基于病毒行为序列的未知病毒分析技术研究

郑 重，王志英，陈项颢，黄 詠，
（国防科学技术大学 计算机学院，湖南 长沙 410073）

摘 要：提出了一种在虚拟执行技术支持下基于病毒行为序列的未知病毒分析检测技术。该技术可以克服病毒特征代码扫描法不能识别未知病毒的特点。在模拟的虚拟执行环境中对该方法进行了测试，测试表明了该方法的可行性和较高的准确性。

关键词：计算机病毒；虚拟执行；行为序列；病毒检测

Unknown Computer Virus Detection Based on Its Behavior Sequence

ZHENG Zhong, WANG Zhi-ying, CHEN Xu-hao, HUANG He
(School of Computer Science, National University of Defense Technology, Changsha, Hunan 410073, China)

Abstract: A algorithm to detect and analyze computer virus based on its behavior sequence under the support of virtual execution technology is presented in this paper. It can overcome the shortage of normal virus scanner, which could not detect unknown virus. Tests in a simulating virtual execution environment indicated that this algorithm is feasible and accurate.

Key words: Computer virus; Virtual execution; Behavior sequence; Virus detection

1 前言

目前的杀毒软件主要依靠“特征码”的方式查杀病毒，这种方法查毒准确、快速，但其要求杀毒软件厂商必须先截获到病毒样本，进行分析后，才能对其进行查杀，而对于一些新病毒以及病毒的新增变种则显得力不从心。并且病毒伪装技术日新月异，如代码的加壳技术等给依靠特征码来进行病毒查杀的方式带来了不小的困难。

主动防御技术作为杀毒软件的辅助功能模块，主要是分析病毒的单个恶意行为，进行资源访问控制，限制未知进程的活动。发现疑似恶意行为时，提示用户，让用户进行判断，这样就把判断病毒行为的任务交给了用户。而普通用户不具有判断病毒的能力，可能存在误操作，给病毒提供了破坏计算机的机会。

而虚拟机技术，可以用软件先虚拟一套运行环境，让病毒先在该虚拟环境下运行，从而观察病毒的执行过程。病毒在虚拟环境下，所有的破坏行为暴露无疑，并且不会破坏真实系统，从而为研究病毒行为提供了一种新的方式。

本文基于虚拟执行的技术，结合对主动防御技术和病毒行为的研究，设计了一种基于病毒行为序列的未知病毒的检测方法。该方法既可以实现对已知病毒的识别，也可以对可疑文件进行分析评判，最终实现对未知病毒的识别。最后，在模拟环境中，对该方法进行了实验测试，达到了良好的实验效果。

2 基于行为序列的未知病毒分析技术原理简介

基于行为序列的未知病毒分析技术可以这样简述：采用软件建立一个病毒执行的虚拟环境，在此环境中运行某病毒，激发该病毒的破坏行为，并根据已有的数据库对病毒行为进行选择性跟踪记录，即这种分析算法是基于病毒序列而非传统的单个恶意行为。每当增加一条行为记录时，将现有的行为序列记录同已建立的行为库进行分析对比，并采用分析算法对该病毒进行判断，从而达到对未知病毒的识别。

3 基于行为序列的未知病毒分析技术

3.1 病毒行为介绍

病毒的代码可以千变万化，并且还可以想方设法地伪装自己的代码，隐藏代码静态特征信息。而病毒的破坏行为却是如出一辙，如设法逃避杀毒软件的侦测，无声无息地运行，悄无声息地盗走用户的信息等等。

3.2 行为序列

一般的主动防御软件正是基于以上总结的病毒行为的特征，在某程序出现了某一疑似恶意行为时，首先进行拦截，然后弹出对话框提示用户操作。由于这些病毒的某些行为，一些正常运行的软件由于运行的需要也会产生，这给未知病毒的检测带来很大的困难。

通过对诸多病毒的分析，我们发现一些行为的序列是病毒特有的，基于单个行为无法判断某个程序的特征，而基于行为的序列却可以将病毒和普通程序明显地分开，即这种识别方法是基于病毒行为的上下文关联分析。

单个行为是指进程出现表 1 中的某具体行为，如对注册表某特定项的修改等。而程序这样的行为很多，如何提取关键的行为是很重要的一个环节。在分析过程中，行为的提取也是基于已有的病毒行为库，只有在行为库中的行为才会被监控和记录。

病毒行为库可建立为一个图（如图 1 所示），图中的每个节点表示在行为库中收集的需要监控和记录的病毒行为。

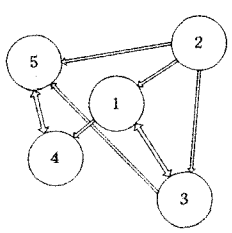

图 1 病毒行为库组织图示意图

图 1 中节点代表某个病毒行为，两点的有向箭头表示这两个行为的关联，如在检测到行为 3 后，若检测到与之有连线的行为 1 或 5，则将该行为加入到追踪序列中。

同时，这些病毒行为库中的节点，根据关联关系组成了不同的病毒行为序列集（见图2），并且这些序列互不包含，所有这些序列组成集合 T_0，即病毒行为序列库。当某程序出现序列集中的行为序列时，则可判断该程序为病毒程序，并可根据历史经验总结给出该病毒的危害方式，以及危害的后果等等。

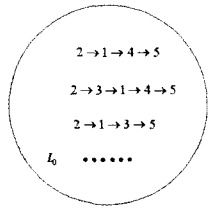

图2 病毒行为序列集示意图

3.3 病毒识别

根据已建立的病毒行为库，当某程序出现库中的某一行为时，则对该进程的行为进行追踪，并根据已有的行为序列规则，在确认病毒之后，直接终止该进程的运行，并实施相关的操作。追踪识别程序行为的算法如下：

(i) $T=\Phi$（Φ 代表空序列），G 代表已建立的行为库；

(ii) 对于某程序的行为 V_i，若顶点 $V_i \in G$，则

当 $T=\Phi$ 时，$T=V_i$；

否则，若存在 $V_x \in G$，使有向边 $(V_x,V_i) \in G$，则 $T=T \rightarrow V_i$；

(iii) 当集合 T 增大时，则检查是否有 $T \in L_0$（即当前行为序列为已确认的危险行为），若有，则立即终止该进程的活动并报警；否则重复(ii),(iii)直至该进程结束。

具体实现可描述为：当某一程序启动时，开始监控其行为，若该行为属于行为库，则将该行为添加到一个序列 T 中，在检测到与集合 T 中的行为有关联的行为时，将该行为也添加到 T 中，直至序列包含在 L_0 中时，便可以确认该程序为病毒程序。注意，可能同时为某一程序建立几个行为追踪序列 T_i。若在程序结束时仍然没有确认其为病毒，则释放为该程序建立的追踪集合 T_i。

3.4 行为库的升级

行为特征库，可在新型病毒研究过程中进行添加，实施行为库的升级，也可由有经验的用户在使用过程中，对于不能准确预测的病毒行为，进行手动添加，由系统分析，并加入到行为库中。

4 模拟验证

4.1 模拟环境

实验中在 VMware 6.0 虚拟机中装 Windows XP 操作系统，并装入文件、进程、注册表、网络监控软件，在运行病毒后，对监控软件的日志进行分析，得出分析结论。

由于在这样的虚拟机中运行病毒会对虚拟机中的操作系统造成破坏，为保持系统在运行某病毒后恢复正常，实验中使用了 VMware 的快照功能，在实验后及时通过快照进行恢复。

4.2 实验结果及分析

实验中，通过对目前较为流行的 245 个病毒的行为进行分析，并结合网络上对病毒行为的特征总结，建立了病毒行为库。并预留了 45 个病毒作为测试集。

在虚拟环境中对测试集进行逐个测试，分析留下的日志痕迹，图3描述了识别每个病毒的行为序列长度。

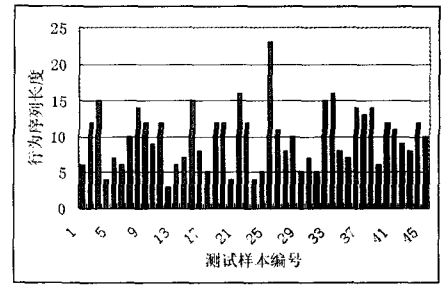

图3 测试样本被判定为病毒时的行为序列长度

一般病毒的行为有上千个，而能判断其为病毒的关键行为却只有 20 个左右。在这些繁琐的行为中筛选出对于判断病毒有用的行为算法显得尤为重要，在试验中，采用了与行为库中的行为进行逐个对比的方法。实验结果表明，这种对比的筛选方法是有效的，同时实验使用基于病毒行为序列的未知病毒检测算法将 45 个病毒均检测出，表明该方法的可行性与较高的检测准确性。

5 结论

基于虚拟执行和病毒行为序列的病毒检测方法可以克服一般的特征码扫描方法的弊端，但是这种方法对系统资源要求更高，执行速度远不及特征码扫描技术。所以基于病毒特征码比对检测的杀毒技术仍是反病毒软件采取的主流技术，虚拟机技术在相当长的时期内仍将是病毒检测的辅助手段。在未来虚拟执行技术更成熟的时候，本文研究的方法将得到更广泛的应用。

参考文献：

[1] 张波云，殷建平等．基于多重朴素贝叶斯算法的未知病毒检测[J]．计算机工程，2006，32(10)：18~21.

[2] C Nachenberg. Computer virus—coevolution. Communications of the ACM, 1997/Vol.40. NO.1.

[3]O Henchiri, N Japkowicz. A feature selection and evaluation scheme for computer virus detection, IEEE International Conference on Data Mining (ICDM),2006.

[4] M Christodorescu, S Jha. Static analysis of executables to detect malicious patterns. Proceedings of the 12th conference on USENIX Security Symposium — Volume 12 , 2003.

[5]T Li, X Liu, H Li. An immune—based model for computer virus detection. Lecture Notes in Computer Science,Volume 3810/2005,p59—71 2005.

[6]Takeshi Okamoto, Yoshiteru Ishida. A Distributed Approach to Computer Virus Detection and Neutralization by Autonomous and Heterogeneous Agents. pp.328, The Fourth International Symposium on Autonomous Decentralized Systems, 1999.

[7] 解析虚拟机杀毒技术 ，http://litiejun.blog.51cto.com/134711/21875, 2009—2—1.

[8] 瑞星 2008 主动防御技术特点和分析详解 http://soft.ccw.com.cn/news/htm2007/20070815_302685.shtml, 2009—2—24.

[9] 揭秘“主动防御”技术，http://www.cnw.com.cn/weekly/htm2006/20060628_47161.shtml, 2009—2—24.

收稿日期：2009—11—03