



基于特征行为分析的木马病毒检测技术的研究^{*}

邹维福¹, 张翼英², 张素香², 杨成月³

(1. 国网泉州供电公司 泉州 362000; 2. 国家电网公司信息通信分公司 北京 100761;
3. 南瑞集团公司厦门亿力吉奥信息科技有限公司 厦门 361009)

摘要: 基于木马病毒行为特性, 提出基于数据挖掘的相似度技术的主动木马病毒检测及预防算法, 从规则化、行为过滤及自学习 3 个方面保证算法的完备性和有效性。首先, 依据木马病毒特征码及特征行为, 进行特征的规则化, 建立初始木马病毒规则库。其次, 建立进程行为特征捕捉及分析过程, 采用聚类分析方法, 完成行为特征规则化。最后, 通过对比规则库及相似度主动对比法, 完成对可疑进程的分析对比, 确定可疑进程性质。分析及实验表明该算法具有自主学习及主动防御特征, 很好地平衡了静态测试技术和动态测试技术的优缺点。

关键词: 木马病毒; 行为检测; 规则化; 安全

doi: 10.3969/j.issn.1000-0801.2014.11.018

Research on Anti-Trojan Malware Mechanism Based on Characteristic Behavior

Zou Weifu¹, Zhang Yiyi², Zhang Suxiang², Yang Chengyue³

(1. State Grid Quanzhou Electric Power Supply Company, Quanzhou 362000, China;
2. State Grid Information & Telecommunication Co., Ltd., Beijing 100761, China;
3. NARI Group Corporation Xiamen Great Power Gio Information Technology Co., Ltd., Xiamen 361009, China)

Abstract: Trojans inject systems and launch various attacks, such as eavesdropping secret information, tampering with system configuration etc., which threatens to system security seriously. A novel anti-Trojan malware mechanism based on characteristic behavior and cosine similarity was proposed. Firstly, according to the initial rules base and application behavior, the mechanism regularized the operations of application, and then, the mechanism invoked rules to judges suspicious behaviors based on current rules base and operational impact. Once the application was considered as Trojan malware, the system would dispatch the appropriate algorithm for processing. The mechanism triggered by sensitive behaviors, and had the active prevention function and self-learning function. The analysis and experiment show the solution can detect Trojan malware effectively.

Key words: Trojan malware, behavior detection, regularization, security

1 引言

网络技术的的发展和普及, 给人们的生活、工作方式带

来了巨大变化, 金融信息、隐私数据等日趋网络化, 网络生活已经成为人们生活的一部分。但是, 由此而带来的安全威胁也日益严重, 如网络欺诈、网络钓鱼、网站后门、拒绝

^{*} 国家高技术研究发展计划(“863”计划)基金资助项目(No. 2011AA05A117)



服务攻击等给网络及用户带来诸多风险。其中,木马病毒在计算机领域中指的是一种后门程序,通常包括服务端(服务器部分)和客户端(控制器部分),通过伪装和隐藏将服务端植入宿主对象,实现对用户计算机或网络资源的获取和控制,以非正常手段骗取正常权限,盗取用户信息、隐私信息及其他数据。

国家计算机病毒应急处理中心发布的《第十二次全国信息网络安全状况与计算机及移动终端病毒疫情调查分析报告》分析显示^[1],经过网络调查,2012年我国计算机病毒感染率为45.07%,通过网络下载或浏览传播病毒的比例占75.42%,其中因为木马等计算机恶意软件感染的事件为56.68%(注:多数病毒采用多种攻击方式,因此各种攻击比例和大于1),如图1所示。

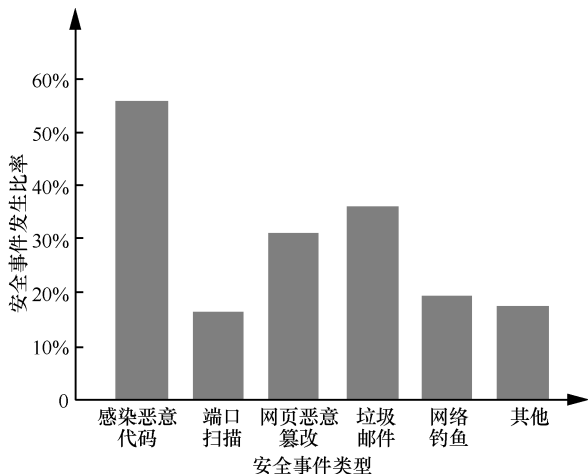


图1 2012年网络安全事件类型比例

在2012年,传播最广的病毒大部分为木马病毒。其中,Troj_Generic系、“代理木马”(Troj_Agent)系、“木马下载器”(Troj_Downloader)系等3系木马病毒分列前三位,见表1。木马病毒隐身植入于目标宿主,造成用户密码、账号被盗、远程控制、系统(网络)效能下降、浏览器等应用被修改等严重后果,使宿主用户面临各种威胁。

本文深入分析木马病毒原理及运行机理,综合静态分析、动态分析优点,提出基于行为分析的木马病毒检测技术及相关方法与工具。

2 木马病毒概念及运行机制

2.1 木马病毒

木马病毒通常具有很强的隐秘性,同时绑定网络、操作系统或一些应用而启动,通过欺骗伪装手段诱导用户激

表1 2012年传播最广的10种病毒^[1]

序号	病毒名称
1	Troj_Generic 及变种
2	“代理木马”(Troj_Agent)及变种
3	“木马下载器”(Troj_Downloader)及变种
4	Backdoor_Rbot 及变种
5	Kido 及变种
6	Troj_Startpage 及变种
7	“QQ 盗号木马”(Troj_QQPsw)及变种
8	Win32_AdWare.Undef 及变种
9	“新鬼影”病毒及变种
10	Download_Patched 及变种

活,向攻击者提供未授权的计算机或网络等资源,使攻击者可以非法获取有关资源。通常,木马病毒具有如下特点。

- 账户身份窃取。采集计算机或网络用户身份信息,如用户名、密码、凭证、cookie 表单等,进而利用身份信息进行非法活动。
- 敏感信息窃取。采集计算机或网络用户相关的敏感信息,包括本地存储介质、网络存储介质等的敏感信息,如金融文件、研究方案、项目信息、图片信息等,发送给木马客户端进行分类分析等,完成了对敏感信息的窃取。
- 操作权限窃取。攻击者通过木马实现对用户注册表、网络服务、监控服务、系统服务等远程控制,发动如拒绝服务攻击、网络钓鱼、垃圾邮件及分发恶意软件等攻击。

2.2 木马特性

目前有多种木马类型存在于各种系统中,它们依据不同的计算机系统执行着不同的攻击。虽然木马种类各异但有着许多共同的特征。主要包括以下几个特征^[2]。

(1)隐蔽性^[3]

作为间谍性程序,隐蔽性是木马最基本的特征。通常包括两种隐藏方式。

- 通信隐藏,主要包括通信信息、网络流量、端口号、通信信道等内容的隐藏。既可以隐藏通信信息,也可以隐藏通信信道的隐藏通道技术,是通信隐藏的关键技术。
- 本地隐藏,主要包括文件隐藏、进程隐藏、Rootkit 隐藏、编译器隐藏和自身加密隐藏等。

(2)欺骗性^[4]

木马病毒经常伪装成可用的或者诱惑的应用或资源

文档,比如图片、工具软件、游戏、功能文档等,用户一旦使用或下载这些应用或文档,实际下载的是木马程序或捆绑了木马程序的软件。

(3) 自启动性

木马病毒在宿主计算机上必须能够自动运行。一种方式是伴随系统自动启动,比如通过修改注册表/配置文档等相关启动项;另一种方式是绑定在系统程序或者其他应用程序上,或者通过更改名称等替换关键文件,如特洛伊 DLL 将木马 DLL 注册为 Windows 系统进程 Svchost 的服务,实现开机自启动。

(4) 自我保护性

木马病毒为了保证顽健性,通过设置自身的守护进程等实现木马的自我保护,如进行备份、删除恢复、进程重启等。

2.3 木马分类

随着操作系统、系统平台及算法技术的发展,木马病毒的制作和传播也发生很大变化。依据木马病毒的攻击行为、攻击影响,可以将木马病毒分成以下几种类型^[5]。

- 系统控制型。此种木马一旦入侵系统并获取部分系统权限,便实施对系统部分功能的控制。如监控用户输入输出端口、篡改系统配置、开放风险服务端口、删除对抗应用、下载非法软件/插件、发动 DoS/DDoS 攻击等,通过控制系统部分功能,实施相应的攻击。
- 数据破坏型。此种木马主要目标是实施数据破坏。一旦入侵成功,则扫描/获取系统相关数据资源,对数据进行破坏,如删除、更改等。通过此种攻击,可能进一步要挟用户赎回数据,避免数据及业务损失。
- 信息盗取型。此种木马一般实行隐蔽攻击,通过扫描系统,获取兴趣数据,如用户银行账户、密码、重要文档、隐私信息等,窃取后实施进一步攻击。

2.4 运行机制

木马病毒属于非复制病毒,一般采用 C/S (client/server, 客户端/服务器) 模式实现攻击。其客户端部署在实施攻击的机器上(常称为控制端),负责管理和控制受控端;而服务端以隐藏的方式运行在俘获的计算机上(常称为受控端),接收控制端控制命令实施攻击操作。随着软硬件技术及网络技术的发展,木马病毒的生成、传播和攻击方式都发生很多变化。木马攻击模型如图 2 所示,木马病毒实施攻击一般分为以下几步。

(1) 攻击者通过恶意代码搜索攻击目标扫描脆弱点,选择漏洞实施攻击,侵入宿主目标,完成入侵过程^[6]。

(2) 触发自启动机制并提取系统权限,创建后门接收攻击指令或实施攻击任务,包括问卷操作权限、动态链接更改、计算机操控权限等。

(3) 受控端通过网络接收攻击指令。

(4) 受控端解析收到的指令,进行指令分析,查询符合攻击条件资源并调用相关模块,实施攻击。

(5) 依据攻击指令要求,进行攻击成果反馈,隐藏并等待接收新的攻击指令。

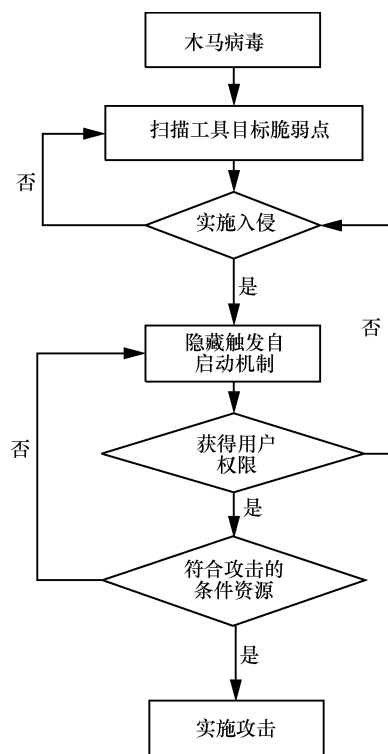


图 2 木马攻击模型

2.5 木马监测技术

由于木马病毒采用多种隐藏技术并利用系统漏洞,因此可以轻易获取系统控制权限,并躲过用户及一般工具的常规检查。木马监测技术主要深度分析各种木马病毒攻击机制,提出检测方案及解决办法^[7,8],通常包括以下几种:

- 网络监测技术,即通过对通信端口、网络连接等监测检查木马攻击行为,包括防火墙、入侵检测等技术;
- 完整性检测技术,即通过检查系统文件或目录的内容或属性等是否与可信版本一致,以发现是否存在木马攻击;



- 特征码扫描技术是将木马病毒的特征代码或特征值(通常为一串二进制字符串)作为扫描依据,因为特征码是木马存在的唯一标识,因此对于已存在的木马效果很好,很多实时扫描工具如 360、瑞星等采用此种技术进行病毒防御,但属于被动防毒,有一定的滞后性^[9];
- 虚拟机技术通过软件虚拟化建立独立的运行环境,对运行任务进行测试与检测,以期检测木马病毒存在性;
- 行为分析技术是一种主动的防御技术,通过启发式方法,依据木马病毒的自身特征及攻击原理,通过检测并分析恶意行为,实现对木马病毒的监测。

3 基于行为分析的木马检测技术

针对木马病毒特性,本文深入分析特征码技术和行为分析技术,立意对木马特征及恶意行为建立关联性,利用余弦相似度进行关联性分析,提出一个基于行为分析的木马监测模型,构建木马行为规则库,通过相似度判别算法,梳理木马恶意行为,实现木马病毒的有效检测。

3.1 模型设计

基于行为分析的木马检测模型主要由 3 部分组成,如图 3 所示。

- 木马病毒规则库建立,通过已知的木马特征码及特征行为,建立初始的木马病毒规则库以备参考及进一步学习添加新的规则。
- 特征行为采集,采用信息挖掘及统计算法,针对可

能的行为进行捕捉,建立行为特征规则。

- 行为过滤指对捕捉的时序下的进程行为特征进行分析^[10],包括两个方面:首先依据当前规则库进行评测,如果与规则库存在规则匹配,则可以直接确定为木马。否则,依据挖掘及分类技术算法进一步判定,依据判定结果进行相关处理。

3.2 行为特征采集

依据木马特征,通过监护进程对应用程序进程进行监视(尤其是木马常用 API 进程的挂钩),采集进程动作。部署监测点,针对热点信息访问,如访问注册表、自启动项、系统关键项、系统权限、重要文献等动作进行采集。依据挖掘和分类技术,记录并积累可疑动作的系列程序,通过关联算法进行识别及动作规则化。监控采集过程如图 4 所示。

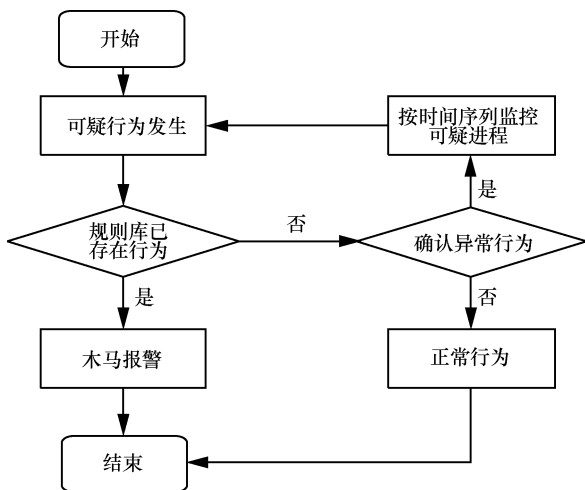


图 4 特征行为监控及采集

3.3 行为特征

针对采集的行为特征,参考木马病毒规则库进行对比分析^[6]。设某目标进程的规则化系列行为集合为 $R=\{r_1, r_2, \dots, r_n\}$, 其中, r_i 为规则化的某特征, $r_i.act$ 为该特征行为; $r_i.key$ 为该特征关键词; $r_i.weight$ 为该行为的特征权重。设木马规则库 $D=\{R_1, R_2, \dots, R_m\}$, 其中, R_i 为已经存在的特征。设 T 为定义的特征行为参考系(依据攻击的时间关联性,新的攻击通常与最近一次攻击相关,因此其缺省值选定为时间上最近的 R_i); $weight_i$ 为 T 行为的权重(包括特征码及特征行为); $threshold$ 为相似度阈值,其值计算如下:

$$T.weight = \frac{\sum_{i=1}^{|T|} T.r_i.weight}{|T|}$$

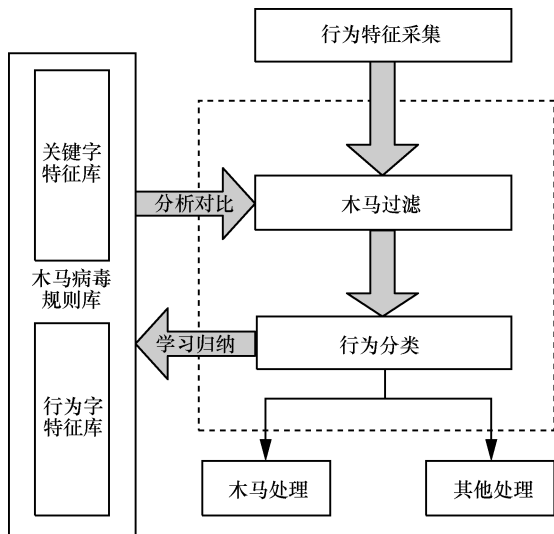


图 3 木马检测模型

$$threshold = \sqrt{\sum_{i=1}^{|D|} (D.R_i.weight - T.weight)^2} \quad (1)$$

式(2)为余弦相似度(cosine similarity)判断函数^[6,9],通过该判断,审查目标进程R是否符合非常规(木马)特征行为。

$$\begin{aligned} cosine(R, T) &= \frac{\langle R \cdot T \rangle}{\|R\| \times \|T\|} \\ &= \frac{\sum_{i=1}^{|T|} R.weight \times weight_i}{\sqrt{\sum_{i=1}^{|T|} R.weight^2} \times \sqrt{\sum_{i=1}^{|T|} weight_i^2}} \quad (2) \end{aligned}$$

依据式(1),行为特征分析如算法1。

算法1 特征行为分析

Input R

for i = 1 to |D| // 遍历规则库; |D|表示规则库规则数目;

if $D_i.R.act = R.r.act$ or $D_i.R.key = R.r.key$ then // 匹配成功

{break;}

else {

if $(cosine(R, T) \leq threshold)$ // 是否与对比特征行为相似

then {

R 为新的木马规则,定性为木马;

$D = D \cup \{R\}$.

} else {

R 为正常动作;

}}

4 安全性

基于木马代码特征码及特征行为算法,利用时空聚类性,依据时间调用序列,进行行为的规则化和归一化量化处理。根据其特性进行相似度分析,可以在木马初次实施攻击阶段就捕捉到其行为特征,建立时间特征序列,不断进行评判。随着系统的多次运行,系统自学习功能会逐渐完善木马病毒规则库内容,提高模型自相似性、准确度、智能性及实时性。

但是,根据判不准定理^[11],由于木马病毒的特殊性以及某些应用功能的相似性,对木马病毒的判别无法达到百分之百准确。木马病毒检测问题无法在多项式时间内得

到满足。只有通过不断对测试分析模型进行训练来提高精确性。但是,可以通过判定算法,实现准确性比较高的监测预警。

通过对获取的木马样本程序及在虚拟机上运行测试,该方法可以有效学习和检测木马特征码及特征行为。尤其针对木马变种,当有效进行特征行为规则化后,可以大大提高检测的速度和准确度。

5 结束语

木马病毒具有隐藏性、欺骗性和巨大的破坏性。本文提出基于挖掘技术的相似度判别算法,从3个方面保证算法的有效性。首先,依据木马病毒特征码及特征行为,进行特征的规则化,建立初始木马病毒规则库。其次,建立进程行为特征捕捉及分析过程,通过对进程特征行为的量化,采用聚类分析方法,完成特征行为规则化。最后,通过对比规则库及相似度主动对比法,完成对可疑进程的分析对比,确定可疑进程性质。安全分析及实验检测表明该算法具有自主学习及主动防御特征,很好地平衡了静态测试技术和动态测试技术的优缺点。

参考文献

- 1 国家计算机病毒应急处理中心. 第十二次全国信息安全状况与计算机及移动终端病毒疫情调查分析报告. <http://www.antivirus-china.org.cn>, 2013
- 2 韩宇, 申永军. 特洛伊木马隐藏技术分析. 微计算机信息, 2012, 28(1):128~129
- 3 蔺聪, 黑霞丽. 木马的植入与隐藏技术分析. 信息安全与通信保密, 2008(7):53~55
- 4 张复生. 基于广谱通信行为分析的木马检测技术研究. 重庆邮电大学硕士学位论文, 2011
- 5 Liu D. Analysis on hiding technologies of Trojan hors. China Science and Technology Information, 2010, 1(1):112~113
- 6 付文, 魏博, 赵荣彩. 基于模糊推理的程序恶意性分析模型研究. 通信学报, 2010, 31(1): 44~50
- 7 李家骥. 计算机木马攻击与检测技术研究及实现. 电子科技大学硕士学位论文, 2011
- 8 李鹏, 王汝传. 基于自相似特性的恶意代码动态分析技术. 南京邮电大学学报(自然科学版), 2012, 32(3):86~90
- 9 邢云冬, 刘胜利. 木马网络通信特征提取模型的设计与实现. 计算机工程与设计, 2010, 31(20):4382~4384, 4446
- 10 程文聪. 面向大规模网络安全态势分析的时序数据挖掘关键技术研究. 国防科学技术大学博士学位论文, 2010
- 11 何鸿君, 罗莉, 董黎明等. 广义病毒的形式化定义及识别算法. 计算机学报, 2010, 33(3):562~568

(下转第 115 页)