基于蜜罐在线恶意网页检测系统研究与设计①

康松林　胡赐元　孙永新 (中南大学 信息科学与工程学院 湖南 长沙 410083)

摘　要： 针对网页安全威胁的动态性、广泛性等特点，设计了一个基于蜜罐在线恶意网页检测系统。该系统使用 URL 数据表来记录网页地址，同时结合蜜罐技术对 URL 数据表不存在或存在但还需进行检测的网页进行综合检测，实时检测出用户需要浏览的网页的安全状态，避免恶意网页的攻击，从而提高人们网络活动的安全性。

关键词： 蜜罐；在线检测；恶意网页；URL 数据表；实时性

Research and Design of On-Line Mal-Webpage Detection System Based on Honeypot

KANG Song-Lin, HU Ci-Yuan, SUN Yong-Xin

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract: As web threats are dynamic and universal, an on-line mal-webpage detection system based on honeypot is designed. In this system, the URL table with some special fields is designed to store URLs, and the honeypot is used to detect those Webs that are not in table or in table but need detection once again. The system can detect the safe state of Webs in realtime with more accuracy.

Keywords: honeypot; on-line detection; mal-webpage; URL tables; real-time

越来越多的入侵事件表明恶意网页已经成为木马病毒传播的最主要手段，网页安全威胁成为了人们网络活动的十大安全威胁之首[1]。虽然现有杀毒软件以及各种网页安全浏览辅助工具对含有已知恶意代码的网页能够有效拦截，但对包含未知恶意技术的网页的检测却总有力不从心的感觉。而且随着黑色网络产业链的盛起以及管理员的安全意识不断增强，网页的安全状态是不断变化的，人们对所关注网页的安全状态实时了解更显重要，为此提出了基于蜜罐的在线恶意网页检测系统，利用蜜罐技术实时检测出用户需要浏览网页的安全状态，并根据用户的需求能够分析并返回该网页的其他安全信息。

1　相关概念

1.1 蜜罐技术

蜜罐技术指的是通过精心设置网络陷阱来引诱攻击者进行攻击并同时监控入侵者在陷阱里的一举一动来分析攻击者的技术[2-4]。一般根据与攻击者的交互程度，可以把蜜罐分为两种[5,6]：高交互蜜罐和低交互蜜罐。前者是使用真实的网络设备和系统等充当诱饵，与攻击者进行高度交互活动以获得更多有关入侵者的信息，后者则通过模拟相关的网络服务来吸引攻击者入侵，以达到安全实用的效果。本文提出的检测系统将利用布置在虚拟机上的操作系统充当高交互蜜罐，使蜜罐能够检测出未知攻击类型的恶意网页，并同时利用 URL 数据表使在线检测的准确性和实时性都能够满足。

1.2 恶意网页

恶意网页指的是能对用户实施恶意行为的网页，一般利用以下三种技术[7]：

(1) URL 重定向技术：当浏览器打开一个 URL 时，网页代码就会指示浏览器自动打开其他的 URL，而不影响对用户显示的内容。它是正常网页为了丰富其内容而引用其他网页而设计的技术，但是当所引用的网页有恶意代码时，正常网页也同样具有恶意性了。通

① 基金项目:国家自然科学基金(60773013)
收稿时间:2009-05-19

常说的挂马就是利用这种技术。

(2) 漏洞利用技术：当浏览器或其上插件有漏洞时，恶意网页就会包含这些漏洞的利用程序以增大对用户的攻击成功几率。0day 漏洞利用技术扮演了一个非常重要的角色，它指的是利用一些没有公布或已公布但还没补丁的漏洞来达到攻击目的的技术，大部分恶意网页的未知攻击类型通常利用了这种攻击技术。

(3) 代码迷惑技术：针对很多检测技术都是基于特征扫描的，一些恶意网页就会使用如 escape()等函数将字符加密而变成不可读的长字符串，以及在脚本里使用 document.write()和 eval()函数进行代码动态注入等各种代码迷惑技术，以逃脱检测软件的扫描。

2 系统整体流程

一个在线的恶意网页检测系统应该满足两条要求：

(1) 实时性：即用户不能等太长的时间。

(2) 准确性：即能准确检测出网页是否有恶意。

这两个要求从本质上说是相互冲突的，很难在高实时性的同时保证高准确率。如何在两者之间选择一个折中的方案，是一个很难的问题。本系统利用在 URL 数据库中查找用户所请求检测的网页地址的恶意程度和能被入侵程度这两个字段，决定是否需要通过蜜罐进行检测，从而加快检测速度，并且在一定程度上保证检测的准确性。系统的整体流程如图 1 所示：

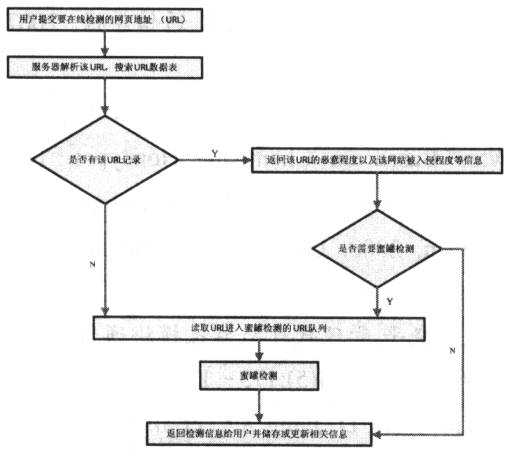

图 1　系统整体流程图

3 组成部件设计

3.1 URL 数据表设计

为了准确描述 URL 的恶意信息，URL 数据表主要包含两个字段：

(1) 恶意程度。表示该 URL 网页的恶意程度，分为三个等级：0 代表正常网页；1 代表该网页是攻击者设计的专门用来入侵用户的恶意网页；2 代表其余类型的恶意网页。这样的划分有助于当数据库中有用户所要查找的 URL 时，快速判断是否需要用蜜罐进一步检测。对 0 和 2 等级的处理，将结合 URL 数据库的另一字段值(被入侵程度)进行决定；对 1 等级的处理，则直接返回给用户该 URL 网页的恶意信息。

(2) 网页被入侵程度。表示该网页的脆弱程度，它是由多因素影响的，如网页源代码的编码安全性、所处平台的脆弱性以及管理员的安全意识强弱等。本系统主要根据该网页的知名度和使用扫描工具的扫描结果来综合判定的。它划分为 2 个等级：0 代表能被入侵程度少，1 代表能被入侵程度大。当然这样划分有点笼统，但它指明了网页以后的安全状态。

3.2 蜜罐检测模块设计

为获得更好的数据捕获与数据控制，本系统将蜜罐布置在虚拟机上。蜜罐检测分为浏览器引擎模块、监控模块和扫描模块。浏览器引擎模块负责启动浏览器软件，以及汇总页面代码的检测情况，监控模块负责对浏览网页的行为进行监控，扫描模块负责网页的脆弱性分析。

3.2.1 浏览器引擎模块

因为在线检查具有针对性，浏览器引擎模块根据用户请求中的浏览器类型以及版本等信息选择相应的浏览器打开 URL，并对页面代码进行如下三方面的检查：

(1) 对脚本进行恶意代码模式匹配。

(2) 对 iframe、script、object 等标签里的 URL 进行检查。

(3) 对已有漏洞特征进行模式匹配。

3.2.2 监控模块

监控模块是在浏览器引擎模块通知浏览器打开 URL 时，对系统进行监控[8]。主要监控这四个方面：

(1) 文件创建、修改等变化。重点监控存放网页的临时文件夹，以及系统目录等。

(2) 注册表的变化。因为这是蜜罐，注册表的任何异动都意味着存在恶意活动。

(3) 进程的创建。重点监控由浏览器进程创建的所有子进程。

(4) 网络行为监控。当有非法访问网络时，就意味着存在恶意活动。

3.2.3 扫描模块

扫描模块是对网页被入侵程度进行分析，主要分析这两个方面[9]:

(1) 知名度分析。 主要根据 Google 的 PageRank (PR)值度量。

(2) 脆弱性分析。利用 Nmap 对网站进行扫描以及其他知名网页脆弱性扫描软件对网页进行扫描，根据综合扫描结果度量脆弱性程度。

4 系统关键技术研究

本系统应用的几个关键技术如下:

4.1 URL 数据表字段值来源及处理

收集知名杀毒软件公司提供的恶意 URL 列表，并将它们的恶意程度标为 1，被入侵程度标为 1。当用户所要检测的网页 URL 不在数据库表中时，插入该 URL 记录，以及检测得到的字段值，当存在该 URL 时，把检测得到的值更新到相应字段。

对系统流程图中是否需要蜜罐检测过程的处理: 当恶意程度为 1 时，直接返回该 URL 的恶意信息; 当恶意程度和被入侵程度都为 0 时，直接返回用户该网页安全的信息。其余都将进入蜜罐检测环节。

4.2 蜜罐检测处理

蜜罐检测中的三个模块的检测结果由浏览器引擎模块统一处理，当浏览器引擎模块发现恶意代码或漏洞特征的模式匹配时，进行报警，返回给用户恶意网页的信息。未发现匹配时，则等待监控模块的信息。当接收到监控模块的报警信息时，浏览器引擎模块询问监控模块的报警原因,定位 URL,返回给用户该 URL 恶意网页的信息。当打开一个网页时，扫描模块对网页进行扫描，将扫描结果发送给浏览器引擎模块，当所扫描的网页的 RP 值大于 5 且脆弱性扫描软件报告无漏洞时，浏览器引擎模块把该 URL 的被入侵程度标为 0，其余情况标为 1。在有报警时把该 URL 的恶意程度标为 2，其余情况为 0。

5 结语

提出了利用蜜罐技术对网页进行在线检测从而判断出所要浏览网页是否安全的检测系统的整体流程以及相关关键部分的研究与设计，利用 URL 数据表的缓冲以及描述网页安全状态字段的设置使得在线检测的实时性和准确性都得到满足。后续工作主要是完成检测系统对恶意活动的分析以便扩充恶意代码匹配库的功能。

参考文献

1 Sophos. Security threat report: 2009 Prepare for this year's new threats.(2008)[2009-03-10] http://www.sophos. com/sophos/docs/eng/marketing_mate-rial/ sophos-security-threat-report-jan-2009-na.pdf

2 Spitzner L. The Value of Honeypots, Part One: Definitions and Values ofHoneypots.(2001-10) [2009- 03-10] http://www.Security focus.Com/infocus/1492

3 熊明辉,蔡皖东.基于主动安全策略的蜜网系统的设计与实现.计算机工程与设计, 2005,26(9):2470－2472.

4 冯朝辉,范锐军.Honeynet 技术研究与实例配置.计算机工程, 2007,33(5):132－134.

5 Dagdee N, Thakar U. Intrusion Attack Pattern Analysis and Signature Extraction for Web Services Using Honeypots. First International Conference on Emerging Trends in Engineering and Technology, Nagpur, 2008. USA: IEEE Computer Society, 2008:1232－1237.

6 诸葛建伟,韩心慧.HoneyBow:一个基于高交互式蜜罐技术的恶意代码自动捕获器.通信学报, 2007, 28(12):8－13.

7 Sun XY, Wang Y, Ren J, et al. Collecting Internet Malware Based on Client-side Honeypot. The 9th International Conference for Young Computer Scientists, Hunan, 2008. USA: IEEE Computer Society, 2008.1493－1498.

8 康松林,樊晓平.管理代理与监控模块通信设计.计算机应用研究, 2007,24(5):173－175.

9 Wang YM, Beck D, Jiang XX, et al. Automated Web Patrol with Strider HoneyMonkeys. Finding Web Sites That Exploit Browser Vulnerabilities. Proc. of the 13th Annual Network and Distributed System Security Symposium, San Diego, 2006.