法是:先获得 KeServiceDescriptorTableShadow 的地址,通过函数的系统调用号在系统服务描述符表中找到对应函数的地址,然后替换原函数的地址为自己的函数地址,若有应用程序调用用户层的上述 API 函数,在内核中会首先调用反键盘记录程序的 hook 函数,在自己的函数中通过 PID 号判断是否是恶意的键盘记录行为,并通知用户有进程调用这些敏感函数,根据用户的选择进行下一步操作。用户层反键盘记录流程如图 4 所示。

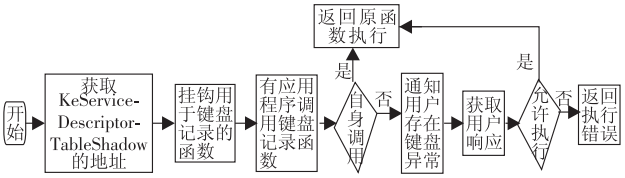

图4　用户层反键盘记录流程

2.2.2　内核层键盘记录及其防御方法的分析

内核层键盘记录的方法主要有 IDT hook、安装键盘过滤驱动、hook 键盘类驱动的分发函数、hook KeyboardClassServiceCallback 函数。

通过对这些键盘记录方法的分析发现,这些方法大部分是 hook 可以获取键盘记录的函数,将原始的函数地址替换成自己的函数地址,然后在自己的函数中实现过滤从而获取键盘记录。为此通过检测这些 hook 点的地址,与原始地址进行比对,如果地址不一致则被挂钩了,用原始地址进行恢复。为了进一步提高隐蔽性,部分键盘记录并不在函数调用点就替换原函数地址,而是使用 inlinehook 技术,即在原函数中修改代码,用一个跳转指令或其他指令跳转到自己的函数中,在自己的函数中进行键盘记录。对于这类方法,通过比对其与原函数代码是否一致来判断,如果不一致则用原函数代码进行恢复。对于安装键盘过滤驱动的键盘记录方法,只需查看设备链上是否有过滤驱动,如果有将其摘除即可。具体实现如下:

针对 IDT hook 型键盘记录的防御方法可通过比对内存中 IDT 的键盘中断处理函数地址与原始的键盘中断处理函数地址是否一致来判断是否有键盘记录程序;键盘过滤驱动型键盘记录的防御方法可通过检测 KeyboardClass0 设备对象的 Attached-Device 域是否为空,若不为空则说明有键盘过滤驱动,将其摘除;hook 键盘类驱动的分发函数型键盘记录的防御方法可通过检测 KeyboardClassRead 函数的调用点 hook 和 inlinehook 进行判断,检测 KbdClass 驱动的读请求分发函数地址是否和从 PE 文件 kbdclass.sys 中找到的地址一致,如果不一致则将其恢复,逐条语句比对 KeyboardClassRead 函数的代码是否和从 PE 文件中得到的一致,如果不一致则用真实代码恢复;hook KeyboardClassServiceCallback 型键盘记录的防御方法和 hook 键盘类驱动的分发函数型键盘记录的防御方法类似,只是检测的函数不同。

内核层防御反键盘记录的整体流程为:首先初始化 ContextData 数据结构,该数据结构中保存一些用于比对的原始数据;接下来依次进行检测键盘中断处理函数并恢复;扫描键盘过滤驱动并恢复;扫描 KeyboardClassRead 函数并恢复;扫描 KeyboardClassServiceClassCallback 函数并恢复。内核层防御反键盘记录整体流程如图 5 所示。

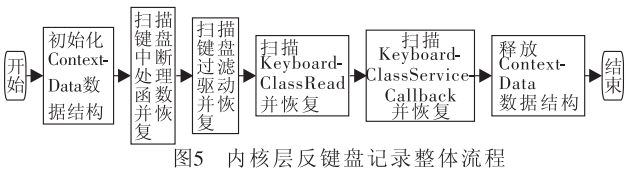

图5 内核层反键盘记录整体流程

2.2.3 键盘记录防御方法的测试

通过键盘记录样本测试对本方法的正确性进行验证。应用层样本为 WH_KEYBOARD 键盘钩子,内核层样本为分别采用 IDT hook、键盘过滤驱动、hook 键盘类读分发函数、hook KeyboardClassServiceCallback 四种技术的键盘记录程序。测试环境为虚拟机,OS 为 XP SP3,采用本方法后会对应用层调用键盘记录的函数进行监控,如图 6 所示。当该应用层样本通过调用 SetWindowsHookEx 进行键盘记录时,就会被监控函数处理,发现是非本身进程所调用就会向用户发出报警。内核层测试过程为首先用上述样本对用户的输入进行记录,然后开启反键盘记录功能。内核层键盘记录检测的结果如图 7 所示,图中阴影部分表明本方法成功地发现了这四种样本攻击。通过测试表明,本方法可以很好地防御这些类型的键盘记录,从而保护了利用木马形式的网络钓鱼攻击。

两种测试结果表明,本文提出的防御方法可以很好地防御利用 XSS 和木马两种形式的新型网络钓鱼攻击。

图 6 应用层监控的用于键盘记录的函数

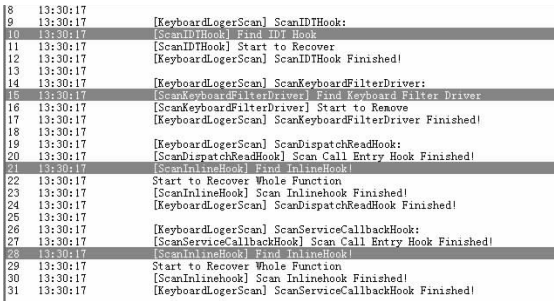

图 7 内核层键盘记录检测结果

3 网络钓鱼攻击的整体防御框架设计与分析

本文提出的防御方法仅针对新型网络钓鱼攻击,对于传统的网页模仿形式的攻击不能检测,而传统的防御方法对新型的钓鱼攻击无能为力。为了更全面地防御各种钓鱼攻击,本文在综合新旧防御方法的基础上,通过互相弥补各自的缺陷提出了一个综合性防御框架。该框架在传统的反钓鱼方法的基础上添加抵御在网页中嵌入恶意代码形式的钓鱼攻击的模块、XSS 检测模块和反键盘记录模块用于防御新型钓鱼攻击。为了管理的方便,使用人机交互模块。通过黑名单数据库、地址分析模块、异常检测模块防御传统的钓鱼攻击。整体防御框架如图 8 所示。

详细检测过程如下:采用多线程方法每当用户打开了一个网页后,首先判断请求的网址是否在黑名单数据库中,如果在,则通过人机交互模块提示用户并关闭页面;否则由异常检测子系统对用户请求的页面进行检测。由地址分析模块对用户请求的地址进行分析,初步判断是否是通过各种域名欺骗技术进行的伪装。如果是则通过人机交互模块提醒用户关闭页面,并将该网址加入到黑名单数据库中;否则将该页面提交给 XSS 攻击检测模块,判断是否是利用 XSS 脚本进行的钓鱼攻击,如果是则提示用户关闭页面,并将该地址加入到黑名单数据库中。如果没有检测到 XSS 形式的钓鱼攻击,则将该页面提交给页面异常检测模块,判断是否是传统的通过网页模仿进行的钓鱼攻击。本文采用基于 SVM 主动学习和适合小样本集的分类方法对网站进行分类判断,如果是则提示用户关闭页面并更新黑名单数据库。再判断页面中是否有表单进行交互,如果有则开启反键盘记录模块,对用户的输入进行保护,否则结束该线程。

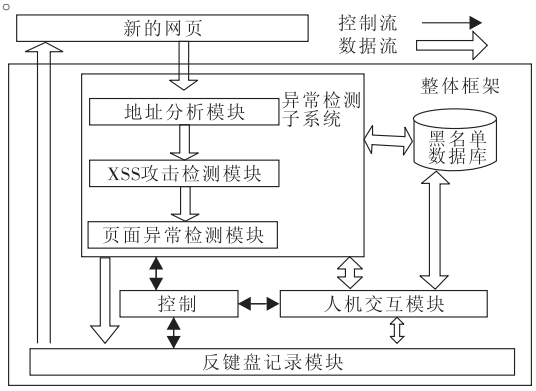

图8 整体防御框架

采用多线程和黑名单的技术可以提高判断速度,采用 SVM 主动学习的方法提高了分类的性能,但由于新增加了新型钓鱼攻击检测模块,势必会降低效率,但相对于用户的重要隐私信息,牺牲一部分的效率是值得的。

4 结束语

本文通过对网络钓鱼攻击和现有的反钓鱼技术进行分析,发现现有的反钓鱼方法对于新出现的钓鱼攻击不能很好地防御。针对新型网络钓鱼攻击给出了防御方法并通过样本测试证明了防御方法的可行性,最后在传统反钓鱼方法的基础上通过添加防御模块提出了一个综合性的反钓鱼框架,弥补了新旧防御方法的各自缺陷,从而提高了检测率,降低了漏报率。

参考文献:

[1] 中国电子商务研究中心. 2010 年中国网络购物安全报告[EB/OL]. (2010-12-22). http://china.toocle.com/cbna/item/2010-12-22/5572682.html.

[2] MA J, SAUL L K, SAVAGE S. Beyond blacklists: learning to detect malicious Web sites from suspicious URLs[C]//Proc of ACM SIGKDD. New York: ACM Press, 2009: 1245-1253.

[3] 黄华军,钱亮,王耀钧. 基于异常特征的钓鱼网站 URL 检测技术[J]. 信息网络安全,2012(1):23-25.

[4] 郑礼雄,李青山,李素科,等. 基于域名信息的钓鱼 URL 探测[J]. 计算机工程,2012,38(10):108-110.

[5] 郭敏哲,袁津生,王雅超. 网络钓鱼 Web 页面检测算法[J]. 计算机工程,2008,34(20):161-163.

[6] PAN Ying, DING Xu-hua. Anomaly-based Web phishing page detection[C]//Proc of the 22nd Computer Security Applications Conference. 2006: 381-392.

[7] 曹玖新,毛波罗,军舟,等. 基于嵌套 EMD 的钓鱼网页检测法[J]. 计算机学报,2009,32(5):922-929.

[8] 何高辉,邹福泰,谭大礼,等. 基于 SVM 主动学习算法的网络钓鱼检测系统[J]. 计算机工程,2011,37(19):126-128.