基于分层权值的恶意程序仿真系统设计与实现

江 雪,朱永强
(电子科技大学 示范性软件学院,四川 成都 610054)

摘 要:动态启发式中对恶意程序的仿真主要通过沙盒技术来模拟实现。沙盒技术由于其本身仿真能力的局限性,其并不能完全准确地仿真真机环境与指令。针对沙盒的仿真能力的缺陷与不足,结合实际工程应用环境,设计了一套使用专用服务器搭配 VMware 虚拟机作为仿真环境的恶意程序仿真系统,并根据动态启发式的判断机制,提出了一种基于行为概率分层的权值赋值算法。通过实验,证明了该系统可以有效查杀各类新型与变种恶意程序以及权值赋值算法的有效性。

关键词:动态启发式;恶意程序;VMWare;行为仿真;概率分层赋值
中图分类号:TP309 **文献标识码**:A **文章编号**:1673-629X(2014)04-0143-04
doi:10.3969/j.issn.1673-629X.2014.04.036

Design and Implementation of One Malicious Program Simulation System Based on Stratification Weights

JIANG Xue,ZHU Yong-qiang
(School of Software,UESTC,Chengdu 610054,China)

Abstract:For simulating malicious programs,the sandbox is the regular tool which the dynamic heuristic often uses. Sandbox technology cannot completely and accurately simulate the real PC environments and instruction. Aiming at the limitations of sandbox's simulation capabilities,combined with the actual application environment,design a simulation system based on VMware virtual machine which is used in malicious programs simulating,and propose a weights assignment algorithm based on hierarchical behavior in probabilistic. The experiments prove that this system can effectively kill many kinds of new and variants of malicious programs,and the weights assignment algorithm is feasible.

Key words:dynamic heuristic;malicious programs;VMware;behavioral simulation;stratified probability assignment

0 引 言

目前在反病毒领域,基于特征码的传统病毒检测技术依然占据主要地位。该方法首先对恶意程序提取其二进制特征,再通过模式匹配[1]的方式,对这些特征进行扫描与定位,但随着黑客技术的盛行,世界范围内的病毒与木马总量大幅度增长,基于特征码的病毒检测技术需要先收集病毒样本,才能提取出病毒特征,而病毒数量的增长,使得对应的分析工作也随之膨胀,且特征码查杀技术的被动响应模式永远滞后于病毒发作,因此具有先天的滞后性。

启发式反病毒[2-5]即针对特征码查杀技术的滞后性,其不依赖于被动更新与响应的特征库,通过分析文件的静态或动态行为,来猜测文件的恶意性。启发式根据所考察的行为特征,又分为静态启发式与动态启发式,静态启发式主要考察文件的 PE 结构特征与反汇编特征,虽然其运行速度较快,但是静态启发式的误报问题往往比较严重,由于动态启发式是获取并分析了文件真正运行时的行为特征,在判断上更为准确。

目前广泛使用的动态启发式技术即沙箱,沙盒本质上,属于一种恶意程序仿真系统,但由于沙箱本身只能简单模拟部分 CPU 指令与运行环境,因此也较难获得较高的查杀率。文中设计了一种使用 VMWare 虚拟机仿真来实现对恶意程序进行仿真的方法,并使用基于行为概率分层的策略,对被仿真行为进行权值确定,以实现对可疑程序性质的分析与判断。

1 基于 VMware 的虚拟机仿真系统架构

上文已经提到,由于沙箱本身的局限性,系统使用

收稿日期:2013-06-13 **修回日期**:2013-09-26 **网络出版时间**:2014-01-28
基金项目:中央政府专项基金:科技部科技型中小企业创新基金(10C26215122841)
作者简介:江 雪(1983-),男,硕士研究生,研究方向为网络安全。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20140128.1152.048.html

真实的物理机,搭配物理机中并行运行的若干虚拟机,作为恶意行为仿真的环境,以获得更真实的文件动态行为特征。

基于 VMware 的恶意程序仿真系统的基本结构图如图 1 所示。

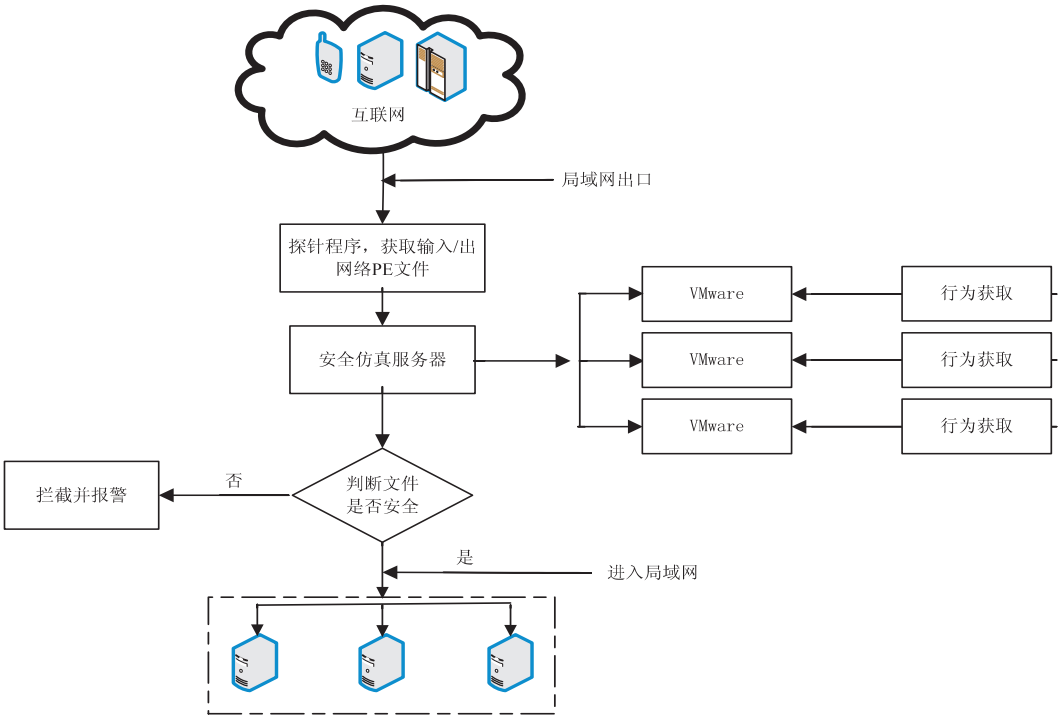

图 1 恶意程序仿真系统的基本结构图

可看到,该恶意程序仿真系统服务器部署在局域网网络出口与接入互联网之间,在服务器内部,运行着若干台 VMware 虚拟机,虚拟机内则为系统的行为获取常驻程序,用于获取被仿真程序的行为信息。进入局域网的可执行 PE 文件,会首先通过恶意程序仿真服务器,对其安全性进行仿真与评估,若文件安全,则允许进入局域网,否则,对此文件进行拦截并预警。

由于 VMware 使用的是基于 VMM 层的虚拟化技术[6],即虚拟计算机的物理资源,因此在虚拟化上层的操作系统,与真机所使用的操作系统基本相同,因此有着更准确的模拟与仿真性能,同时,由于 VMware 本身提供了方便快捷的快照及恢复功能,因此即使仿真操作系统被恶意程序破坏,也可以迅速地对其进行修复与还原。

2 恶意程序的常见关键行为模型

2.1 恶意程序感染计算机后的常见行为模型

一个恶意程序感染计算机后,往往会有几类最基本的行为模型,分别为:启动行为、文件行为、网络行为与隐藏行为,以下分别简单介绍:

(1)启动行为:启动行为是指恶意程序在植入目标计算机之后的二次启动,也就是达到驻留目标计算机,并自动其功能的目的。

(2)文件行为:文件是恶意程序的一个载体,是行为分析的主要依据。恶意程序从植入到任务完成清理痕迹的过程中,关于文件的操作有:文件创建、文件移动、文件重命名、文件删除、文件打开(盗取)等。

(3)网络行为:恶意程序在成功驻留系统,并顺利启动起来后,往往需要反弹回联,寻找控制端。其在主机上发生的与网络有关的行为包括:域名请求、域名回复处理、IP 连接、端口扫描、端口监听、发送邮件等。

(4)隐藏行为:恶意程序为了保证自己的生存,往往需要将自身的各类信息隐藏,具体包括进程隐藏、注册表隐藏、服务隐藏、端口隐藏、文件隐藏等。

2.2 恶意程序感染计算机后的部分具体可获取行为举例

下面列举了一些恶意程序常见的隐藏或者破坏的具体行为:

(1)进程类行为。

任务管理器可以显示出当前系统中所有的进程信息,而恶意程序往往需要将自身的进程信息隐藏,以下介绍一种基于 Hook 技术的进程隐藏方式。

Hook 技术可以监控系统中所有进程的 API 调用,任务管理器之所以能够显示出系统中所有的进程,是因为其调用了 EnumProcesses 等 API 函数。恶意程序通过 Hook 技术,可以对系统中所有程序的进程检测 API 的调用进行监控,在任务管理器(或其他调用了列举进程函数的程序)调用 EnumProcesses 函数时,木马

会提前获取此消息,因此便在函数将结果(列出所有进程)返回给程序前,将自身的进程信息从返回结果中抹去,以实现对自身信息的隐藏。

类似的,为了实现自身某种目的,木马还可以使用其他多种进程类恶意行为。

(2)内核入侵行为。

为了监听特定的进程信息,木马可以设置“钩子”的位置有很多,如 ntdll. dll 中的函数,但这并不保险,因为这些函数处于 ring3 级别,木马所设置的钩子可以被一个普通程序轻易绕过或消除。

因此,很多恶意程序选择在 SSDT/Shadow SSDT 下挂钩或者安装文件过滤驱动等,这些动作受到 ring0 保护,普通程序无法查知或修改。获得 ring0 的运行级别方法有很多,可以通过安装驱动(动态加载或者静态写注册表服务键),操作

[url = file://\\Device\PhysicalMemory]\Device\PhysicalMemory[/url]

以设置调用门,修改 win32k. sys 等。如果能将仿真程序置于 ring0 层,即可获取到恶意程序加载驱动或者试图进入 ring0 层的消息,进而发现可疑的恶意程序。

(3)工作机制行为。

一旦恶意代码通过传输到达了宿主计算机,会执行一个其相应的工作,工作可以采用许多形式。以下简单介绍几种标志性的恶意程序工作机制。

后门:这种类型的工作机制允许对计算机进行未经授权的访问。它可能提供完全访问权限,但也可能仅限于某些访问权限,例如,通过计算机上的端口 21 启用文件传输协议(FTP)访问。如果攻击可以启用 Telnet,黑客则可以将已感染计算机用作 Telnet 攻击在其他计算机上的临时区域。

信息窃取:一种恶意软件的工作机制旨在窃取信息。它可能会提供一种将信息传回恶意软件作恶者的机制。这种情况可以以多种形式发生,例如,传输可以自动进行,从而使恶意软件获取本地文件或信息;利用键盘侦听,获取用户所按的键(以便获取用户名和密码)等。

3　恶意程序仿真系统的基本工作原理与核心技术

该恶意程序仿真系统,在本质上属于启发式反病毒中的“动态启发式”,其原理与沙盒技术相同,但是仿真性能要远远优于传统的沙盒。系统的核心技术为:分析并获取恶意程序的关键性行为[7-11];使用虚拟机内常驻程序对运行可疑程序进行分析与行为提取;对不同恶意行为赋予合理权值,当某文件危险行为权值叠加超过某一阈值时,则对此文件预警。

3.1　恶意程序仿真的基本原理与监控机制

恶意程序可以轻易地对自身进行变形或免杀,来改变自身的特征值,但无论使用何种手段,恶意程序想要达到某种破坏目的,就必然要暴露出一些特定的行为,如:信息窃取类木马往往监听用户的键盘输入消息,以获取用户的各类账号与对应密码,并回联至某 IP 进行自动文件传输等。可以说,恶意程序或者木马想要实现某种功能,必定会有一些“关键步骤”,动态启发式,包括基于沙盒技术的与基于虚拟机仿真技术的,其本质,都是去监视这一类“关键步骤”。

根据对恶意程序主要行为的提取与分类,可建立对应的行为监控点,并编写对应的监视程序,以常驻形式运行在虚拟机中,对导入虚拟机内部的可疑文件进行仿真,该虚拟机仿真系统主要监控如下几个区域:

(1)进程类信息监控,用于监控木马的进程类行为,如 Hook API 等。

(2)注册表类行为,用于监控注册表中指定的敏感位置与对应键值。

(3)驱动与内核类行为,用于获取是否有可疑程序试图进入 ring0 级或安装驱动。

(4)信息窃取或监控行为,用于侦测是否有可疑的信息窃取动作,如监听键盘,打开 FTP 协议回传文件,固定频率抓屏(很有可能是远程监控动作)等。

由于 VMware 虚拟机本身可以通过快照轻松地实现快速还原,因此即使在虚拟机中运行真实的木马对当前操作系统造成了破坏,也可以快速方便的将虚拟机与内部挂载的仿真操作系统通过原始快照恢复到被破坏前的状态。

3.2　基于行为概率分层赋值的行为权值确定算法

由于虚拟机仿真本身属于“动态启发式”,本身是一种猜测性的行为,因此虽然优于传统的静态启发式,但虚拟机仿真还是会存在部分误报。

基于启发式的反病毒方式,往往会对其监控的行为赋予一个危险程度权值,一旦某程序的危险程度权值叠加超过了某一预设阈值,则对此文件进行警告,如果权值与阈值设定不合理,则会使启发式反病毒的准确性降低,如阈值设定过高,将会降低反病毒系统的报出率,而阈值过低,又会产生较多的误报,而每个行为设定的权值的合理性,将从根本上决定启发式的准确性。

在对权值赋值的时候,传统的做法是将在所有木马中出现的共有特征的权值取比较大的值,而其他的非共有特征则赋予较小的权值。所谓的共有特征,即为样本中所有木马共有的行为,而非共有特征即为只在部分木马中出现的行为,如在文献[12]中所述的“上面列出的几个木马,一共有 6 个共有特征,这样就

可以给其中的共有特征每个分配 0.1 的权值,其他 9 个特征共用 0.4 的权值”的方法。

由于近年来网络中恶意程序的数量与行为的快速增加,这种传统的方法已经很难适用,最明显的,在大样本恶意程序下,已经很难找到所有恶意程序的“共有特征”了,而各个行为在样本中的出现概率也出现了明显的分层。

据此情况,该系统在加权算法上,提出了一种基于行为概率分层赋值的算法[13],对行为权值进行赋值,具体算法如下:

设恶意行为的权值区间为[MIN,MAX](MAX 与 MIN 都为整数),其在样本中出现的概率最大值为 K_{max},最小值为 K_{min}。

(1) 确定每个权值区间的范围绝对值:

$$M = \left\lceil \frac{K_{max} - K_{min}}{MAX - MIN + 1} \right\rceil$$

(2)确定每个行为对应的恶意权值:

$$B_k = MIN + \left[\frac{L_k - K_{min}}{M} \right]$$

其中,L_k 为当前所考察行为 k 的样本发生率;B_k 为此行为的对应恶意权值。

依此算法,可依次确定各个启发式检测行为的恶意权值。

对于预警阈值,则可在行为权值确定后,通过对恶意程序样本与正常文件进行试验,选取最佳值,理论上,应找到报出率与误报率之间的合理平衡。

4 实验与分析

以下通过实验,证明该项目恶意代码检测系统的有效性以及特点。

实验样本共分为三批,分别为:

样本一:为实验前两个星期所收集恶意程序样本,共 507 个。

样本二:为实验前一个星期所收集恶意程序样本,共 639 个。

样本三:为实验本星期所收集恶意程序样本,共 573 个。

作为对比,该项目选择了某款基于特征值的开源反病毒软件,将其特征库更新至测试当天,同样分别对三批样本进行测试,得到实验测试结果如表 1 和图 2 所示。

通过实验数据,可以看到,基于特征值的反病毒软件,其查杀率随着样本的发现时间递减,即越是新近发现的木马样本,基于特征值的反病毒软件的报出率越低,但也并不是完全不能发现,这体现出了基于特征值查杀的“先收集,再分析,后查杀”的滞后性。

表 1 恶意程序查杀实验测试数据表

样本空间	样本一		样本二		样本三	
总数	507		639		573	
测试数据	报出数	报出率/%	报出数	报出率/%	报出数	报出率/%
该项目恶意程序仿真系统	462	91.1	561	87.8	493	86.0
某特征码查杀反病毒软件	431	85.0	354	55.4	216	37.7

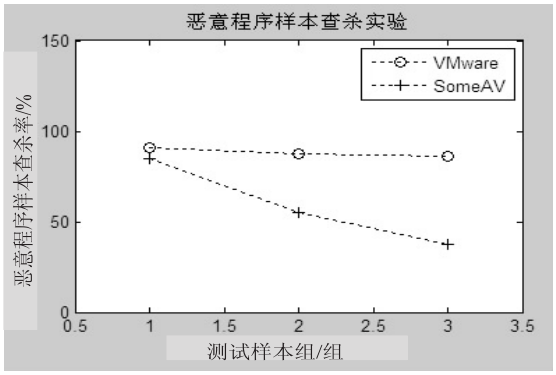

图 2 恶意程序查杀实验测试数据图

而基于 VMware 的恶意代码仿真程序,其查杀率基本上为一条直线,这是由于基于 VMware 的恶意代码仿真程序在本质上属于动态启发式,其查杀性能只取决于提取行为的准确性,而与时间因素无关。

此外,可以看到,几个样本空间下,基于 VMware 的恶意代码仿真程序对恶意程序的查杀性能,都要明显优于基于特征码的查杀。这是因为该系统在本质上属于动态启发式,其关注与监视的行为,往往都是恶意程序不能隐藏的动作与行为,对部分免杀恶意代码仿真的恶意程序进行分析,发现其中部分是使用了反虚拟机技术,导致恶意代码仿真无法对其进行监控,如何发现此类文件,将成为下一步工作的重点。

当然,相对于传统的特征值查杀,该系统要消耗更多的物理资源,其使用了服务器,来作为虚拟仿真环境的载体,相应地会提高系统的硬件成本。

5 结束语

文中所设计的恶意程序仿真系统,可以搭配相应的硬件与服务器,部署于各类机构与部门的网络出口,通过前端的探针系统,可以获取输入或输出所监控网络内的各类可执行文件,如邮件附件、QQ 传输文件等,并对此类文件进行仿真与判断,一旦发现恶意程序,则可进一步地进行监控、跟踪或拦截,从而阻止恶意程序侵入所保护的局域网,保证网络内部环境的安全。下一步可进行的工作是继续优化仿真行为库,同时提升前端的数据获取能力以及发现带有反虚拟机功能的恶意程序,以提高恶意行为仿真的覆盖面与准确性。

(下转第 150 页)

成为一种有效的机制,以满足网络节点信任。其优势在于:

(1)没有人可以使用 DAA 公共密钥去确定节点的身份,从而保证 Ad Hoc 节点是可靠的;

(2)仿真实验证明:由于签发的 DAA 与 TPM 绑定的为一次性有效证书,所以有且仅有一次发行机会,因此不会给系统造成瓶颈;

(3)DAA 证书可以颁发给制造商,也可以发给购买的平台。很容易促进基于可信计算的 Ad Hoc 网络的安全性。

4 结束语

如果 Ad Hoc 技术被滥用,会导致大量的网络犯罪,所以文中提出基于可信计算的 Ad Hoc 网络。在节点访问网络之前,使用可信计算理论去验证和监控网络,可以确保网络节点可信,使 Ad Hoc 网络更安全。未来的工作将集中在 Ad Hoc 网络认证评估、TPM 和其他平台研究。随着可信计算的发展,Ad Hoc 将达到一个新的水平。

参考文献:

[1] Abraham J. A survey of intrusion detection for ad-hoc network [J]. Journal of global research in computer science, 2013, 4(4): 182-185.

[2] Noorman J, Agten P, Daniels W, et al. Sancus: Low-cost trustworthy extensible networked devices with a zero-software trusted computing base[C]//Proc of 22nd USENIX security symposium. [s. l.]: [s. n.], 2013.

[3] Brickell E F F, Li J. Apparatus and method for direct anonymous attestation from bilinear maps: U. S., 20, 130, 080, 771 [P]. 2013-03-28.

[4] Deodi P, Shrivastava S, Bhatele M. Security issues in monitoring medical disease through vehicular Ad Hoc network[C]// Proceedings of all India seminar on biomedical engineering 2012 (AISOBE 2012). India: Springer, 2013: 147-152.

[5] 金 伟,刘方爱,王晓洁. 基于 NS 的 Ad hoc 网络路由协议仿真研究[J]. 计算机技术与发展,2010,20(1):63-66.

[6] 沈 奔,秦 军,万 丽. 无线 Ad Hoc 网络中 AODV 路由算法的研究与改进[J]. 计算机技术与发展,2011,21(3):150-153.

[7] Kale M R A, Gupta S R, Prmit R B. An overview of Manet Ad Hoc network [J]. International journal of computer science and applications, 2013, 6(2): 223-227.

[8] Barton M, Kwon T J. Security method for mobile Ad Hoc networks with efficient flooding mechanism using layer independent passive clustering (LIPC): U. S., 20, 130, 145, 461[P]. 2013-06-06.

[9] 李奕男,钱志鸿,刘 影,等. 基于博弈论的移动 Ad hoc 网络入侵检测模型[J]. 电子与信息学报,2010,32(9):2245-2248.

[10] 刘孜文,冯登国. 基于可信计算的动态完整性度量架构[J]. 电子与信息学报,2010,32(4):875-879.

[11] 宋 成,孙宇琼,彭维平,等. 改进的直接匿名认证方案[J]. 北京邮电大学学报,2011,34(3):62-65.

[12] 陈小峰,冯登国. 一种基于双线性映射的直接匿名证明方案[J]. 软件学报,2010,21(8):2070-2078.

[13] 刘雪艳. 基于 ID 可证安全的带权限身份认证方案[J]. 计算机工程,2011,37(4):151-153.

[14] 李 睿,徐秋亮. 一种构造并发不可延展零知识的新方法[J]. 计算机学报,2012,35(4):682-692.

[15] 刘景美,王新梅. Schnorr 签名方案的一种攻击[J]. 计算机科学,2006,33(7):141-142.

(上接第 146 页)

参考文献:

[1] Hua N, Song H, Lakshman T. Variable-stride multi-pattern matching for scalable deep packet inspection[C]//Proc of INFOCOM 2009. [s. l.]: IEEE, 2009: 415-423.

[2] 杨 玲,孟传良. 基于启发式分析的木马检测技术研究[J]. 现代机械,2006(4):61-63.

[3] 唐彰国,李焕洲,钟明全,等. 基于网络通信指纹的启发式木马识别系统[J]. 计算机工程,2011,37(17):119-121.

[4] 孙玉星,黄松华,黄 皓,等. 自治网络中信任信誉模型的安全现状研究[J]. 计算机科学,2009,36(4):5-11.

[5] Martignoni L, Stinson E, Fredrikson M, et al. A layered architecture for detecting malicious behaviors[C]//Proc of RAID'08. Cambridge, USA: [s. n.], 2008: 78-97.

[6] 杨卫平. 面向虚拟机的网络入侵检测系统[D]. 武汉:华中科技大学,2008.

[7] Acharya S, Mills B N, Abliz M, et al. Optwall: A hierarchical traffic-aware firewall[C]//Proc of NDSS. [s. l.]: [s. n.], 2007: 528-533.

[8] 王倍昌. 走进计算机病毒[M]. 北京:人民邮电出版社,2010.

[9] 张仁斌,李 钢,侯整风. 计算机病毒与反病毒技术[M]. 北京:清华大学出版社,2010.

[10] 辛 毅,方滨兴,贺龙涛,等. 基于通信特征分析的蠕虫检测和特征提取方法的研究[J]. 通信学报,2007,28(12):1-7.

[11] 钟明全,李焕洲,唐彰国,等. 基于网络驱动技术的木马通信检测系统[J]. 计算机工程,2010,36(9):150-152.

[12] 单长虹,张焕国,孟庆树,等. 一种启发式木马查杀模型的设计与分析[J]. 计算机工程与应用,2004(20):130-132.

[13] 刘晓石,陈鸿建,何腊梅. 概率论与数理统计[M]. 北京:科学出版社,2006.

万方数据 WANFANG DATA 文献链接

基于分层权值的恶意程序仿真系统设计与实现

作者： 江雪， 朱永强， JIANG Xue， ZHU Yong-qiang
作者单位： 电子科技大学 示范性软件学院,四川 成都,610054
刊名： 计算机技术与发展
ISTIC
英文刊名： Computer Technology and Development
年，卷(期)： 2014(4)

本文链接：http://d.wanfangdata.com.cn/Periodical_wjfz201404036.aspx