

文章编号:2095-6134(2015)05-0714-07

# 一种基于可信计算的涉密文件抗丢失技术

李凤海<sup>1†</sup>, 张佰龙<sup>1</sup>, 杜 皎<sup>1</sup>, 宋 衍<sup>1</sup>, 李 爽<sup>2</sup>

(1 信息保障技术重点实验室, 北京 100072; 2 江苏信息职业技术学院, 江苏 无锡 214101)

(2014 年 8 月 20 日收稿; 2015 年 1 月 19 日收修改稿)

Li F H, Zhang B L, Du J, et al. An anti-lost scheme for confidential files based on trusted computation[J]. Journal of University of Chinese Academy of Sciences, 2015, 32(5): 714-720.

**摘 要** 分析软件的安全威胁和涉密文件的安全威胁, 概述可信计算中软件行为动态度量技术和涉密文件数字水印技术, 提出一种基于可信计算软件动态度量和文件水印标签技术相结合的涉密文件抗丢失技术方案. 最后对该技术模型的特点进行了总结.

**关键词** 可信计算; 涉密文件; 软件动态度量

**中图分类号:** TP309      **文献标志码:** A      **doi:** 10. 7523/j. issn. 2095-6134. 2015. 05. 020

## An anti-lost scheme for confidential files based on trusted computation

LI Fenghai<sup>1</sup>, ZHANG Bailong<sup>1</sup>, DU Jiao<sup>1</sup>, SONG Yan<sup>1</sup>, LI Shuang<sup>2</sup>

(1 Science and Technology on Information Assurance Laboratory, Beijing 100072, China;

2 Jiangsu Information and Technology College, Wuxi 214101, Jiangsu, China)

**Abstract** This paper analyzes the security threats of software and the confidential file, provides an overview of software dynamic measure technology in trusted computing and digital watermarking technology for confidential files, and proposes an anti-lost scheme for confidential files by combining software dynamic measurement and digital watermarking technique. Finally, the technical characteristics of the model are summarized.

**Key words** trusted computing; confidential file; software dynamic measurement

随着全球网络信息化程度的高速发展, 计算机数据安全问题成为日益突出的问题, 特别是近几年, 从“斯诺登”事件到“苹果 iOS 后门”事件的频频曝光, 说明计算机数据信息安全问题已经上升到国家安全层面; 为了应对各种信息安全威胁, 国家的党、政、军各部门构建了各种高等级安全网络来防止信息泄密, 但事实已经证明: 高级持续威胁 (APT) 正在逐渐成为针对高价值数据目标, 最主流的攻击形式. 这种攻击手段正在日益威胁组

织机构的核心数据和重要数据资产<sup>[1]</sup>.

APT 攻击将更加盛行, 网络窃密风险加大. APT 攻击具有极强的隐蔽能力和针对性, 传统的安全防护系统很难防御<sup>[2]</sup>. 传统的防御手段如防火墙、VPN、防病毒、入侵检测、防水墙、内网管控、可信计算机等, 已经难以应对目前出现的威胁, 现有的技术在对抗未知威胁时出现明显的滞后性, 在这种情况下, 传统防御手段已经不能确保涉密文件的安全, 甚至在物理隔离的高等级安全网络

† 通信作者, E-mail: fenghai67@sina.com

环境下,即使加密的涉密文件也存在可能被盗取的风险。

可信计算的出现一定程度上提高了系统的安全性,特别是提高了涉密文件存取的安全性。可信计算通过对系统重要资源进行数据完整性度量,可以杜绝多数非法恶意软件的存在,从而确保系统启动或运行初始阶段是安全的。但是,即使软件运行前未被恶意代码修改,在其运行后仍然有可能遭受各类动态攻击,如利用各类 0-Day 漏洞和病毒攻击。可见,目前的可信计算方案难以保障系统的动态可信,也不能保证涉密文件的动态访问<sup>[3]265</sup>。

如何判断软件系统的动态安全可信运行和涉密文件的安全访问,是确保涉密文件不能被盗取的关键。因此,有必要在采用可信计算技术保证软件静态完整性的基础上,从软件动态行为分析入手,采取软件的动态行为可信度量技术,结合涉密文件的安全属性,确保涉密文件动态访问的安全和可信,从而增强涉密文件的抗丢失性。

本文提出一种基于可信计算动态度量技术和涉密文件水印标签技术相结合的涉密文件抗丢失技术方案,能够有效对抗针对涉密文件的攻击。

## 1 安全威胁分析

### 1.1 软件安全威胁分析

随着软件规模的不断扩大,软件的开发、集成和应用变得越来越复杂,这导致软件产品总会含有一些已知或未知的缺陷。这些缺陷不仅对软件系统的安全可靠运行构成威胁,而且对涉密文件信息的安全存储及安全访问构成严重威胁;另一方面,软件的开发和运行环境从传统静态封闭的状态变成互联网环境下动态开放的状态,越来越多的软件漏洞和缺陷被不断发现并被恶意攻击者甚至国家利益集团所利用<sup>[3]241-242</sup>。

软件安全是计算机系统安全的核心和关键,软件安全正面临前所未有的严峻挑战。为了保障软件的安全运行,目前业界采用多种安全防护手段和技术。典型的包括:数据完整性验证技术、反病毒技术、主动防御技术、主机入侵检测技术、防火墙技术、内网管控技术等。病毒或恶意代码数量的增长和攻击手段的提高,已经给目前各类传统安全防护手段带来极大挑战。目前,各安全公司日均截获的恶意代码数量已经超过 6 万,各安全公

司在恶意代码识别和特征提取上非常吃力,传统的基于特征值的恶意代码查杀技术已经很难确保用户的计算机系统安全。与此同时,在已经出现的计算机病毒中,大量病毒具备对抗和逃避反病毒软件查杀和进行防火墙穿透的功能,目前,大部分恶意程序已经开始与各类防护软件进行直接对抗,争夺对计算机终端的控制权,计算机终端面临极其严重的安全隐患,已有的网络终端防护手段受到严峻挑战<sup>[3]241-242</sup>。

### 1.2 涉密文件安全威胁分析

近两年来,病毒整体数量有所减少,但攻击的目标针对性越来越强,所造成的损失却大大增加。现实情况是,大部分涉密文件并没有经过专业的文件加密系统软件进行加密处理,经过加密处理的涉密文件仅占很少的一部分,但即使是经过加密处理的涉密文件,存储在物理隔离的高安全等级网络或主机中,仍然有被盗取的风险。光盘交换、无线注入、硬件木马、U 盘交换等等,这并不是道听途说,已经发现的 APT 攻击表明,某些黑客组织能够把高等级持续性安全威胁的恶意代码注入到物理隔离的网络中<sup>[4]</sup>,伺机盗取涉密文件,甚至针对文件加密系统软件定制攻击,在即使文件加密的情况下,仍能还原或旁路加密操作,把涉密文件明文盗取,传统的文件加密手段也受到严重威胁。

### 1.3 必要性分析

如何提高现有计算机系统的安全防护能力,确保计算机系统的动态运行安全和存储涉密信息安全,已成为当前信息安全保障工作中的关键问题。

理论上讲,采用可信计算技术的计算机软件硬件系统可以解决大多数计算机安全问题,但由于大量软件漏洞的存在,使得采用可信计算技术的可信计算机并不安全,可信计算机上运行的软件不安全;即使对涉密文件采用透明加密和数字水印技术进行加密和标签,也无法保证涉密文件不被盗取,因此,在可信计算机上采取数字水印技术加密存储的涉密文件也不安全。

软件动态行为可信技术是可信计算必须解决的一个关键性问题,数字水印技术是防止文件丢失的主要技术手段,软件动态行为分析与数字水印技术相结合,是保证操作涉密文件的软件安全、增强涉密文件抗丢失的一种有效方式。

软件动态行为可信技术可以确保操作涉密文件的软件正确运行,即软件主体的行为“总是以预期的方式,达到预期的目标”;涉密文件通过数字水印技术加密并添加涉密标签后,可以确保涉密文件的内容被合法使用,即涉密文件的内容“总是以合法的方式被阅读或使用”;采用软件动态行为可信技术与涉密文件数字水印技术的技术方案,可以确保操作涉密文件的软件总是以预期合法的方式,操作涉密文件的内容,从而提高了涉密文件的抗丢失性。

软件动态度量技术曾经是访问控制/可信计算/可信软件领域的重要研究内容,在 7 到 10 年前学术界曾投入相当大的精力予以研究,但是最终研究成果都停留在概念讨论、框架设定或特殊场景的解决方案层面,没有出现具有普适意义的实用化方法;相比之下,数字水印技术比较成熟,2 种技术结合,可以大大缩小软件动态度量的空间,软件行为集合相对较小,是可信技术在涉密文件抗丢失需求上成功的典型应用。

因此,研究软件动态行为可信与数字水印技术相结合的技术与模型,对于确保涉密文件抗丢失是十分必要的。

2 软件动态行为可信技术概述

2.1 软件动态行为可信的内涵

在确保软件数据完整性的前提下,如何确保软件的行为总是以预期的方式,朝着预期的目标运行,这就是软件动态行为可信问题。软件动态行为可信是衡量软件十分可信的重要依据,也是可信软件追求的最终目标。

目前软件之所以频频遭受攻击,引发无数安全问题,其根本原因在于:软件设计之初没有或者无法充分考虑软件自身运行环境及外来干扰的复杂性,导致软件存在设计缺陷或漏洞,在面临异常环境时无法按照预期行为轨迹运行,从而造成软件动态行为可信性受到影响。

软件动态行为可信性无法得到保证,将会导致如下问题<sup>[3]247-248</sup>:

- 1) 软件停止运行或无法完成指定的功能;
- 2) 软件(特别是重要控制软件)运行异常导致重大安全事故发生;
- 3) 软件漏洞被利用,导致软件运行时的所在系统被非法控制。

可见,确保软件动态行为可信性具有重大意义。

提高软件可信性的方法,主要有形式化方法、软件测试方法、过程管理方法以及软件监控方法等。目前,国内外的可信软件研究主要集中在软件缺陷分析、软件可信性度量与建模、可信性设计以及可信性验证、可信环境的构建与评估、可信软件的演化与控制等方面。

目前,国内外在软件行为监控方面具备一定的技术基础,在软件静态可信性认证、评估和可信模型构建方面具有一定的研究成果,但在基于软件运行实时监控的软件动态可信认证理论与技术上,还没有比较成熟的研究成果出现,且没有形成比较完善的软件动态可信认证思路和体系,特别是在对软件预期行为的提取和表征,以及软件的动态可信认证模型等方面还有待深入研究。

2.2 软件动态行为可信度量的基本思想

软件动态行为可信度量的基本思想如图 1 所示。每一个软件都有自己的正常预期行为集合,首先对软件的预期行为进行描述和抽取,获得“软件行为描述集”。之后对软件的预期行为进行摘要处理,获得软件的“软件行为特征码”。然后,将软件行为特征码与二进制程序可执行程序进行融合处理,最后发布。这样,每个软件副本都携带自身软件预期行为的描述与标识信息。为了方便软

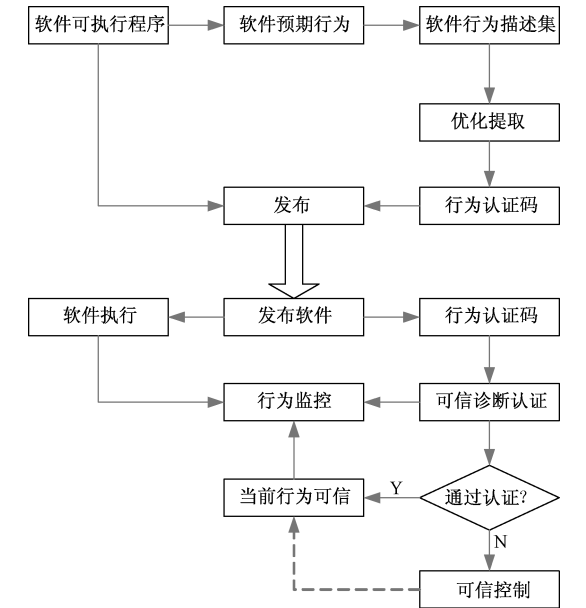


图 1 软件动态可信认证的基本思想

Fig.1 Software dynamic measurement method



件运行时准确获得软件的实际行为运行信息,在软件自身和运行系统上实施软件行为监控点监控机制.当软件实际运行时,通过设立的软件行为监控点对其实际行为进行监控,在指定的监控点获得实际软件行为.然后通过软件动态可信诊断认证中心根据软件实际行为和软件行为特征码的匹配关系,及时发现并控制软件的非预期行为,进而保障系统中的软件实时行为符合预期,确保系统的动态可信.

为准确描述软件本体,软件度量值分为软件静态度量值和软件行为动态度量值,大体的分类如图 2 所示.

1) 软件静态度量值

用于描述软件静态数据情况,确保软件运行前不被恶意代码篡改或替换.

2) 软件动态行为度量值

用于描述软件运行时作为主体,依靠其自身的功能对客体的操作或动作,一般以软件的 API 函数调用作为刻画软件行为的主要方面,尤其是一些与系统安全相关操作的 API 函数.从系统安全角度出发,在 Windows 操作系统下,一般软件行为主要包括:系统行为、文件操作、网络行为、进程行为等.

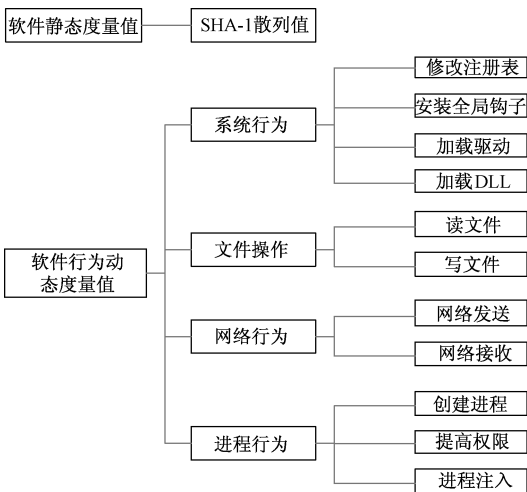


图 2 软件度量值分类

Fig. 2 Classification of software measure values

3 涉密文件数字水印技术概述

涉密文件一般通过文件加密系统进行加密处理,同时为涉密文件添加数字水印标签,配合终端或内网管控系统,防止涉密文件被盗取.

数字水印一般在操作系统的内核以文件过滤驱动程序的形式实现,下面以 Word 应用软件在 Windows 操作系统下操作 Test. doc 文件为例,说明实施加密 Test. doc 并添加数字水印的过程,来说明涉密文件数字水印技术的原理.如图 3 所示.

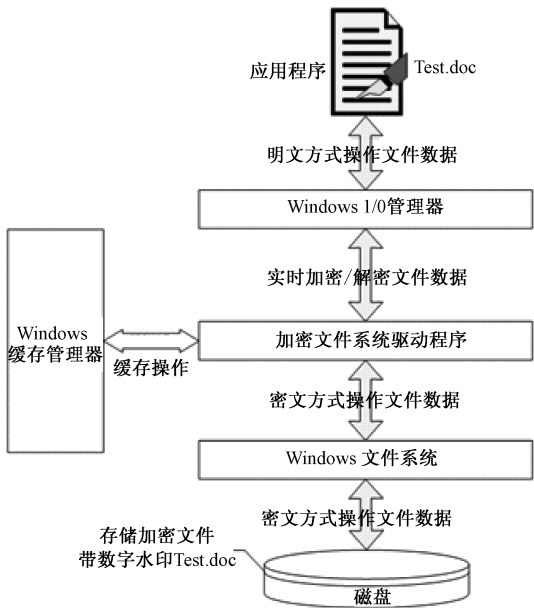


图 3 添加数字水印原理

Fig. 3 Digital watermarking method

在 Windows 操作系统下,应用层软件的所有文件操作都是向 Windows I/O 管理器发出的,再由 Windows I/O 管理器将操作定位到具体某个文件系统上,并由其完成相应的操作.应用软件发送或接收 Windows I/O 管理器的数据都是明文数据.也就是说,对用户来说,用 Word 应用软件操作加密和添加数字水印的 Test. doc 文件,与操作未加密的文件是完全一样的,没有任何区别.整个加/解密的过程对用户是完全透明的.

添加数字水印是由文件系统驱动程序实现的,能够为涉密文件提供实时的、透明的加/解密服务和添加/去掉数字水印服务.并且,文件系统驱动程序负责对 Windows 核心的缓存管理器的操作,从而实现和 Windows 系统缓存管理器协调工作,共同为文件系统提供高速、实时、透明的数字水印服务.例如,用 Word 操作 Test. doc,当用户发出存盘操作时,实际是向 Windows I/O 管理器发出写文件操作,Windows I/O 管理器发送明文数据到文件系统驱动程序,文件系统驱动程序调用加密引擎加密数据并添加数字水印,并将密文传

递给 Windows 文件系统,由 Windows 文件系统负责把密文传递给磁盘文件存储系统,所以,存储在磁盘上的 Test. doc 文件实质上是以密文的形式存储的;读操作正好与此相反. 也就是说,Windows I/O 管理器与 Windows 文件系统之间操作的数据都是密文数据. 此外,由于加密文件系统驱动程序需要与 Windows 核心的缓存管理器协调管理文件缓存,而 Windows 核心的缓存管理器无法管理密文,所以还要把文件的明文交给缓存管理器管理,这样才能实现文件系统驱动程序与 Windows 的缓存管理器协调工作.

## 4 基于可信计算的涉密文件抗丢失初步方案

### 4.1 基本思想

方案的中心思想是以添加了数字水印的涉密文件为核心,对操作涉密文件的应用程序动态行为进行度量,对操作涉密文件的应用程序行为进行深度解析,对涉密文件的读写操作进行全程监控,阻断任何针对涉密文件的非法软件行为,有效防止针对涉密文件的非法盗取.

### 4.2 初步方案

#### 1) 基于可信计算技术的涉密文件抗丢失模型

基于可信计算技术的涉密文件抗丢失模型如图 4 所示,重点对操作涉密文件的应用程序进行行为判断,对一个应用程序进程的行为判断分两阶段,一个是行为学习阶段,另一个是动态度量阶段.

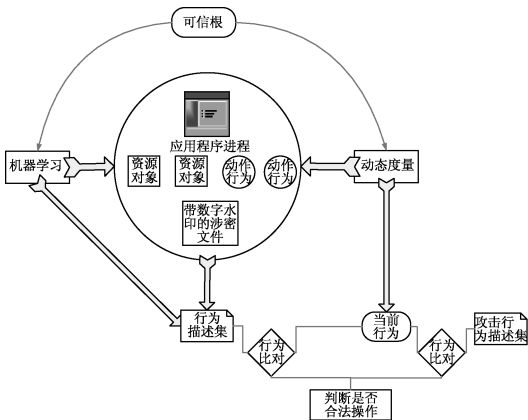


图 4 基于可信计算技术的涉密文件抗丢失模型

Fig. 4 An anti-lost scheme for confidential files based on trusted computation

应用程序进程在计算机系统内存中,通常包括若干资源对象,并通过资源对象访问信息数据,最后执行一些动作行为. 理论上,特定的应用程序进程,在特定的环境下,其行为特征是固定的.

行为学习阶段,是在一定时间内通过对特定应用程序进程通过机器学习算法进行自动分析,提炼出应用程序进程的特定行为描述集. 动态度量阶段,是在特定应用程序进程运行过程中,实时监控进程行为,并实时描述出当前行为表示,把当前的行为表示与此进程的行为描述集进行比较. 如果当前行为是行为描述集中行为;则认为当前行为是正常行为;如果当前行为不在行为描述集中,则与攻击描述行为集进行比对;如果符合攻击描述集,则认为是攻击行为;如果无法判断此行为是否具有攻击性,则提交用户界面交互处理. 攻击行为描述集是攻击经验的描述集合,是对多种攻击分析总结后的结果.

#### 2) 软件行为描述

软件的静态行为可以用度量值度量,而软件的动态行为如何表示,以及攻击行为如何描述,这是自动化抗攻击工具实现的关键. 传统的软件行为和攻击行为分析都是计算机安全专家人工分析,主观判断,判断水平取决于专家水平,而自动化的机器表示软件行为和攻击行的规范化表示语言,能为机器学习和机器分析打下坚实基础.

具体内容包括:研究软件行为的定义、分类和获取;研究软件行为的存储数据结构;研究软件行为的规范化表示;建立软件行为规则库及快速查询方法.

目前,还没有统一规范的软件行为描述语言,但具体实现中,多数以三元组表示,即

软件行为描述语言三元组:对象,动作,结果,软件行为描述集是三元组的集合.

#### 3) 软件行为评估模型及机器学习方法

在软件行为规范化定义的基础上,研究软件行为的评估模型,基本思路是采用基于层次分析方法,分层分解的软件行为信任评估模型. 软件行为是十分复杂的,软件行为分为正常行为(可信行为)和非正常行为(攻击行为或非可信行为),从 2 种行为类型的比例来看,非正常行为占用较少的比例. 重点研究软件非正常行为评估模型,描述软件非正常行为规范表示,用于机器学习方法,通过机器学习方法在一定时间内,对软件行为进

行评估和学习最终形成软件的行为模式集。

4) 软件行为动态度量方法

软件行为从时间上看,可分为短期行为和长期行为,软件的短期行为可以通过软件评估模型获得,软件的长期行为可以通过长时间机器学习获得,软件的动态行为度量方法是描述软件行为的可信程度,也就是判断软件行为是否有攻击行为。

具体实现上,需要开发软件行为监视器,以超级进程的形式对计算机系统内读取涉密文件的进程进行度量,并同软件行为模式集进行匹配,从而判断软件行为的可信程度。

5) 基于可信计算的涉密文件抗丢失原型系统

基于进程的可信和软件行为可信,实现对进程静态度量和软件行为动态度量,最终建立可信的加密文件系统原型。

原型系统的功能,首先是能够对涉密文件进行透明加密或解密,同时,为涉密文件添加数字水印,所谓的数字水印是一段结构化的数据(通常为 512 字节),结构中至少包括:数字水印唯一标示序列、文件密级、用户身份信息等,数字水印通常在文件的末尾处。

其次是对操作涉密文件的进程进行度量功能,提取操作涉密文件的应用程序的静态度量值,以及软件操作动态度量集。

再次是监控功能,监控操作涉密文件的应用程序的动态和带有数字水印的涉密文件读写操作。

最后,报警和日志功能,对非法操作涉密文件的行为,进行报警和日志记录处理,必要时可以通过人工干预判断进行处理。

4.3 关键技术实现

1) 软件行为表示

本方案重点描述操作文档文件的软件行为,主要软件包括 MicroSoft Office、PDF Reader、NotePad、WordPad、金山 WPS 等,主要文件类型为 .doc、.txt、.pdf 等。具体实现上,描述软件行为的结构如图 5 所示。

进程名称	进程路径	进程HASH值	操作对象类型	操作对象值	访问条件
------	------	---------	--------	-------	------

图 5 软件行为描述结构

Fig. 5 Structure of software action representation

```
typedef struct tag_SoftBehaviorEntry_
{
    ULONG    Index;
    CHAR
    ProcessName[ MAX_NAME_LEN ];
    CHAR
    ImageFullPath[ MAX_NAME_LEN ];
    UCHAR
    ImageSimpleHash[ 32 ];
    UCHAR    OperateType[ 32 ];
    UCHAR
    ObjectName[ MAX_NAME_LEN ];
    ULONG    AccessFlag;
}SOFTBEHAVIOR_ENTRY, * PSOFTBEHAVIOR_
ENTRY;
```

例如,描述 Word 软件的行为规则:

①文件访问

01 Word C:\Program Files\Microsoft Office\Office14\Winword. exe “hashvaluexxx” “. DOC” “test. doc”允许

②注册表访问

02 Word C:\Program Files\Microsoft Office\Office14\Winword. exe “hashvaluexxx” “REG” “HKEY \_ LOCAL \_ MACHINE \ SOFTWARE \ Microsoft”允许

③网络访问

03 Word C:\Program Files\Microsoft Office\Office14\Winword. exe “hashvaluexxx” “NET” “TCP/IP”不允许

④剪切板访问

04 Word C:\Program Files\Microsoft Office\Office14 \ Winword. exe “ hashvaluexxx ” “CLIPBOARD” “”加密等。

2) 形成规则集

刻画出软件进程的行为表示后,需要自动生成软件行为规则集合,操作涉密文件的进程,规则集中,我们重点关注其文件访问、注册表访问、网络访问、剪切板访问、移动存储访问等与文件失泄密有关的操作。具体工程实现上,采用 HookApi 技术与文件系统过滤驱动技术相配合,在操作系统的内核层和应用层分别挂接或过滤我们关注的进程操作。HookApi 技术采用开源的 EasyHook 开发

包,挂接注册表、剪切板和网络访问的所有操作,采用微软的 MiniFilter 模型,过滤文件系统和移动存储设备访问操作.

应用层开发软件规则获取程序,控制在一定时间内,获取我们关注的进程行为规则,并把规则放入本地 ACES 数据库中,目前的技术实现中,需要人工运行关注的应用程序,例如 WinWord.exe,最后需要人工校对或微调,才能形成比较精简的规则集. 规则集生成过程如图 6 所示.

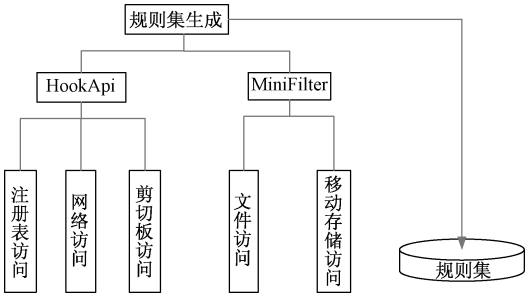


图 6 生成规则集过程

Fig. 6 Procedure of rule-set generation

3) 软件行为动态度量

软件行为动态度量,同样是采用 HookApi 技术与文件系统过滤驱动技术相配合,在操作系统的内核层和应用层分别挂接或过滤我们关注的进程操作. 只不过是运行态时,实时获取软件行为规则后,需要与规则集进行匹配,从而判断软件行为是否可信.

4) 涉密文件数字水印技术实现

涉密文件添加数字水印技术是通过文件系统过滤驱动程序,在操作系统内核通过 MiniFilter 模型实现的,文件的数字水印为 512 字节的结构,添加到文件的末尾,其结构如图 7 所示.

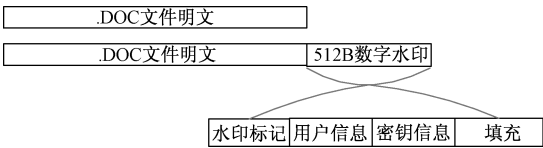


图 7 文件数字水印

Fig. 7 Watermark embedded in a digital file

```
typedef struct _FILE_WATERMARK_FLAG_  
{  
    UCHAR WaterMarkFlag[MAX_FLAG_SIZE];  
};
```

```
UCHAR    UserName[NAME_SIZE];  
ULONG    FileSecLevel;  
ULONG    CryptoAlgId;  
UCHAR    CryptoKeyLen;  
...  
UCHAR    Version;  
LARGE_INTEGER    FileSize;  
  
UCHAR    ped[1];  
} FILE_WATERMARK_FLAG,  
* P FILE_WATERMARK_FLAG;
```

通过文件系统过滤驱动,不仅实现了对指定文件类型的透明加密,也为涉密文件添加了数字水印标签,配合软件行为动态度量,可以只关注“指定进程的特定行为”与“带标签的文件”,确保非法进程不能操作涉密文件,阻止其通过网络、注册表、剪切板、移动载体等途径泄密,从而较好地实现了涉密文件的抗丢失性.

5 结束语

本文重点关注操作涉密文件的应用程序进程和涉密文件本身,相对于监控整个操作系统来说,监控空间比较小,可以在涉密文件全生命周期内,对操作涉密文件的应用程序进程进行深度行为分析,效率和准确度较高,能够阻断非法程序或恶意代码对涉密文件的盗取,并到达实用的程度. 原型系统如何进一步优化效率,并能够到达很好的涉密文件抗丢失效果,还需要进一步研究.

参考文献

[ 1 ] 赛迪网. 解读 RSA2013 热点话题:高级持续威胁( APT ) [ EB/OL ]. ( 2013-03-11 ) [ 2014-07-30 ]. [http://tech.ccidnet.com/art/1099/20130308/4784303\\_1.html](http://tech.ccidnet.com/art/1099/20130308/4784303_1.html).  
[ 2 ] 搜狗百科. 2011 年中国互联网网络安全态势报告 [ EB/OL ]. ( 2014-09-10 ) [ 2015-01-10 ]. <http://baike.sogou.com/v73515609.htm>.  
[ 3 ] 张焕国,赵波. 可信计算 [ M ]. 武汉:武汉大学出版社,2011.  
[ 4 ] eNet硅谷动力. 2013 年我国互联网网络安全态势综述: CNCERT 观点 [ EB/OL ]. ( 2014-07-28 ) [ 2014-08-10 ]. [www.cert.org.cn/2013 Network Security Situation.pdf](http://www.cert.org.cn/2013%20Network%20Security%20Situation.pdf).