

基于浏览器的钓鱼网站检测技术研究

梁雪松

(四川教育学院物理系, 四川 成都 610041)

【摘要】目前, 网络钓鱼攻击给互联网用户带来严重的威胁。为了应对这种威胁, 许多软件厂商与组织提出了各种反钓鱼策略。论文针对基于浏览器的钓鱼网站检测技术进行了分析研究。

【关键词】网络钓鱼; 黑名单; 白名单; 启发式分析

【中图分类号】TP393.08 【文献标识码】A 【文章编号】1009-8054(2007) 11-0053-03

Techniques of Detecting Phishing Sites Based on Browser

LIANG Xue-song

(Dept. of Physics, Sichuan College of Education, Chengdu Sichuan 610041, China)

【Abstract】Nowadays, phishing attacks have brought serious threats to Internet users. To cope with these threats, many software vendors and organizations have put forward a variety of anti-phishing schemes. In this paper, some techniques for detecting phishing sites based on browser are proposed.

【Keywords】phishing; blacklist; whitelist; heuristics

0 引言

网络钓鱼是一种网络诈骗手法, 主要通过电子邮件、网页等途径散布虚假信息, 诱骗不知情的网络用户连上仿冒的网站(也称为钓鱼网站), 比如假冒的网上银行、在线购物等, 骗取用户的网银账号和密码等重要信息。据反钓鱼工作组(APWG)统计: 2007年3月共发生24,853次钓鱼攻击, 比上月份多出1000多次; 钓鱼网站达到了2,0871个, 其中位于中国(包括香港、台湾)的钓鱼网站数仅次于美国与韩国^[1]。

近年来, 出现了大量商业和免费的反钓鱼工具。许多工具以浏览器插件的形式存在, 比如Microsoft Phishing Filter^[2]、Google Safe Browsing^[3]、TrustWatch^[4]、SpoofGuard^[5]、雅虎助手等, 使用黑名单、白名单、启发式分析等多种技术识别钓鱼网站, 当用户访问的网站被其判定为钓鱼网站时, 会通过弹出安全提示框等途径警示用户。本文针对目前基于浏览器的反钓鱼工具所使用的几种钓鱼网站

检测技术作了如下研究。

1 钓鱼网站检测技术

1.1 黑/白名单技术

黑名单技术是将所有已经发现的钓鱼站点记录到一个地址列表(即所谓的黑名单)中, 据此判断用户所访问的网站是否为钓鱼网站。黑名单技术实现简单, 问题在于及时更新黑名单十分困难, 因为钓鱼网站往往只会存在一段时间。据APWG统计, 钓鱼网站的平均在线时间仅有4天^[1]。因此黑名单必须在尽可能短的时间内进行更新, 才能有效地阻塞钓鱼网站。白名单与黑名单正好相反, 其中列出了可信的网站地址, 用于判断用户所访问的网站是否为合法网站。

1.2 启发式分析

启发式分析采用了评分的方法对网站分类, 以求减小误判的可能性。它使用一组已知钓鱼网站的特征, 对网站进行特征分析, 每项分析结果都会被赋予一个权值, 如果权值总和超过了给定的阈值, 那么一个网站就会被判定为钓鱼网站。目前启发式分析主要从以下方面对网站进行分析。

(1) URL分析。钓鱼网站的URL往往具有欺骗性, 因而分析网站的URL是辨别网站真伪的重要手段。目前常用的分析方法主要有:

相似域名检测。钓鱼网站的域名往往在发音或形式

收稿日期: 2007-05-30

作者简介: 梁雪松, 1972年生, 男, 讲师, 工程硕士, 研究方向: 计算机教学和计算机教育。

上与真实网站相近,期望用户不会发现它们之间的差异而上当。针对相似域名问题,目前一种常用的解决方法是设置一个基于域名的白名单,如果一个网站的域名与白名单中某个合法网站的域名相似,有可能就是钓鱼网站。

检查基于IP的URL。钓鱼网站的URL往往使用IP代替主机名,主要一个原因在于许多钓鱼网站是通过僵尸主机建立的,而这些主机通常未申请域名,因此可对网站URL中的IP进行反向域名解析,检查其域名是否存在。

端口检测。一些ISP阻塞了个人用户的80端口,防止个人用户私自架设网站以进行某些非法活动(比如架设钓鱼网站)。为了绕过上述检测技术,一些钓鱼网站就使用了非80端口。因而从URL中所获取的网站端口号也能作为识别钓鱼网站的一个依据。

检查URL中“.”的数目。Google、Yahoo、AOL等合法网站出于某些原因,提供了URL重定向服务。比如“http://rd.yahoo.com/*http://www.sohu.com/”,使用了Yahoo的重定向服务,将URL重定向至搜狐网。不幸的是,网络欺诈者往往利用这些重定向服务,对用户隐藏其URL的真实目的地。一些钓鱼网站使用了互联网上免费的泛域名解析服务。泛域名解析是指在DNS服务器中使用通配符“*”,将一组域名解析为同一IP,例如下列的DNS记录:

```
*.mybank.com IN A 192.168.1.1
```

将所有以“.mybank.com”结尾的域名都解析为192.168.1.1。利用泛域名解析服务,钓鱼网站可以使用欺诈性的域名,比如“www.ccb.cn.mybank.com”,可以使许多用户误以为是中国建设银行网站。另外,通过不断更改子域名,以图绕过黑名单技术的检测。采用上述两种欺骗技术的URL有一个明显的特征,即URL中“.”的数目往往大于4。因此一个网站的URL中“.”的数目大于4,有可能就是钓鱼网站。此外,检测URL中是否包含@字符,检查是否存在IDN(国际化域名)欺骗等,也是目前常用的URL分析方法。

(2) 域欺骗检测。所谓域欺骗(Pharming)是指利用域名劫持、DNS缓存投毒、修改用户主机中的hosts文件等形式,将用户对合法网站的访问请求重新导向钓鱼网站。为了应对域欺骗,一些基于浏览器的反钓鱼工具主要采取以下两种技术,以检测用户是否被劫持到了钓鱼网站。

维护一个合法网站主机名与IP地址的映射表。检查当前网站URL中的主机名及其解析出的IP地址是否与映射表匹配。

根据网站URL中的主机名解析出的IP地址,再使用可信的DNS服务器进行反向域名解析,判断解析出的主机名与URL中的主机名是否相同。

(3) 分析SSL证书信息。合法的商业网站通常会对安全敏感的网页,比如用于提交用户账号、密码的网页,启用SSL安全连接机制,以防止信息在传输过程中被窃听、篡改。与之相对应的是,绝大多数仿冒的网站未采用SSL连接,或者使用伪造的或过期的SSL证书认证的连接。基于上述原因,安全敏感的网页的SSL相关信息,包括是否启用了SSL安全连接、颁发SSL证书的CA是否权威可信、SSL证书是否过期、证书中的识别名(Distinguished Names, DN)是否与网站的身份相符等等,也可作为识别网站真伪的一个依据。需要注意的是,有些合法的网站建立了自己的CA,并使用自签的SSL证书,因而该方法可能会引起误判。

(4) 查询第三方数据库。

查询WHOIS数据库。WHOIS数据库存放了互联网中各域名的注册信息,包括域名所有者的相关资料(包括名称、地址、联系方式、管理员等)、域名的注册时间、有效期限等。这些信息有助于识别钓鱼网站。由于钓鱼网站的生存期较短,因而其域名往往是新近注册的,或具有较短的有效期。另外域名注册信息不够详细的网站,比如未提供必要的联系方式、管理员信息等等,也极可能是仿冒的。然而值得注意的是,网络欺诈者可以假借被仿冒的合法机构的名义注册域名,因而WHOIS数据库的可靠性存在问题。

查询网站的网络排名。相对于合法网站而言,钓鱼网站的流量、访问量、站点流行度(或反向连接数)等方面低得多,因而根据这些数据计算出的网站排名也较低。基于上述理由,可以通过查询Alexa获取网站的流量、访问量、站点流行度、综合排名等信息,也可以从Google、Yahoo等知名的搜索引擎中获取网站的排名,作为识别钓鱼网站的一种依据。但这种方法同样也存在一些问题,网络欺诈者可能通过某些欺诈手段,来提高自己的钓鱼网站的排名。此外,该方法也不利于识别新建立的合法网站。

查询网站分类目录。网站分类目录是一个人工编辑管理的网站目录集合,为网络用户或搜索引擎提供网站查询服务,同时也有利于目录内网站的推广。目前较权威的网站分类目录有DMOZ(The Open Directory Project: 开放目录项目)、Yahoo!、门户搜索引擎目录搜狐等。如果一个网站希望登录某个权威的分类目录,必须首先提交登录申请,该分类目录的工作人员对提交申请的网站的各方面(包括网站规模、网站内容等)进行严格的审查,通过后才能最终登录到该分类目录中。由此可见,通过查询权威的分类目录,也有利于判别一个网站的真伪。这种方法的问题在于,许多合法的网站由于某些原因并未登录到权威的分类目录中。

(5) 分析网页内容。合法网站会在其网页的HTML元素中声明自己的网站名、所属机构名等身份信息。这些元素主要有<TITLE>标记、<META NAME="KEYWORDS"/> "DESCRIPTION"/> "COPYRIGHT">标记中的CONTENT属性、<IMG/INPUT/AREA/OBJECT>标记中的ALT属性等。而钓鱼网页往往是直接在合法网页的基础上稍加改动而来的,同时为了迷惑用户,在其HTML元素中仍会保留合法网页中的身份信息。基于上述原因,我们可使用关键词匹配等方法,查询网页的HTML某些元素中是否引用了其他合法网站的身份信息。

钓鱼网站通常会使用被仿冒网站的商业标志logo。因此,一个合法网站的logo出现在与该网站无关的某个网页中,那么该网页就可能是一个钓鱼网页。然而考虑到一个合法网站的logo也可能被其他某些合法网站引用,比如Visa、Paypal等网站的logo会经常出现在一些合法的在线购物网站中,因而该方法主要针对敏感网页进行检测,以减少误判的可能性。基于成本等方面的考虑,钓鱼网站一般只会模仿或克隆合法网站中的部分网页,因而其钓鱼网页往往包含了大量指向外部域(通常是被仿冒网站的域)的链接,或者不指向任何目的地的空链接,比如:。由此可见,一个网页中包含的异常链接数目越多,其可疑的程度也就更大。

利用网页内容检测钓鱼网站的另一种常用方法是近似网页检测技术——用于比较两个网页的相似程度。目前该技术主要采用多指纹相似性比较算法。算法基本思想是:计算出网页的一组指纹(Fingerprint),若两个网页拥有一定数量的相同指纹,认为这两个网页的内容重叠性较高,即二者是近似的。这种算法需要布置一个动态的指纹数据库,存放合法网站的首页或敏感网页的指纹集。

(上接第52页)

果需要重新设置用户的权限,只需改变其所关联的角色即可,这样就简化了授权管理操作,减少了授权出错的可能性,有效地保障了系统的安全。系统运行时,在工作流引擎的统一调度下,用户执行相应的操作,并与指定的其他用户通过交互类进行信息传递,形成了端到端的业务流程。


3 结语

Web2.0增值业务安全运营支撑系统必将为Web2.0网站的不断发展提供有力的支持。如何根据业务特点,正确

2 结语

本文分析了基于浏览器的反钓鱼工具所采用的多种钓鱼网站检测技术。需指出的是,目前的钓鱼网站检测技术还不成熟,黑名单技术时效性差,而启发式分析基于已知钓鱼网站的特征,容易被更先进的欺骗手段绕过,导致了反钓鱼工具对钓鱼网站存在一定程度的漏检^[3]。因此,反钓鱼攻击仍是一项非常艰巨的工作。

参考文献

- [1] Anti-Phishing Working Group. Phishing Activity Trends Report[EB/OL]. http://antiphishing.org/APWG_Report_March_2007.pdf, 2007.
- [2] Microsoft corp. Microsoft Phishing Filter: A New Approach to Building Trust in E-Commerce Content [EB/OL]. <http://www.microsoft.com/downloads/>, 2005.
- [3] Robichaux, Paul. Gone Phishing: Evaluating Anti-Phishing Tools for Windows [EB/OL]. <http://www.3sharp.com/projects/antiphishing/gonephishing.pdf>, 2006.
- [4] GeoTrust corp. GeoTrust Introduces Industry's First Secure Consumer Search Service[EB/OL]. http://www.geotrust.com/about/news_events/press/PR_TrustedSearch_092605s.pdf, 2006.
- [5] Chou N, Ledesma R, Teraguchi Y, et al. Client-side defense against web-based identity theft[R]. Proceedings of the Network and Distributed System Security Symposium, 2004. 

架构系统,合理运用各种技术,是随着Web2.0增值业务的发展需要进一步研究的问题。

参考文献

- [1] TM Forum. GB920 Release1.0, The NGOSS Approach to Business Solutions[S]. 2005.
- [2] TM Forum. GB921 Release6.0, eTOM Solution Suite Release 6.0[S]. 2006.
- [3] 陈龙,张春红,云亮,等. 电信运营支撑系统[M]. 北京:人民邮电出版社,2005:46~72.
- [4] 赵卫东,陈杰. 基于对象的角色工作流模型研究[J]. 计算机工程,2004,30(5):87~89. 