

(12) 发明专利

(10) 授权公告号 CN 101989322 B

(45) 授权公告日 2012. 11. 21

(21) 申请号 201010551270. 3

审查员 孙国辉

(22) 申请日 2010. 11. 19

(73) 专利权人 北京安天电子设备有限公司

地址 100085 北京市海淀区农大南路 1 号硅谷亮城 2B-521

(72) 发明人 肖梓航

其他发明人请求不公开姓名

(51) Int. Cl.

G06F 21/00(2006. 01)

G06F 21/22(2006. 01)

(56) 对比文件

CN 101645119 A, 2010. 02. 10,

WO 03/096607 A1, 2003. 11. 20,

CN 101685483 A, 2010. 03. 31,

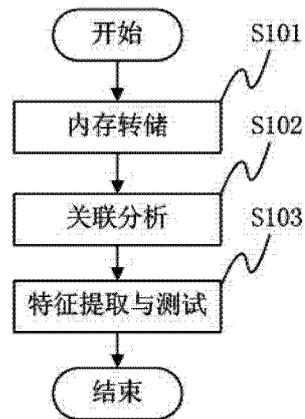
权利要求书 2 页 说明书 8 页 附图 6 页

(54) 发明名称

自动提取恶意代码内存特征的方法和系统

(57) 摘要

本发明公开了一种自动提取恶意代码内存特征的方法,包括:运行恶意代码对新产生的线程信息进行内存转储,生成转储文件;对转储文件进行关联分析并分组;对分组的转储文件提取特征并进行测试处理;系统包括:内存转储模块,用于运行恶意代码对新产生的线程信息进行内存转储,生成转储文件;关联分析模块,用于对转储文件进行关联分析并分组;特征提取与测试模块,用于对分组的转储文件提取特征并进行测试处理。本发明整个方案都是自动化的流程,无需人工参与,以线程为基本处理对象,实现了细粒度的更精确和全面的内存特征提取,不再依赖于分析人员的经验,最终获得的内存特征具有较低的误报率和极低的漏报率。



1. 一种自动提取恶意代码内存特征的方法,其特征在于,包括:

步骤 a、运行恶意代码,对新产生的线程信息进行内存转储,生成转储文件;所述的内存转储过程至少执行两次,每次都生成一批转储文件;

步骤 b、对转储文件进行关联分析并分组,具体包括:

比较同一批转储文件中的每两个转储文件的相似性,如果相似,则删除其中一个转储文件,保留另外一个转储文件;

比较所有保留的转储文件中的每两个转储文件的相似性,将相似的转储文件归为一组;

步骤 c、对分组的转储文件进行特征提取与测试处理。

2. 如权利要求 1 所述的自动提取恶意代码内存特征的方法,其特征在于,步骤 a 的具体步骤包括:

a1、对没有任何恶意代码运行的操作系统当前所有线程创建快照,记录所有线程的信息,其中至少包括线程 ID;

a2、在该操作系统中运行恶意代码;

a3、经过预设的时间后,遍历该操作系统中的当前所有线程,对于每一个线程,在快照中进行搜索,找出快照中没有的新产生的线程;

a4、查询该线程的入口点地址,根据该地址查询入口点所在的内存块;

a5、读取该线程入口点所在的内存块,将该线程入口点所在的内存块的内容转储为二进制文件,并记录该线程入口点在文件中的相对偏移字节数。

3. 如权利要求 1 所述的自动提取恶意代码内存特征的方法,其特征在于:步骤 b 中,判断两个转储文件是否相似,具体包括:

比较转储文件的大小是否相等,如果不相等,则不相似;

比较线程入口点在转储文件中的相对偏移字节数是否相等,如果不相等,则不相似;

对比转储文件的内容,如果相同内容占所有内容的百分比超过一个固定的阈值,则认为它们相似,否则不相似。

4. 如权利要求 1 所述的自动提取恶意代码内存特征的方法,其特征在于,步骤 c 中,对一组转储文件进行特征提取与测试处理具体步骤包括:

c1、对于包括多个转储文件的分组,将同组转储文件互相对比,从线程入口点开始遍历,直到同组所有转储文件有相同数据,取一段该相同数据,将这段数据作为待定特征;

c2、在预先配置的没有运行所述恶意代码的操作系统中,用该待定特征进行内存扫描;如果扫描到病毒,则为误报,舍弃该待定特征,转到步骤 c1 取下一条待定特征;如果没有扫描到病毒,则通过误报测试;

c3、在预先配置的运行了所述恶意代码的操作系统中,用该待定特征进行内存扫描;如果没有扫描到病毒,则为漏报,舍弃该待定特征,转到步骤 c1 取下一条待定特征;如果扫描到病毒,则通过漏报测试,该待定特征成为正式内存特征。

5. 如权利要求 4 所述的自动提取恶意代码内存特征的方法,其特征在于,步骤 c 中,对每组转储文件进行特征提取与测试处理,得到所述恶意代码的所有正式内存特征。

6. 一种自动提取恶意代码内存特征的系统,其特征在于,包括:

内存转储模块,用于运行恶意代码,对新产生的线程进行内存转储,生成转储文件;所

述内存转储模块对所述恶意代码进行内存转储的处理至少执行两次,每次都生成一批转储文件;

关联分析模块,用于对转储文件进行关联分析并分组,具体包括:

比较同一批转储文件中的,每两个转储文件的相似性,如果相似,则删除其中一个转储文件,保留另外一个转储文件;

比较所有保留的转储文件中的每两个转储文件的相似性,将相似的转储文件归为一组;

特征提取与测试模块,用于对分组的转储文件进行特征提取与测试处理。

7. 如权利要求 6 所述的自动提取恶意代码内存特征的系统,其特征在于,所述内存转储模块中,进行一次内存转储处理具体包括:

对没有任何恶意代码运行的操作系统当前所有线程创建快照,记录下所有线程的信息,其中至少包括线程 ID;

在该操作系统中运行要提取特征的恶意代码;

经过预设的时间后,遍历该操作系统中的当前所有线程,对于每一个线程,在快照中进行搜索,找出快照中没有的新产生的线程;

查询该线程的入口点地址,根据该地址查询入口点所在的内存块;

读取该线程入口点所在的内存块,将该线程入口点所在内存块的内容转储为二进制文件,并记录该线程入口点在文件中的相对偏移字节数。

8. 如权利要求 6 所述的自动提取恶意代码内存特征的系统,其特征在于,所述关联分析模块中,判断转储文件是否相似,具体包括:

比较转储文件的大小是否相等,如果不相等,则不相似;

比较线程入口点在转储文件中的相对偏移字节数是否相等,如果不相等,则不相似;

对比转储文件的内容,如果相同内容占所有内容的百分比超过一个固定的阈值,则认为它们相似,否则不相似。

9. 如权利要求 6 所述的自动提取恶意代码内存特征的系统,其特征在于,所述特征提取与测试模块中,对一组内存转储文件进行特征提取与测试处理具体包括:

对于包括多个转储文件的分组,将同组转储文件互相对比,从线程入口点开始往后遍历,直到同组所有转储文件有相同数据,取一段该相同数据,将这段数据作为待定特征;

在预先配置的没有运行所述恶意代码的操作系统中,用该待定特征进行内存扫描;如果扫描到病毒,则为误报,舍弃该待定特征,重新取下一条待定特征;如果没有扫描到病毒,则通过误报测试;

在预先配置的运行了所述恶意代码的操作系统中,用该待定特征进行内存扫描;如果没有扫描到病毒,则为漏报,舍弃该待定特征,重新取下一条待定特征;如果扫描到病毒,则通过漏报测试;该待定特征成为正式内存特征。

10. 如权利要求 9 所述的自动提取恶意代码内存特征的系统,其特征在于,所述特征提取与测试模块对每组转储文件进行特征提取与测试处理,得到所述恶意代码的所有正式内存特征。

## 自动提取恶意代码内存特征的方法和系统

### 技术领域

[0001] 本发明涉及计算机安全技术,尤其涉及自动提取恶意代码内存特征的方法和系统。

### 背景技术

[0002] [0002] 恶意代码(包括木马、蠕虫、病毒等)是信息安全领域最严重的威胁,如何有效地发现、检测、清除、防御、遏制计算机与网络中的恶意代码,是该领域最核心的问题之一。

[0003] 恶意代码在计算机中呈现为两种形态:静态的文件,是它的宿主;动态的进程和线程,是它的实际执行体。在检测和清除过程中,既要检测出恶意代码的宿主文件,将文件删除,又要检测出它创建的进程和线程,将其终止。

[0004] 对恶意代码的检测,通常采用特征匹配的方法。使用恰当的特征进行检测,是降低检测的误报率和漏报率、提高检测精确度的关键。对应于静态和动态两种场景的检测需求,分别有文件特征和内存特征。

[0005] 根据内存特征对进程和线程进行检测,有很高的现实意义:一方面,如果只是删除了恶意代码所在文件,而进程和线程仍然在系统中运行,它所产生的危害并未消除;另一方面,很大一部分恶意代码采用了“加壳”等技术,来对抗文件特征检测技术,此时就需要通过进程和线程的检测结果,反过来判断文件是否为恶意代码的宿主。

[0006] 在内存特征的提取方面,目前一般采用手工提取的方法。流程如下所述:

[0007] 步骤 S10:运行恶意代码,观察它创建了哪些进程,对每个新建进程执行步骤 S11;

[0008] 步骤 S11:使用动态调试工具,或者内存转储工具,获得该进程的内存内容,并将其转储为文件;

[0009] 步骤 S12:使用反汇编工具,将转储的内存文件反汇编,得到它的汇编代码;

[0010] 步骤 S13:分析汇编代码,寻找其中恶意代码特有的攻击代码;

[0011] 步骤 S14:在上述特有攻击代码中,寻找一段适当长度的代码,使得其中不包含需要重定位的部分;

[0012] 步骤 S15:将这段代码对应的二进制数据作为该进程的待定内存特征;

[0013] 步骤 S16:测试待定内存特征,如果有误报或漏报,则将其抛弃,返回步骤 S13 重新提取;如果没有误报和漏报,则将其作为该进程的内存特征。

[0014] 现有方案存在以下不足之处:在监视恶意代码创建的进程和线程方面,依赖于人工观察,难以保证监视结果的完整性,并且只能做到进程粒度,对于在已有进程中创建新线程的恶意代码,无法监视;对汇编代码的分析,非常依赖于分析人员的经验积累,而且也需要大量的时间;对误报和漏报的判定依赖于人工观察,有可能判断不准确。

### 发明内容

[0015] 针对以上不足,本发明要解决的技术问题是提供一种自动提取恶意代码内存特征

的方法和系统,实现自动提取恶意代码内存特征,以线程为基本处理对象,最终获得的内存特征具有较低的误报率和极低的漏报率。

[0016] 为了解决上述技术问题,本发明提供一种自动提取恶意代码内存特征的方法,包括:

[0017] 步骤 a、运行恶意代码并进行内存转储,对新产生的线程信息进行内存转储,生成转储文件;

[0018] 步骤 b、对转储文件进行关联分析并分组;

[0019] 步骤 c、对分组的转储文件进行特征提取与测试处理。

[0020] 进一步的,步骤 a 中,一次内存转储过程具体包括:

[0021] a1、对没有任何恶意代码运行的操作系统当前所有线程创建快照,记录所有线程的信息,其中至少包括线程 ID;

[0022] a2、在该操作系统中运行要提取特征的恶意代码;

[0023] a3、经过预设的时间后,遍历该操作系统中的当前所有线程,对于每一个线程,在快照中进行搜索,找出快照中没有的新产生的线程;

[0024] a4、查询该线程的入口点地址,根据该地址查询入口点所在的内存块;

[0025] a5、读取该线程入口点所在的内存块,将该线程入口点所在的内存块的内容转储为二进制文件,并记录该线程入口点在文件中的相对偏移字节数。

[0026] 进一步的,对所述恶意代码进行内存转储的过程至少执行两次,每次都生成一批转储文件。

[0027] 进一步的,步骤 b 中,对转储文件进行关联分析具体为判断两个转储文件是否相似,具体包括:

[0028] 比较转储文件的大小是否相等,如果不相等,则不相似;

[0029] 比较线程入口点在转储文件中的相对偏移字节数是否相等,如果不相等,则不相似;

[0030] 对比转储文件的内容,如果相同内容占所有内容的百分比超过一个固定的阈值,则认为它们相似,否则不相似。

[0031] 进一步的,步骤 b 具体包括:

[0032] 比较同一批转储文件中的每两个转储文件的相似性,如果相似,则删除其中一个转储文件,保留另外一个转储文件;

[0033] 比较所有保留的转储文件中的每两个转储文件的相似性,将相似的转储文件归为一组。

[0034] 进一步的,步骤 c 中,对一组转储文件进行特征提取与测试处理具体步骤包括:

[0035] c1、对于包括多个转储文件的分组,将同组转储文件互相对比,从线程入口点开始遍历,直到同组所有转储文件有相同数据,取一段该相同数据,将这段数据作为待定特征;

[0036] c2、在预先配置的没有运行所述恶意代码的操作系统中,用该待定特征进行内存扫描;如果扫描到病毒,则为误报,舍弃该待定特征,转到步骤 c1 取下一条待定特征;如果没有扫描到病毒,则通过误报测试;

[0037] c3、在预先配置的运行了所述恶意代码的操作系统中,用该待定特征进行内存扫描;如果没有扫描到病毒,则为漏报,舍弃该待定特征,转到步骤 c1 取下一条待定特征;如

果扫描到病毒,则通过漏报测试,该待定特征成为正式内存特征。

[0038] 进一步的,步骤 c 中,对每组转储文件进行特征提取与测试处理,得到所述恶意代码的所有正式内存特征。

[0039] 本发明还提供了一种自动提取恶意代码内存特征的系统,包括:

[0040] 内存转储模块,用于运行恶意代码,对新产生的线程进行内存转储,生成转储文件;

[0041] 关联分析模块,用于对转储文件进行关联分析并分组;

[0042] 特征提取与测试模块,用于对分组的转储文件进行特征提取与测试处理。

[0043] 进一步的,所述内存转储模块中,进行一次内存转储处理具体包括:

[0044] 对没有任何恶意代码运行的操作系统当前所有线程创建快照,记录下所有线程的信息,其中至少包括线程 ID;

[0045] 在该操作系统中运行要提取特征的恶意代码;

[0046] 经过预设的时间后,遍历该操作系统中的当前所有线程,对于每一个线程,在快照中进行搜索,找出快照中没有的新产生的线程;

[0047] 查询该线程的入口点地址,根据该地址查询入口点所在的内存块;

[0048] 读取该线程入口点所在的内存块,将该线程入口点所在内存块的内容转储为二进制文件,并记录该线程入口点在文件中的相对偏移字节数。

[0049] 进一步的,对所述恶意代码进行内存转储的处理至少执行两次,每次都生成一批转储文件。

[0050] 进一步的,所述关联分析模块中,对转储文件进行关联分析具体为判断转储文件是否相似,具体包括:

[0051] 比较转储文件的大小是否相等,如果不相等,则不相似;

[0052] 比较线程入口点在转储文件中的相对偏移字节数是否相等,如果不相等,则不相似;

[0053] 对比转储文件的内容,如果相同内容占所有内容的百分比超过一个固定的阈值,则认为它们相似,否则不相似。

[0054] 进一步的,所述关联分析模块的功能具体包括:

[0055] 比较同一批转储文件中的,每两个转储文件的相似性,如果相似,则删除其中一个转储文件,保留另外一个转储文件;

[0056] 比较所有保留的转储文件中的每两个转储文件的相似性,将相似的转储文件归为一组。

[0057] 进一步的,所述特征提取与测试模块中,对一组内存转储文件进行特征提取与测试处理具体包括:

[0058] 对于包括多个转储文件的分组,将同组转储文件互相对比,从线程入口点开始往后遍历,直到同组所有转储文件有相同数据,取一段该相同数据,将这段数据作为待定特征;

[0059] 在预先配置的没有运行所述恶意代码的操作系统中,用该待定特征进行内存扫描;如果扫描到病毒,则为误报,舍弃该待定特征,重新取下一条待定特征;如果没有扫描到病毒,则通过误报测试;

[0060] 在预先配置的运行了所述恶意代码的操作系统中,用该待定特征进行内存扫描;如果没有扫描到病毒,则为漏报,舍弃该待定特征,重新取下一条待定特征;如果扫描到病毒,则通过漏报测试;该待定特征成为正式内存特征。

[0061] 进一步的,所述特征提取与测试模块对每组转储文件进行特征提取与测试处理,得到所述恶意代码的所有正式内存特征。

[0062] 本发明的有益效果是:

[0063] 本发明可以实现完全的自动化,无需人工操作,也不依赖于病毒分析人员的专业知识和分析经验;

[0064] 本发明所提取内存特征是基于线程的,有更细的粒度;当恶意代码是远程注入型木马时,可以只清除线程,而不终止其宿主进程,这样就将病毒清除工作对系统的影响降至最低;

[0065] 通过转储恶意代码衍生线程的内存内容,提取的特征来自于其衍生线程入口点往后的内存内容,是其实际执行的代码,即而已代码,具有代表性;

[0066] 通过多次转储与相似性判断,消除重定位产生的漏报,消除转储时系统新建线程产生的误报;并通过实际环境测试,进一步降低特征的误报率和漏报率,最终获得高质量内存特征。

#### 附图说明

[0067] 为了更清楚地说明本发明或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0068] 图 1 为本发明自动提取恶意代码内存特征方法的整体流程图;

[0069] 图 2 为本发明自动提取恶意代码内存特征方法的内存转储流程图;

[0070] 图 3 为本发明自动提取恶意代码内存特征方法的转储文件相似度判断流程图;

[0071] 图 4 为本发明自动提取恶意代码内存特征方法的关联分析实施例流程图;

[0072] 图 5 为本发明自动提取恶意代码内存特征方法的特征提取与测试流程图;

[0073] 图 6 为本发明自动提取恶意代码内存特征方法的特征提取与测试实施例流程图;

[0074] 图 7 为本发明自动提取恶意代码内存特征的系统示意图。

#### 具体实施方式

[0075] 为了使本技术领域的人员更好地理解本发明实施例中的技术方案,并使本发明的上述目的、特征和优点能够更加明显易懂,下面结合附图对本发明中技术方案作进一步详细的说明。

[0076] 本发明提供了一种自动提取恶意代码内存特征的方法和系统,可以实现恶意代码内存特征提取的自动化流程,并且得到的内存特征具有一定的代表性、误报率低、漏报率极低,有较高的实用价值,可用于杀毒软件、计算机安全辅助软件或杀毒引擎的特征库,作为恶意代码检测和清除的判断依据。

[0077] 首先介绍本发明提供的自动提取恶意代码内存特征的方法,具体实施步骤如图 1

所示,包括三个阶段:

[0078] S101、内存转储;运行恶意代码并进行内存转储,对新产生的线程信息进行内存转储,生成转储文件;

[0079] S102、关联分析;对转储文件进行关联分析并分组;

[0080] S103、特征提取与测试;对分组的转储文件进行特征提取与测试处理。

[0081] 内存转储阶段 S101 中,内存转储的具体实施方式如图 2 所示,包括:

[0082] S201、对没有任何恶意代码运行的操作系统当前所有线程创建快照,记录所有线程的信息,其中至少包括线程 ID;

[0083] S202、在该操作系统中运行要提取特征的恶意代码;

[0084] S203、经过预设的时间后,遍历该操作系统中的当前所有线程,对于每一个线程,在快照中进行搜索,找出快照中没有的新产生的线程;

[0085] 其中,等待预设时间的目的是让恶意代码完全激活、所有行为都开始执行,包括创建衍生线程;

[0086] S204、查询该线程的入口点地址,根据该地址查询入口点所在的内存块;

[0087] 在 Windows 中,这两个操作可以通过系统提供的 NtQueryInformationThread() 和 VirtualQueryEx() 接口函数来实现;

[0088] S205、读取该线程入口点所在的内存块,将该线程入口点所在的内存块的内容转储为二进制文件,并记录该线程入口点在文件中的相对偏移字节数。

[0089] 上述 S201 至 S205 为一次内存转储过程,内存转储阶段 S101 中对所述恶意代码进行内存转储的过程至少执行两次,每次都生成一批转储文件。

[0090] 关联分析阶段 S102 中,对转储文件进行关联分析具体为判断两个转储文件是否相似;

[0091] 相似是指认为两个转储文件从执行相同代码的线程中转储而来;例如,恶意代码开启多个线程通过相同的系统函数调用来向网络发送数据,这些线程就执行相同的代码,其转储文件是相似的;之所以是相似而不是相同,是因为重定位,即一段代码的多次运行(表现为不同的线程),内存中的内容不一定完全一致;

[0092] 判断相似的具体方法如图 3 所示,包括:

[0093] S301、比较转储文件的大小是否相等,如果不相等,则不相似;否则执行 S302;

[0094] S302、比较线程入口点在转储文件中的相对偏移字节数是否相等,如果不相等,则不相似;否则执行 S303;

[0095] S303、对比转储文件的内容,如果相同内容占所有内容的百分比超过一个固定的阈值,则认为它们相似,否则不相似;

[0096] 在实践中,这个阈值可以使用 95%,这是一个经验指导值;是则相似,否则不相似。

[0097] 关联分析阶段 S102 具体包括:

[0098] 比较同一批转储文件中的每两个转储文件的相似性,如果相似,则删除其中一个转储文件,保留另外一个转储文件;

[0099] 比较所有保留的转储文件中的每两个转储文件的相似性,将相似的转储文件归为一组;

[0100] 关联分析阶段 S102 的一个具体实施例如图 4 所示,可以划分为分两个阶段:

- [0101] 同一批转储文件阶段：
- [0102] S401、读取同一批转储文件；
- [0103] S402、两两进行相似性比较；
- [0104] S403、如果相似，则认为是执行相同代码的线程，删除其中一个，留下另外一个；
- [0105] S404、判断所有文件是否比较完，是则执行 S205，否则执行 S201；
- [0106] 不同批次转储文件阶段：
- [0107] S405、读取不同批次转储文件；
- [0108] S406、两两进行相似性比较；
- [0109] S407、相似的文件归为一组；
- [0110] S408、判断所有文件是否比较完，是则结束，否则执行 S206。
- [0111] 经过步骤 S401 至 S408，多次转储得到的二进制文件被分为不同的组，每一组中的文件是相同功能线程在多次转储中得到的，具有相同的文件大小、入口点相对偏移字节数，并有较高的相似度。
- [0112] 特征提取与测试阶段 S103 中，对一组转储文件进行特征提取与测试处理的具体实施方式如图 5 所示，包括：
- [0113] S501、对于包括多个转储文件的分组，将同组转储文件互相对比，从线程入口点开始遍历，直到同组所有转储文件有相同数据，取一段该相同数据，将这段数据作为待定特征；
- [0114] S502、在预先配置的没有运行所述恶意代码的操作系统中，用该待定特征进行内存扫描；如果扫描到病毒，则为误报，舍弃该待定特征，转到步骤 c1 取下一条待定特征；如果没有扫描到病毒，则通过误报测试；
- [0115] S503、在预先配置的运行了所述恶意代码的操作系统中，用该待定特征进行内存扫描；如果没有扫描到病毒，则为漏报，舍弃该待定特征，转到步骤 c1 取下一条待定特征；如果扫描到病毒，则通过漏报测试，该待定特征成为正式内存特征；
- [0116] 上述步骤 S501 至 S503 是对一组转储文件进行特征提取与测试处理，特征提取与测试阶段 S103 中对每组转储文件进行特征提取与测试处理，得到所述恶意代码的所有正式内存特征。
- [0117] 下面给出特征提取与测试处理阶段 S103 的一个具体实施例，如图 6 所示，包括：
- [0118] S601、取上述同组的转储文件，如果一组中只有一个文件，则跳过，不予考虑；
- [0119] 跳过是为了降低误报率。因为内存转储的步骤 S102 中，等待了一段时间，这段时间操作系统可能创建了其他与恶意代码无关的线程，也被转储了；但这样的偶然事件在多次转储中都发生的概率不大，因此如果一组中只有一个文件，就认为对应于这样的事件，应该不予考虑，以避免从系统线程中提取到特征，产生误报；
- [0120] S602、在对转储文件分组时，已经确保同一组的文件有相同的大小和入口点偏移，并且有较高的相似度；从入口点偏移开始往后遍历，将同组文件互相对比，直到在某一个偏移处取得一段数据，使其在同组所有文件都完全相同；数据长度由系统要求的特征长度决定，例如 128 字节；将这段数据作为待定特征；
- [0121] S603、在一个预先配置的没有运行该恶意代码的操作系统中，使用普通的内存扫描技术，用待定特征进行扫描；

[0122] S604、如果扫描到有病毒,则为误报,舍弃该待定特征,转入步骤 S402 继续取下一条待定特征;如果没有扫描到有病毒,则通过误报测试;

[0123] S605、在一个运行了该恶意代码的操作系统中,使用普通的内存扫描技术,用待定特征扫描;

[0124] S606、如果没有扫描到有病毒,则为漏报,舍弃该待定特征,转入步骤 S402 继续取下一条待定特征;如果扫描到有病毒,则通过漏报测试;

[0125] S607、待定特征成为该恶意代码的正式内存特征之一,将其录入病毒特征库;

[0126] S608、对该恶意代码转储的每一组文件,执行步骤 S401 到步骤 S407,得到该恶意代码的所有正式内存特征。

[0127] 其中,步骤 S603 到步骤 S606,涉及将待定特征送入不同操作系统环境中进行内存扫描,这些工作可以通过虚拟机技术与脚本语言结合,实现自动处理,而无需手工操作。

[0128] 本发明还提供了一种自动提取恶意代码内存特征的系统,如图 7 所示,包括:内存转储模块 101,用于运行恶意代码,对新产生的线程进行内存转储,生成转储文件;

[0129] 关联分析模块 102,用于对转储文件进行关联分析并分组;

[0130] 特征提取与测试模块 103,用于对分组的转储文件进行特征提取与测试处理。

[0131] 其中,内存转储模块 101 中,内存转储处理功能具体包括:

[0132] 对没有任何恶意代码运行的操作系统当前所有线程创建快照,即记录下所有线程的信息,至少包括线程 ID;

[0133] 在该操作系统中运行要提取特征的恶意代码;

[0134] 经过预设的时间后,遍历该操作系统中的当前所有线程,对于每一个线程,在快照中进行搜索,找出快照中没有的新产生的线程;

[0135] 查询该线程的入口点地址,根据该地址查询入口点所在的内存块;

[0136] 读取该线程入口点所在的内存块,将该线程入口点所在内存块的内容转储为二进制文件,并记录该线程入口点在文件中的相对偏移字节数。

[0137] 内存转储模块 101 对所述恶意代码进行内存转储的处理至少执行两次,每次都生成一批转储文件。

[0138] 关联分析模块 102 中,对转储文件进行关联分析具体为判断转储文件是否相似,具体包括:

[0139] 比较转储文件的大小是否相等,如果不相等,则不相似;

[0140] 比较线程入口点在转储文件中的相对偏移字节数是否相等,如果不相等,则不相似;

[0141] 对比转储文件的内容,如果相同内容占有所有内容的百分比超过一个固定的阈值,则认为它们相似,否则不相似。

[0142] 关联分析模块 102 的功能具体包括:

[0143] 比较同一批转储文件中的,每两个转储文件的相似性,如果相似,则删除其中一个转储文件,保留另外一个转储文件;

[0144] 比较所有保留的转储文件中的每两个转储文件的相似性,将相似的转储文件归为一组。

[0145] 特征提取与测试模块 103 中,对一组内存转储文件进行特征提取与测试处理具体

包括：

[0146] 对于包括多个转储文件的分组，将同组转储文件互相对比，从线程入口点开始往后遍历，直到同组所有转储文件有相同数据，取一段该相同数据，将这段数据作为待定特征；

[0147] 在预先配置的没有运行所述恶意代码的操作系统中，用该待定特征进行内存扫描；如果扫描到病毒，则为误报，舍弃该待定特征，重新取下一条待定特征；如果没有扫描到病毒，则通过误报测试；

[0148] 在预先配置的运行了所述恶意代码的操作系统中，用该待定特征进行内存扫描；如果没有扫描到病毒，则为漏报，舍弃该待定特征，重新取下一条待定特征；如果扫描到病毒，则通过漏报测试；该待定特征成为正式内存特征。

[0149] 特征提取与测试模块 103 对每组转储文件进行特征提取与测试处理，得到所述恶意代码的所有正式内存特征。

[0150] 通过以上具体实施方式的描述，本发明通过转储恶意代码衍生线程的内存内容，获得提取来源；通过多次转储与相似性判断，消除重定位产生的漏报，消除转储时系统新建线程产生的误报；通过从入口点偏移开始选取多次转储内容中的相同部分，获得具有代表性的待定特征；通过实际环境测试，进一步降低特征的误报率和漏报率，最终获得高质量内存特征。

[0151] 当然，本发明还可有其他多种实施例，在不背离本发明精神及其实质的情况下，熟悉本领域的技术人员当可根据本发明作出各种相应的改变和变形，但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

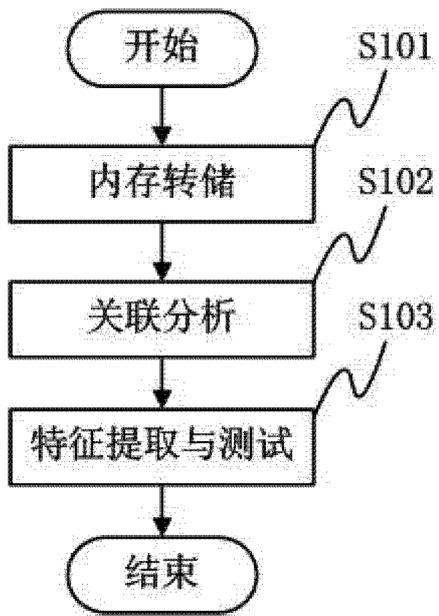


图 1

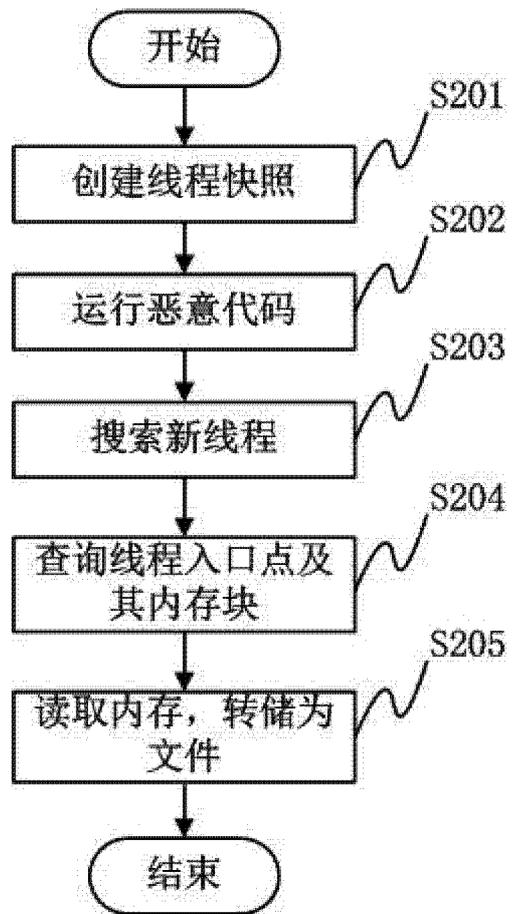


图 2

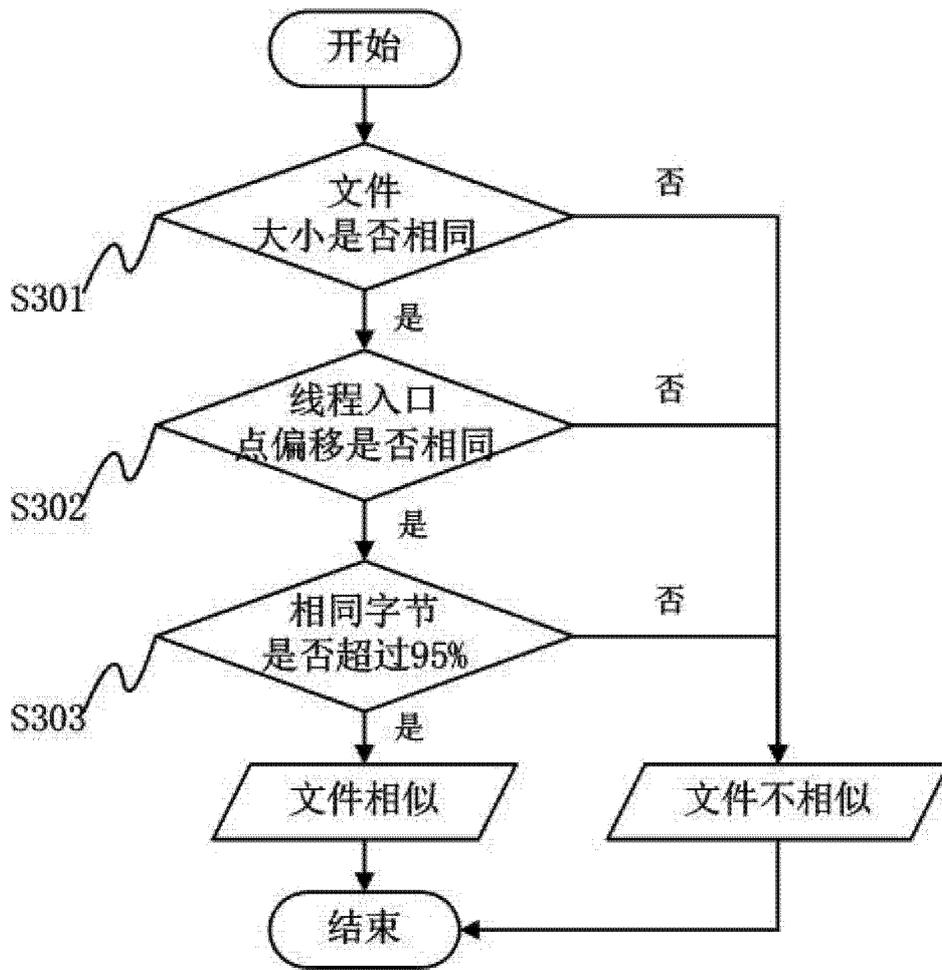


图 3

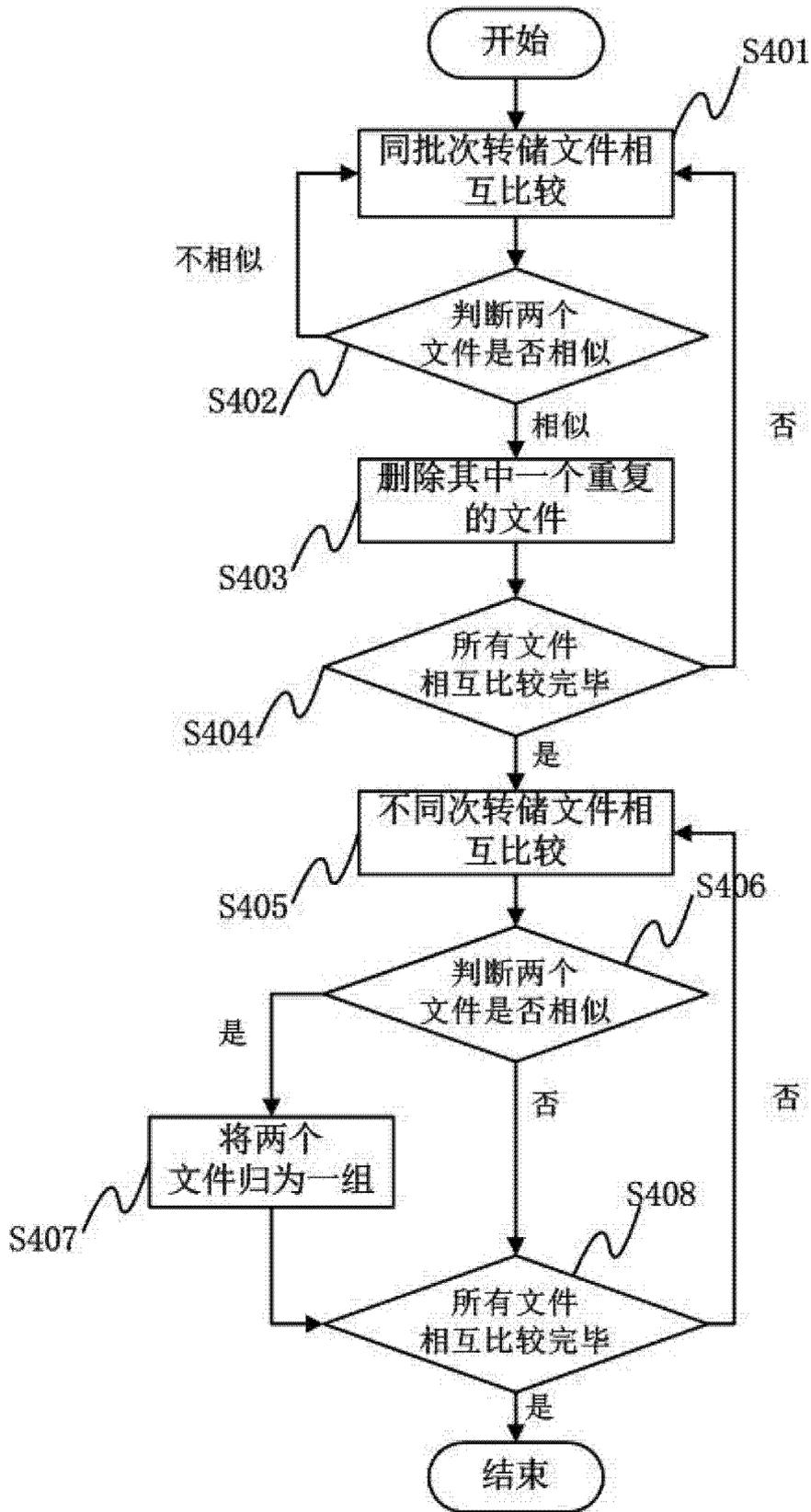


图 4

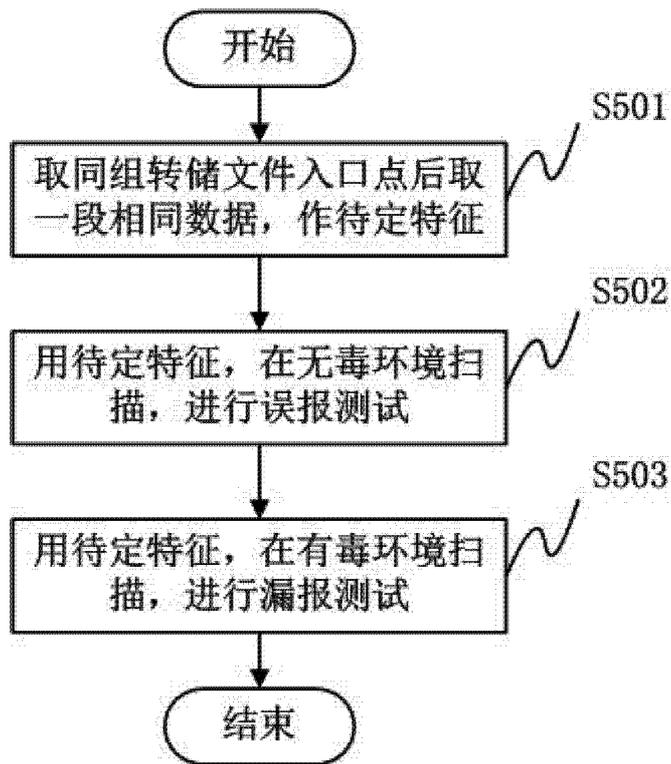


图 5

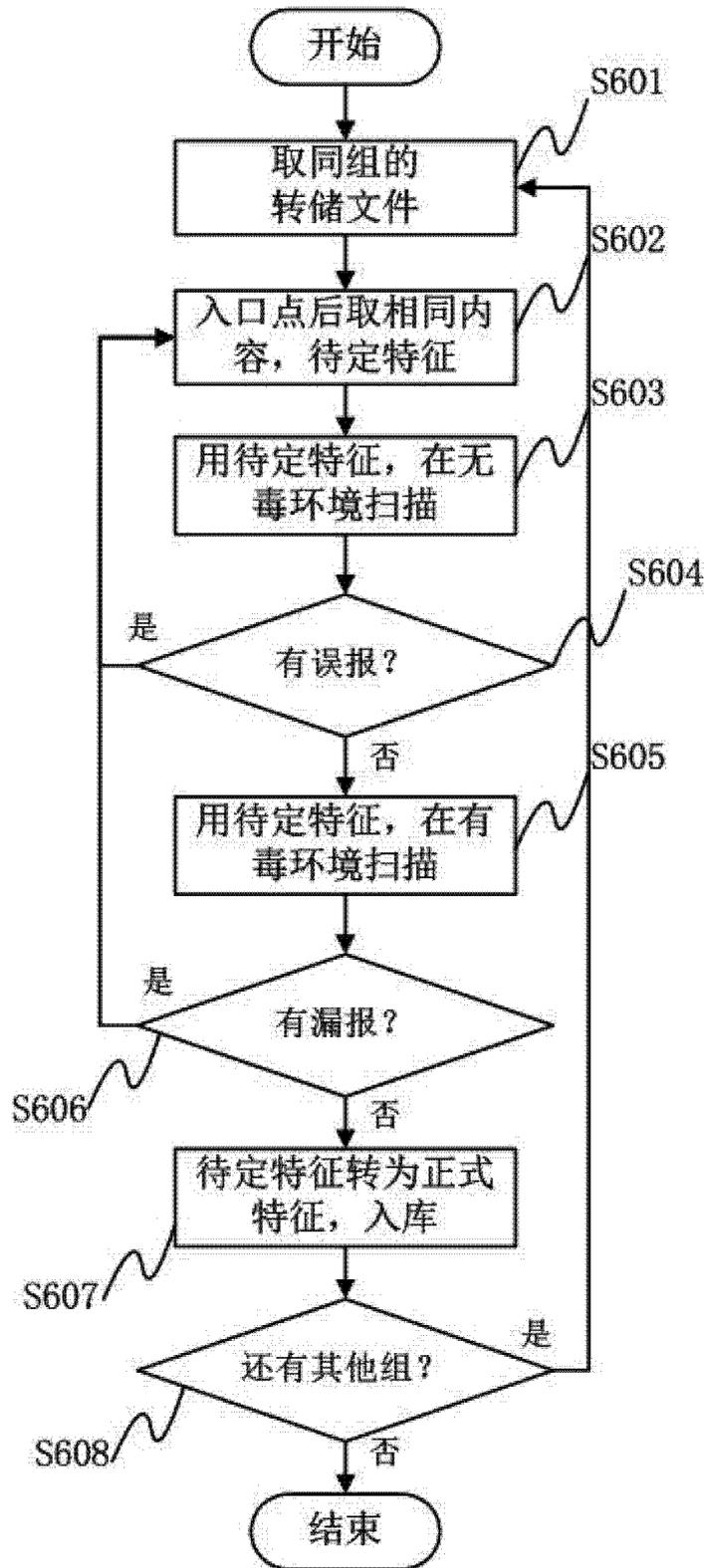


图 6

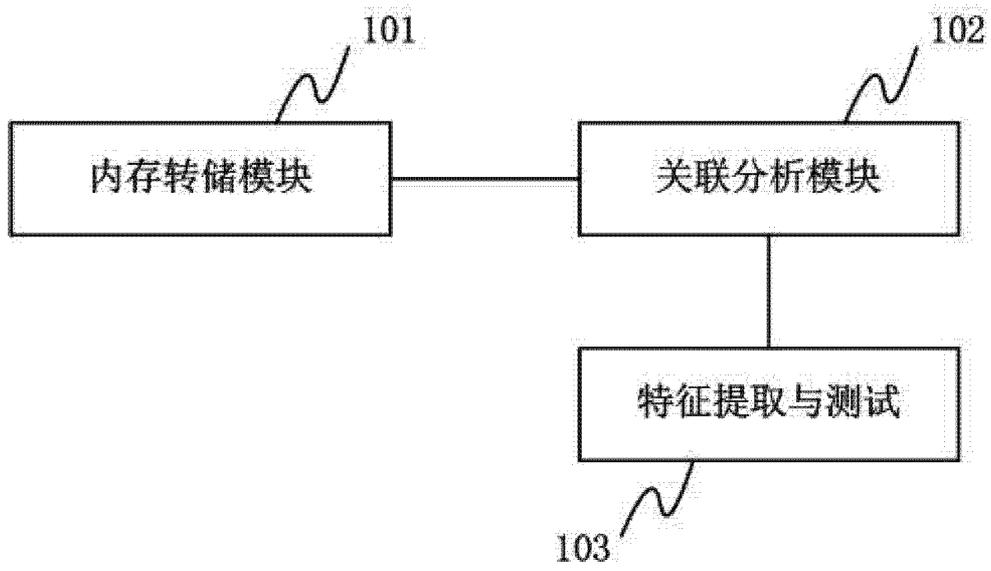


图 7