基于行为分析的木马检测系统

颜会娟　秦杰

河南工业大学信息科学与工程学院　河南　450001

摘要：本文通过对木马及木马检测技术的研究，提出了基于行为分析的木马检测技术。主要对木马的行为特征进行抽象描述，首先根据一定的规则建立一个行为特征数据库，并结合启发式分析器来进一步分析判断被检测的程序是否是木马，同时做相应的处理。实验表明，与传统的木马检测技术相比，该算法准确率高，实时性强，占用系统资源少。

关键词：木马；行为特征；系统调用；行为分析

0　前言

本文提出基于行为分析的木马检测技术，该系统的模型如图 1 所示。该模型设计符合标准化的入侵检测体系结构规范(CIDF)结构。该系统包含行为对象监控模块，行为特征数据库模块，行为启发式分析模块，系统响应模块。其中行为对象监控模块是对系统中程序有可能操作的对象进行实时监控、记录并将结果交给行为启发式分析模块进行处理；行为特征数据库是用来记录木马行为特征的，并且对木马的每种行为根据一定的规则设定一个权值，以便在行为分析阶段判断正在运行的程序是否是木马程序；行为启发式分析模块主要是结合行为特征数据库对行为监控模块传来的信息进行分析，进一步做出判断；系统响应模块是对行为分析模块所做的判断给予相应的处理。

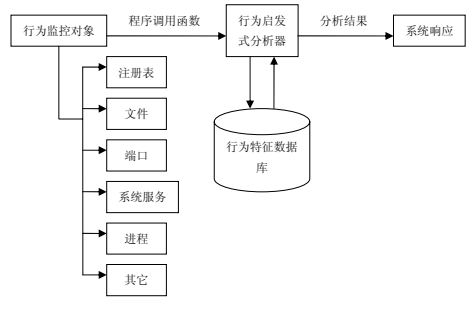

图 1　系统模型

1　模块设计

1.1　行为对象监控模块

对木马的行为进行分析之前首先把行为有可能操作的对象进行分类总结，即木马程序可能对哪些对象做操作。本文中列出了几种比较常见的行为对象，如注册表、文件、端口、进程、系统服务等。对木马这些行为对象有可能做的操作进行抽象描述(如表 1 所示)。

表 1　木马行为特征描述

行为对象	木马行为特征描述
注册表	修改注册表的自启动项、关联项及其他一些具有自启动功能的项
文件	修改系统文件，如 wini.ini，system.ini 等；捆绑自启动文件；拷贝 Autorun.inf 文件；拷贝文件到系统目录；拷贝文件给自启动项；新增、删除文件等行为
端口	打开、关闭端口
进程	创建进程，远程线程注入，隐藏进程，关闭安全软件，调用 CMD 运行可执行程序等
系统服务	修改自己的路径，把自己注册为系统服务进行运行
其它	记录键盘操作、屏幕截取操作

在本策略中对这些行为对象的监控使用 Windows API HOOK 技术。下面对 Windows API HOOK 在本策略中的应用做简单的介绍。

API HOOK 是指应用程序在调用真正的系统 API 前先被截获，进行一些处理再调用真正的 API 函数。从本质上讲，API HOOK 技术改变了一个正常程序的执行路径，它可以拦截或者监听可执行代码在执行过程中的一些信息，可以更改操作系统的行为，可以帮助设计者了解系统内部结构和运作机制等，被广泛应用于时间追踪和修改系统行为等操作中。

由于木马是由一段段的程序代码组成的，要在计算机里运行，都要调用系统中不同的 API 函数。所以可以利用 Windows API HOOK 技术监控、拦截这些木马程序的 API。

本文由国家自然科学基金项目(No.60373003)与河南工业大学校基金项目(No.2006BS009)资助。
作者简介：颜会娟(1984-)，女，硕士研究生，研究方向：网络安全。秦杰(1968-)，男，副教授，硕士生导师，研究方向：Web 数据库、信息检索、网络安全。

下面以监控注册表和文件操作为例来分析、拦截木马程序。

监控注册表：要实现注册表监控，首先要在系统服务调度表中找到要替换的函数指针，将这个指针指向我们自己定义的函数，当调用相应的函数指针时，首先就会执行我们自定义的函数，这样就实现了系统服务的挂接。以监控木马写注册表入自启动项为例。要对注册表进行修改，使用的是 RegCreateKey 函数。它的函数原型是：

Private Declare Function RegCreateKey Lib"advapi32.dll" Alias"RegCreateKey"(Byralhkey As long, ByvalpSubkey As String, phkResult As long) As long

该函数有三个参数：第一个参数是该键的根键的预定义值；第二个参数是主键名，包括其路径；第三个参数是长整型的指针，如果函数返回 0(成功)，它将保存该键的句柄。当 RegCreateKey 调用 Ntdll.dll 这个动态链接库中 *NtCreateKey* 函数，而此时 *NtCreateKey* 在设置好参数以后，就会通过软中断进入到内核态，请求相应的服务。这时因为在 Ntdll.dll 中到处以 *Zw* 开头的系统服务占位函数在内核中都有一个相应的以 *Zw* 开头函数与之对应，所以系统就会根据中断服务号在系统服务调度表中查找到 *ZwCreateKey* 函数，然后执行调用这个函数。那么，我们在利用钩子函数时，将系统服务调度表中的函数指针指向我们自定义的函数 *MyZwCreateKey*，此时，执行系统服务调用时就会先执行函数 *MyZwCreateKey*，在这个函数中进行我们所需要的预处理，然后再根据具体情况，执行原来的 *ZwCreateKey* 函数，或者跳过这个函数，直接执行后面的代码。

监控文件操作：这里以监视系统中对文件的拷贝操作。使用的是 *CopyFile* 函数，它的函数原型是：

Declare Function CopyFile Lib "kernel32" Alias "CopyFileA" (ByVal lpExistingFileName As String, ByVal lpNewFileName As String, ByVal bFailIfExists As Long) As Long

其中该函数有三个参数：第一参数 lpExistingFileName 代表源文件名；第二个参数 *lpNewFileName* ，代表目标文件名；第三个参数 *bFailIfExists* 是长整型的指针，如果设为 TRUE(非零)，那么一旦目标文件已经存在，则函数调用会失败，否则目标文件被改写。同样，要监控这种类型的文件操作，系统首先定义一个关于文件复制的挂钩函数 *myCopyFile*，我们在利用钩子函数时，将系统服务调度表中的函数指针指向我们自定义的函数 *myCopyFile*，此时，执行系统服务调用时就会先执行函数 *myCopyFile*，在这个函数中进行我们所需要的预处理，然后再根据具体情况，执行原来的 *CopyFile* 函数，或者跳过这个函数，直接执行后面的代码。

同样，对其他的操作类似上述的情况，我们可以分别进行设计。

1.2 行为特征数据库

行为特征数据库主要集合了木马各种行为的特征抽象，要完整的描述各种木马行为，并且方便程序设计者调用分析，在本策略中所建立的行为特征库由一组特征向量 $N_A=\{A_1,A_2,\cdots,A_n\}$，其中 A_n 代表一个表示一个行为的特征向量。对于每个行为的特征向量可以描述为如下几个部分：

（1）行为特征抽象描述 P_i，以修改注册表自启动项为例，我们可以描述为“ P_i =修改注册表自启动项”。

（2）调用函数 C_i：由于木马上述行为在木马程序上一般表现为不同的 API 函数的调用，所以在对木马行为进行抽象描述时可描述为： C_i =行为对应的 API 名称(即函数名称)。同样以修改注册表自启动项为例，因为木马在程序在进行修改注册表时要调用 RegCreateKey 函数，所以，可以把该行为调用 API 描述为相应的函数“RegCreateKey”。

（3）在本策略中用到了启发式分析器分析程序行为，主要是对木马的各种行为进行分析、提炼，在这里对每种木马行为进行反复比较分析，根据一定的规则给它们分别设定一个权值 B，它们的权值体现了相应木马行为在木马行为特征中的重要程度。

在这里，以修改注册表自启动项为例，它在行为特征库中的描述(P_i , C_i , B)可以描述为：(修改注册表自启动项，RegCreateKey, B)，对其它的行为也可做类似的总结。

1.3 行为启发式分析模块

该模块对行为对象监控模块传来的数据进行分析，具体实现如下：首先对从行为对象监控模块传来的数据结合行为特征数据库中的对应行为的权值，计算各可疑行为的加权值之和，然后用这个加权值之和跟事先预定的阀值进行比较，如果权值之和大于阀值，就可确定此行为是木马行为。

本策略根据各个可疑指标的可疑程度为每一个行为制定权值，并且把某个程序在此次运行中表现的不同行为的权值进行相加，将得到的值与事先规定的阀值进行比较，如果权值之和比临界值大，就确定此程序为木马程序。具体的计算公式如下：

$$M=a_1*b_1+a_2+b_2+\cdots\cdots+*a_n*b_n$$

其中 a_i 为自重权数，表示每个行为特征的权值，并且满足 $a_1+a_2+\cdots\cdots+a_n=1$，规定 a $\in[0,1]$，a_i 的取值体现了相应行为特征在木马各行为特征中的重要程度。n 代表某个程序在此次运行中表现的可疑行为的个数；M 为此次运行的程序为木马程序的可疑程度值。

加权系数的确定原则是使预测误差的均方值最小，在这里利用过去 k 个取样值来进行预测，也可以称为 k 阶线性预测。

设 $y(n)$ 的预测值用 $\hat{y}(n)$ 表示，则有 $\hat{y}(n)=-\sum_{i-1}^{k}a_i * y(n-i)$, $n \in N$

式子中，N 为正整数；$-a_i$ 表示加权系数，称为预测系数。

预测误差为 $e(n)$：$e(n)=y(n)-\hat{y}(n)=\sum_{i-1}^{k}a_i * y(n-i)$, $a_0=1, n \in N$

预测误差可以按均方规则确定：$\xi_{\min}=E[e^2(n)]$

1.4 行为响应模块

对启发式分析器分析的行为，系统就可以做出以下响应：如果确定此行为是木马行为时就直接调用查杀模块进行删除；如果不能确定是木马行为时，就弹出警告窗口，询问用户是否允许、禁止或删除此程序；如果判断是正常行为，系统就允许此程序继续运行。

2 实验结果及分析

为了验证本系统的性能，本实验通过在 Windows XP 下的 C++语言实现编程。实验数据来源于美国国防高级计划研究署的入侵检测测试数据集 DARPA。为了方便实验，本系统从中随机选取了 4 组数据进行实验，其中第 1 组包括 10000 条数据，其中包含 100 条异常数据；第 2 组包括 20000 条数据，其中包含 200 条异常数据；第 3 组包括 30000 条数据，其中包含 300 条异常数据，第 4 组包括 40000 条数据，其中包含 400 条异常数据。经过实验结果显示，第一组报警数为 78，第二组报警数为 169，第三组报警数为 271，第四组报警数为 374。通过实验分析可以得出本系统检测木马的检测率、误报率和准确率，如图 2 所示。

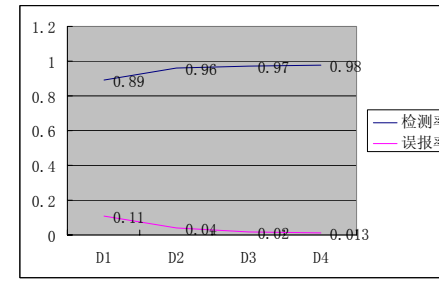

图 2 实验结果

从图 2 中可以看出，本系统在检测木马病毒时具有很高的检测率，同时误报率也较低。

3 结束语

本文从行为分析的角度提出了基于行为分析的木马检测技术，对木马的行为特征进行了透彻分析，并结合启发式分析对被测试程序进行分析判断，使其能够有效识别已知木马和未知木马的入侵。

参考文献

[1]金山:2008 年中国互联网安全报告.

[2]LYMAN J In search of the world's costliest computer virus 2008.

[3]HUGHES L.DELONE G Viruses,worms,and Trbjan horses: Serious crimes,nuisance.or both 2007.

[4]尹清波,张汝波,李雪耀等.基于线性预测与马尔可夫模型的入侵检测技术研究[J].计算机学报.2005.

[5]HUGHES L.DELONE G Viruses,worms,and Trbjan horses: Serious crimes,nuisance.or both 2007.

[6]陈友,沈华伟等.一种高效的面向轻量级入侵检测系统的特征选择算法[J].计算机学报.2007.

[7]顾雨捷.用于行为分析的反木马的模糊分类算法研究[D].学位论文.2008.

The Trojan horse Detection System Based on Behavioral Analysis

Yan Huijuan, Qin Jie

College of Information Science and Engineering,Henan University of Technology, Henan,450001,China

Abstract: Through studying techniques of the Trojan and anti-Trojan, this paper presents the Trojan-detection technology based on behavioral analysis. Through the abstract description of the Trojan's behavior, according to certain rules to establish a behavior feature database, and combining the heuristic analyzer to further analysis and judge whether the program is the Trojans, then do the appropriate processing at last. Comparing with the traditional technology of the Trojan horse detection, the experiments show this algorithm has high accuracy rating, and is effective and efficient in real time; what's more, it takes up little system resources.

Keywords: Trojan;behavioral features;system call;behavioral analysis