学校编码：10384　　分类号_____密级_____
学号：X2010230638　　UDC_____

厦门大学

工　程　硕　士　学　位　论　文

基于沙箱技术的恶意代码分析系统的设计与实现

Design and Implementation of Malicious Code Analysis System Based on the Sandbox Technology

冉向阳

指导教师姓名：　吴　清　强　副教授
专　业　名　称：　软　件　工　程

论文提交日期：　2012　年　10　月
论文答辩日期：　2012　年　11　月
学位授予日期：　2012　年　　月

答辩委员会主席：_____
评　　阅　　人：_____

2012 年　月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为(　　)课题(组)的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。)

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　√　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

摘 要

随着信息技术的发展和互联网的普及，人们的工作生活越来越离不开网络，同时恶意代码对计算机系统和网络的威胁也日益严重，造成的危害越来越大。恶意代码已经不再是最初以出名为目的，现在更多的转向通过不正当手段来获取个人隐私、攫取非法利益，而且恶意代码也成为了国家和政治组织之间达到政治目的的一种手段。税务系统的信息化已经有了长足的发展，但也同样面临着病毒、蠕虫、木马等恶意代码的威胁。目前恶意代码的制作人员的知识越来越丰富，采用的技术越来越高超，非常难以防御，因此恶意代码的检测和分析成为信息安全领域中非常重要的防御手段。

本文首先介绍了恶意代码的发展历史，种类、采用的技术以及目前恶意代码的防御技术，并且介绍了恶意代码的分析技术：静态分析和动态分析技术，通过对两种方法的分析对比，认为动态分析技术是今后恶意代码分析的主要发展方向。沙箱技术即虚拟机技术，在虚拟机中运行可执行程序不会对实体主机造成影响并且可以回滚到初始状态，所以沙箱技术可以应用到恶意代码的行为分析当中。本文采用沙箱技术构建了一个恶意代码分析系统，其中采用了内核 Hook 技术、虚拟机技术、驱动编程、内核调制技术等。实现了系统监控、虚拟机控制、进程的获取和监控、文件、注册表、驱动和网络监控以及日志分析模块。对恶意程序的各种行为进行监控，并将监控结果进行数据过滤和分析以给出可视化的分析结果，实现其监控和分析过程的自动化。

经过测试，本系统可以对恶意代码的各种恶意行为能够得出较为准确的分析，能够为研究恶意代码的行为特征提供特征库，为恶意代码的研究与分析、检测与防御提供有力支持。

关键词：恶意代码；沙箱；行为分析

Abstract

With the development of information technology and the spread of the Internet, the network becomes more and more inseparable from people's work and life, but meanwhile the threats of malicious code on computer systems and networks are also becoming increasingly serious, causing greater and greater harm. Malicious code no longer sticks to its initial purpose of becoming famous, but is utilized in improper means to obtain personal privacy and obtain illegal profits; moreover, the malicious code has also become a means to achieve political objectives for nations and political organizations. The informatization of the tax system has been long developed, but it still faces the threat of malicious codes like viruses, worms and Trojan. At present, the producers for malicious codes are getting richer and richer knowledge and applying higher and higher technology, which is very difficult to defense; therefore, the detection and analysis of malicious codes has become a very important defense means in the field of information security.

This dissertation first introduces the development history and types of malicious codes, as well as the applied technologies and the current defense technologies for malicious codes and further describes the analysis technologies of malicious codes: static analysis and dynamic analysis technologies. By analysis and comparison of the two methods, the author puts forward that the dynamic analysis technology is the future development direction of the malicious code analysis. Sandbox technology also known as virtual machine technology will not affect the physical host machine when operating the executable program in the virtual machine and can make it roll back to the initial state. Therefore, the sandbox technology can be applied in the behavior analysis of the malicious codes. This dissertation adopts the sandbox technology to build a malicious code analysis system, which applies kernel Hook technology, virtual machine technology, driven programming technology, kernel modulation technology and others.

Thus system monitoring and control of virtual machine, access and monitoring of process, monitoring of files, registry, driver and network as well as log analysis module are realized. Monitor for the various behaviors of the malicious program and the visual analysis results can be obtained by conducting data filtering and analysis for the monitoring results to achieve the automation of monitoring and analysis process.

After testing, this system can accurately analyze the behaviors of various malicious codes. Hence it can provide feature library for the study of the behavioral characteristics of the malicious codes and strong support for the research and analysis, detection as well as the defense for malicious codes.

Key words: Malicious Code; Sandbox; Behavioral Analysis

目 录

Contents

第一章 绪论

1.1 研究背景

随着计算机、互联网技术的发展，人们的工作、生活越来越离不开信息技术带来的生产、生活方式的转变，工作中需要计算机来处理各种文档、各种业务、收发电子邮件等。生活中需要计算机来上网冲浪，获取各种信息、发微博、观看视频、网络游戏等。互联网技术的发展改变了企业与消费者相互沟通及交互的方式，已成为信息共享与商务交易的主要平台。同时，互联网变得日益复杂、没有界限。2011 年 7 月 19 日，中国互联网络信息中心（CNNIC）在京发布了《第 28 次中国互联网络发展状况统计报告》显示：截至 2011 年 6 月，中国网民规模达到 4.85 亿，较 2010 年底增加 2770 万人；互联网普及率攀升至 36.2%，较 2010 年提高 1.9 个百分点[1]。

自税务系统实行金税工程以来，通过“金税工程”一期、二期的建设，国税系统的网络建设已经覆盖了全国区县（含）以上国税机关，形成了总局、省局、地市局、区县局的四级广域网，成为国税系统的网络通信支撑平台。在进行网络建设的同时，税务系统在各种硬件配备上也有了一定规模：拥有小型机 1000 多台，其中国税约 800 台，地税约 200 台；PC 服务器 15000 多台，其中国税约 10000 台，地税约 5000 台；PC 机 25 万台，其中国税 16 万台，地税 9 万台；已经实现计算机化管理的基层征收单位 2.2 万多个，其中国税约 1.2 万个，地税约 1 万个；通过计算机管理的纳税户超过 1000 万，80%以上的税款通过计算机征收。另外，在税务系统信息化建设过程中形成了 3 万人左右的信息技术队伍，成为整个税务系统信息化建设的中坚力量[2]。

高速发展的信息技术给人们的工作生活带来巨大地便利的同时时，也受到病毒、木马、黑客等种种安全威胁和影响。在这其中恶意代码是当今互联网的主要安全威胁。在信息安全事件中，绝大部分用户的经济损失是由恶意代码造成的。当前，各种计算机病毒、特洛伊木马、蠕虫(Worm)、逻辑炸弹以及间谍软件(Spyware)等恶意软件呈广泛蔓延趋势，其主要危害包括窃取用户敏感信息、发送垃圾邮件、控制受害主机发动 DDoS 攻击等。目前恶意代码并不仅仅

局限于国内，出现了许多跨国攻击情况。国家互联网应急中心(CNCERT)发布的《2011 年中国互联网网络安全态势报告》指出，2011 年，我国遭受境外网络攻击持续增多，境外有近 4.7 万个 IP 地址作为木马或僵尸网络控制服务器，控制我国境内近 890 万台主机，比 2010 年控制主机数增长近 1 倍，其中美国以 9500 多个 IP 地址控制我国境内近 885 万台主机，高居榜首[3]。并且，恶意代码经过多年的发展演化后，其数量也变得非常庞大,它们的破坏力和感染力较初期阶段得到了显著的增强。赛门铁克的报告指出，虽然漏洞数量有所减少，但恶意网络攻击的数量仍在继续增长。赛门铁克在 2011 年一共拦截了 55 亿次恶意攻击，较 2010 年增长了 81%；每天拦截的网络攻击数量增加了 36%，恶意代码数量增加到 4.03 亿个[4]。恶意代码造成的信息安全问题呈爆发状态，它不仅对用户的数据安全存在着非常严重地威胁，而且让各个企业以及国家都遭受了经济上的巨大损失。

江民科技 2011 年上半年网络安全信息报告显示[5]，2011 年上半年，最为活跃的病毒类型仍旧为木马病毒，其共占据所有病毒数量中 60%的比例，如图 1-1 所示。其次，分别为蠕虫病毒和后门病毒。这三种类型的病毒共占据所有病毒数量中 83%的比例，可见目前网民面临的首要威胁仍旧来自于这三种传统的病毒类型。在木马病毒中，盗号木马较去年有明显的减少，取而代之的则是以恶意推广为主要目的的木马和脚本型木马病毒。

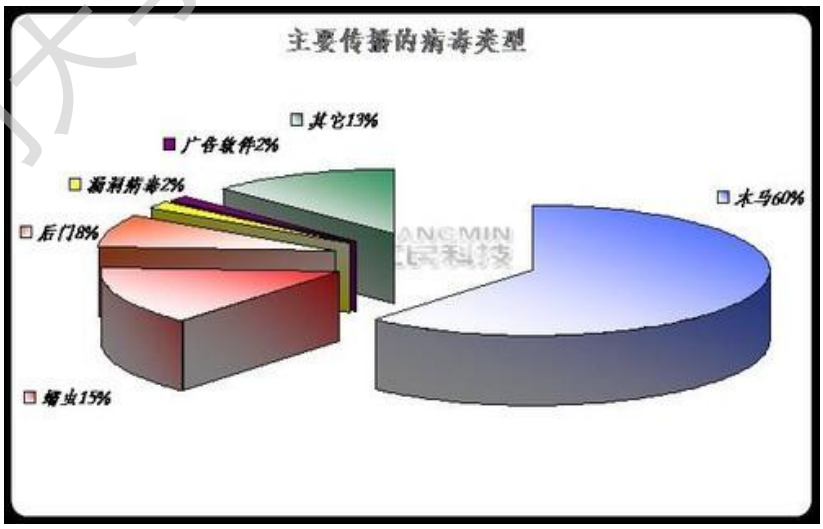

图 1-1 江民科技 2011 年上半年信息安全报告主要传播病毒类型

根据河北省国家税务局安全防护系统数据统计，计算机病毒仍是全省国税

系统安全所面临的最大威胁。2011 年全省国税系统计算机感染病毒 160 万次，较 2010 年增加了 57 万次，病毒种数 5272 种，较 2010 年减少了 1106 种，感染计算机台数 15544 台，较 2010 年增加了 9166 台。病毒的主要形式是蠕虫和木马，占总数的 90.19%。

经研究发现，基于病毒库的防御检测方法已经无法赶上网络上由原本单一的病毒或木马经过加壳、加花、组装等手段后成为“变种”木马数量的增长速度，其增长速度非常快，数量巨大；并且无法抵抗新近出现的未知病毒。因此通过检测恶意代码行为上的特征和惯用手段来建立高效的防御手段已经成为当务之急。

因此，通过对恶意代码的行为研究和监控，不仅可以有针对性地发现当前木马的行为特征，而且还能及时发现新的攻击技术与形式，这些都对恶意代码检核以及防御具有重要意义。

1.2 主要研究工作

本文设计并实现一个基于沙箱技术的恶意程序行为分析系统 Taxbox，对恶意程序进行行为的自动分析。具体工作如下：

1.在虚拟机层设计和实现了监控控制模块和网络流量捕获模块。

2.在主机层设计和实现了基于 VIX 的 VMware 控制模块和数据分析模块，并实现了系统的可视化。

3.采用 SSDT Hook 技术对进程、文件和驱动加载行为进行监控，发现进程的创建、终止行为，线程的创建行为，文件的创建、覆盖、删除行为以及驱动加载行为等。

4.采用注册表回调函数对注册表行为进行监控，发现恶意代码在执行过程中新建注册表键、删除注册表键、设置注册表值、删除注册表值等行为。从而实现对恶意程序的进程、注册表、文件和网络等行为的综合分析。本文设计开发的恶意程序行为自动分析系统能够全面、自动地批量分析恶意程序的行为，使用方便，效率高。

1.3 论文组织结构

本文第一章为绪论，给出了本文的项目背景，阐述了恶意代码的危害和研究恶意代码的重要性。介绍了本文的主要工作，给出了本文的组织结构。

第二章介绍了恶意代码概述及其相关原理、关键技术、恶意代码的分析技术，并且对沙箱技术的原理、实现方法以及系统采用的关键技术等进行了阐述。

第三章提出 TaxBox 系统的需求分析，介绍了各个功能模块的需求。阐述了系统的整体架构，从概念和层次上阐述了系统的结构，给出了系统的工作流程。

第四章对系统的关键功能进行了具体的设计。

第五章内容为各个功能模块的具体实现。

第六章 TaxBox 测试与验证，介绍了测试环境，对恶意代码的测试给出了结果并对其恶意行为做出了结论。

第七章为总结与展望。总结了文本的工作和不足，展望了后续的工作。