基于程序行为分析的文件防篡改软件的设计与实现①

Design and Implementation of Files Tamper-Proofing Software Based on Process Behavior Analysis

吴　标　赵　方　(北京林业大学 信息学院 北京 100083)

摘　要： 实现了一个基于程序行为分析的高主动防御能力的文件防篡改软件，详细介绍了其结构和工作原理。该软件把文件操作拦截、程序行为分析、进程内存扫描、文件完整性检查 4 项技术有机结合，向被动防御模式的文件防篡改软件结构中注入主动因子，提高软件的主动防御能力。

关键词： 程序行为分析 主动防御 文件过滤驱动 文件完整性检查 文件防篡改

收稿时间:2009-02-21

1　引言

近年来，类似某机构网站的网页被改，某重要数据文件被改的事件发生频率惊人[1]，此类事件对机构的经济和形象的破坏力极强，保护文件数据势在必行，于是文件防篡改软件应运而生。目前的文件防篡改软件的功能主要有两大部分：文件操作权限控制和文件完整性检查。文件操作权限控制功能通过拦截文件操作实现，在 Windows 上有文件过滤驱动和 Hook 文件操作 API 两个技术选择。文件完整性检查一般采用密码水印技术，此加密过程会产生较高的资源消耗[2]。

从防范模式的角度来看，文件防篡改软件中的文件操作权限控制部分属于被动防御性质，文件完整性检查部分属于主动防御性质，但由于其高运行消耗的原因，所以此部分只在有限的情况下启动，如在一个文件被读取前，因而此部分提供的主动防御能力就很有限。所以综合来看，目前的文件防篡改软件侧重于被动防御模式，并没有重点采用主动防御技术[3]，主动防御能力比较欠缺。实际上，被动防御模式的文件防篡改软件已经可以达到的很高安全防范程度了，但是并非完全没有疏漏。比如，属于被动防御模式的文件过滤驱动拦截文件操作可以同样地被文件过滤驱动突破，Hook 文件操作 API 更是容易被突破。当被动防御这一层被突破的时候，如果不采取主动防御方法主动地追踪篡改的根源，被篡改的文件就在毫不知情的情况下被传播出去，并且永远不会发现已经发生的篡改行为。

为了弥补被动防御模式的文件防篡改软件的不足，本文实现了一个基于程序行为分析的文件防篡改软件，该软件把文件操作拦截、程序行为分析、进程内存扫描、文件完整性检查 4 项技术有机结合，提高软件的主动防御能力。

2　软件设计

2.1 总体结构

本软件的大体工作原理：

① 先添加文件为本软件的保护目标文件，设置进程文件操作权限；

② 采用文件过滤驱动技术进行拦截文件操作，在允许文件操作进行前先检查完整性，当发现文件被篡改或者非授权访问后先对该进程进行非授权文件路径字符串引用扫描，再进行全局的非授权文件路径字符串引用扫描和全局文件完整性检查；

③ 把进程的特征系统调用作为程序行为分析的根据，建立可疑程序行为特征库，记录进程的特征系

① 基金项目:国家科技支撑项目(2006BAD10A03)

统调用并检查其是否匹配可疑程序行为特征库中的记录，发现可疑程序行为后对该进程进行非授权文件路径字符串引用扫描，再对被非授权引用文件路径的文件进行完整性检查。

④ 定时进行非授权文件路径字符串引用扫描、文件完整性检查。

本软件的工作原理如图 1 所示，虚线框内为本软件的运行代码，分为主程序、文件过滤的驱动程序、Hook 特征系统调用的 DLL 程序 3 部分，主程序在一个独立的进程中运行，文件过滤的驱动程序在共享的系统空间中运行，通过 DeviceIoControl 函数、事件、核心态与用户态共享内存和主程序通信，Hook 特征系统调用的 DLL 程序在调用者进程的用户空间内运行，通过 Windows Sockets API 和主程序通信。

本软件结构的软件结构如图 1 所示：

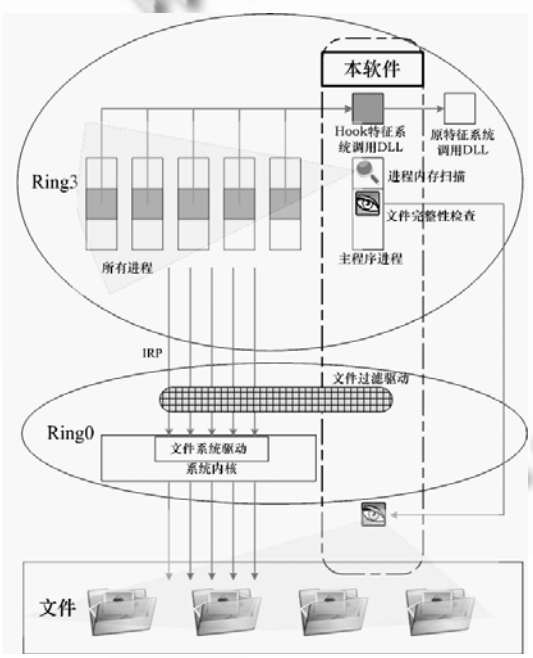

图 1　工作原理示意图

2.2 核心模块设计

2.2.1 文件操作控制模块

本模块功能是拦截对保护目标文件的操作请求并进行相应处理。在允许文件操作进行前先对此文件进行完整性检查，防止被篡改的文件传播出去。如果发现此文件被篡改则意味着软件的被动防御层已经被突破，其他文件可能也被篡改，所以此时应该主动出击，搜寻破坏进程和其他被篡改的文件，调用扫描模块的一般扫描子模块进行全局的非授权文件路径字符串引用扫描和全局的文件完整性检查。如果文件未被篡改接着进行非授权文件访问检查，如是非授权访问文件，则此进程有可能是破坏进程，提示用户发现非授权访问文件并询问用户是否授予该进程对此文件的此操作权限并允许此次访问通过。因为破坏进程一般不只篡改一个文件，所以用户选择否就调用非授权访问文件进程扫描子模块对该进程进一步扫描，发现其他的破坏行为并总结其程序行为特征入库。再调用扫描模块的一般扫描子模块进行全局的非授权文件路径字符串引用扫描和全局的文件完整性检查。此模块的流程如图 2 所示。

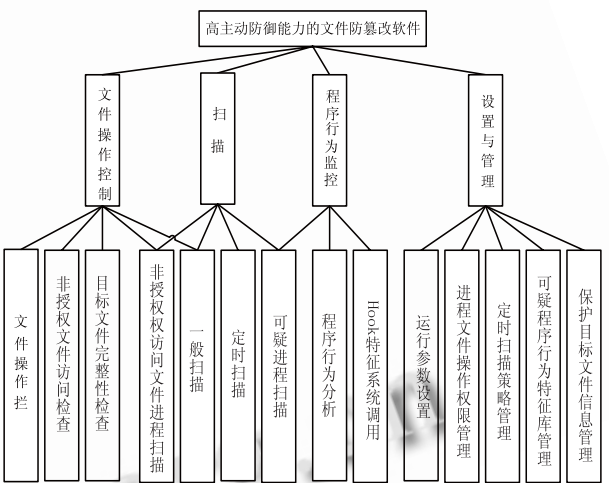

图 2　软件结构图

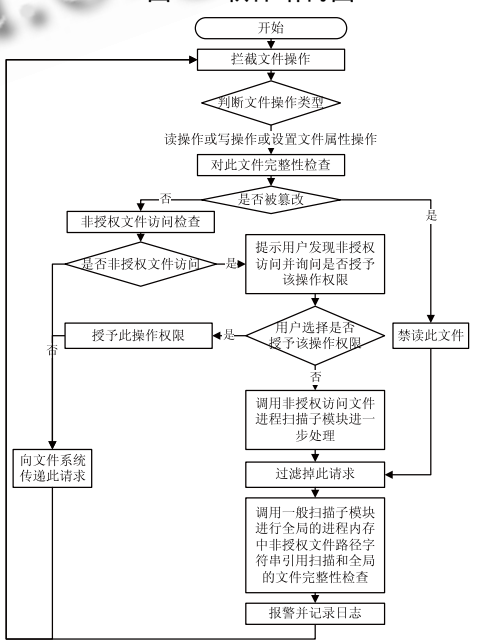

图 3　文件操作拦截模块的流程图

进程权限检查子模块

根据设定的进程的文件操作权限判断当前进程的文件操作是否被授权。

2.2.2 扫描模块

扫描模块按照 4 种方式（非授权访问文件进程扫描、可疑进程扫描、一般扫描、定时扫描）启动扫描，是软件主动防御能力的体现点。前 3 种方式用于提供给文件操作控制模块和程序行为监控模块调用。

① 非授权访问文件进程扫描子模块

本模块在文件操作控制模块发现非授权访问时被调用，功能是对非授权访问进程进一步扫描，发现其他的破坏行为并总结其程序行为特征入库。因为破坏进程一般不只篡改一个文件，所以如果是非授权的进程访问文件的情况还要先调用非授权文件路径引用扫描子模块搜寻进程内存中的非授权访问的文件路径字符串，并对找到的所有非授权访问的文件路径进行文件完整性检查，从而发现其可能篡改的所有文件，再调用程序行为分析模块的总结程序行为子模块把 Hook 特征系统调用子模块所记录的此进程的特征系统调用记录序列总结成一个可疑程序行为特征记录并储存到可疑程序行为特征库，为以后的可疑的程序行为的判断提供更多依据。

② 扫描可疑进程子模块

本模块用于提供给程序行为监控模块发现可疑进程时调用，功能是调用非授权文件路径引用扫描子模块扫描可疑进程。

③ 一般扫描子模块

本模块功能是进行全局的非授权文件路径字符串引用扫描和全局文件完整性检查。两部分同时开始，在独立的线程中运行。

④ 定时扫描子模块

本模块分别进行定时的全局的非授权文件路径字符串引用扫描和全局文件完整性检查。

⑤ 非授权文件路径引用扫描子模块

本模块功能搜寻进程内存中的所有保护目标文件路径字符串和用户自定义的敏感关键字，并调用进程权限检查模块排除授权访问的文件路径，得到非授权访问的文件路径，并逐一进行文件完整性检查。

⑥ 全局文件完整性检查子模块

本模块功能为预先生成保护目标文件的报文摘要并储存，进行检查时生成保护目标文件当前的报文摘要与原始正确的报文摘要比较，如果不相同则说明文件被篡改。文件的报文摘要的生成方法为：把文件内容加上文件最后修改时间，再加上文件属性，用 MD5 算法加密生成[4]。

⑦ 进程内存扫描子模块

本模块的功能是扫描进程的内存中是否包含保护目标文件名和用户自定义的敏感关键字。

2.2.3 程序行为监控模块

本模块旨在实现有针对性地主动出击，检测并排除风险。其工作原理为：调用 Hook 特征系统调用子模块 Hook 特征系统调用，记录下进程调用这些 API 的痕迹，再调用程序行为分析子模块统计判断其是否符合可疑程序行为特征库的记录，如符合则调用扫描模块的扫描可疑进程子模块扫描此进程。此模块流程图如图 4 所示。

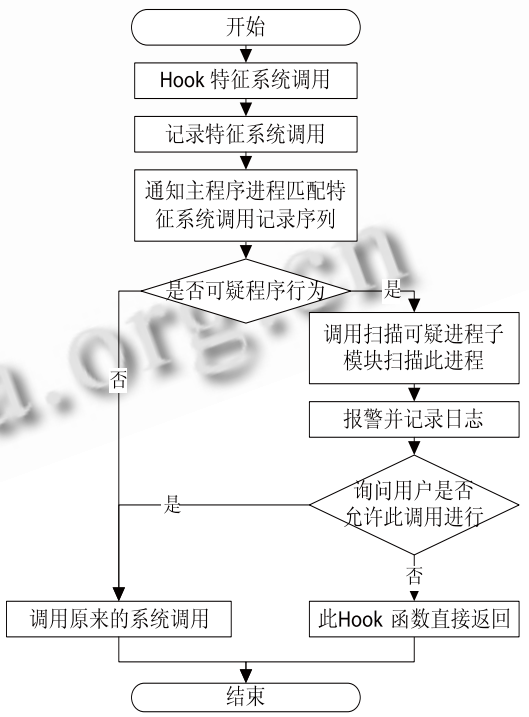

图 4　程序行为监控模块流程图

① Hook 特征系统调用子模块

本模块的功能是 Hook 特征系统调用，当进程调用特征系统调用时，其对应的 Hook 例程先记录此次调用，再通知上层的程序行为监控模块对该进程进行

分析。每个特征系统调用记录包括进程 ID、函数名、调用时间。本软件监控的特征系统调用见表 1。

表 1　Hook 特征系统调用列表[5]

类型	函数名
文件操作	WriteFile、ReadFile、SetFileAttributes、GetFileAttributes、CopyFile、CreateFile、MoveFile
内存操作	memcpy、memmove、memset、memcmp、ReadProcessMemory、WriteProcessMemory
进程线程操作	OpenProcess、CreateProcess、TerminateProcess、ShellExecute、WinExec、LookupPrivilegeValue、AdjustTokenPrivileges、CreateRemoteThread
钩子	SetWindowsHookEx、UnHookWindowsHookEx、CallNextHookEx
驱动服务操作	OpenSCManager、CreateService、OpenService、StartService、ControlService、DeviceIoControl
网络通信	Socket、listen、send、recv、sendto、recvfrom
DLL 力劳动操作	LoadLibrary、LoadLibraryEx、GetProcAddress

② 程序行为分析子模块

本模块有 2 个功能：匹配程序行为——按照可疑程序行为特征库判断指定特征系统调用记录序列是否为可疑程序行为；总结程序行为——把指定进程的特征系统调用记录序列总结成一个可疑程序行为特征记录并入库。前者用于提供给上层的程序行为分析模块调用，后者用于提供给非授权访问文件进程扫描子模块调用。

特征系统调用记录和可疑程序行为特征库记录的数据定义如下：

名字：特征系统调用记录
描述：记录进程一次特征系统调用的基本信息
定义：特征系统调用记录 = 进程 ID+,+函数名+,+调用时间
位置：

名字：可疑程序行为特征库记录
描述：描述可疑程序的特征系统调用特征的表达式
定义：可疑程序行为特征库记录 = 组合函数调用频率描述单元|函数调用频率描述单元 +0{关系符+组合函数调用频率描述单元|函数调用频率描述单元}100+!+信任进程列表
信任进程列表 = (+进程名+0{,+进程名}100+)
组合函数调用频率描述单元 = <+组合函数调用描述+,+最长时间+,+最少次数+>
组合函数调用描述 = (+函数名+0{,+函数名}100+)
函数调用频率描述单元 = (+函数名+,+最长时间+,+最少次数+)
关系 = 并且|或者|时间先后
并且 = &
或者 = |
时间先后 = -
位置：可疑程序行为特征库

以下列出 4 条通用的可疑程序行为特征记录，用户可以调整其中数字和信任进程列表：

• 频繁的文件操作行为，30 分钟 10 次以上为频繁。其表达式为：(WriteFile,30,10)| (ReadFile,30,10)|(SetFileAttributes,30,10)|(GetFileAttributes, 30,10) | (CopyFile,30,10)|(CreateFile,30,10)|(MoveFile,30,10)!(IE,System,Explorer,inetinfo,svchost,services,winlogin,csrss,smss)

• 表 1 的内存操作函数中 ReadProcessMemory、Write zProcessMemory、OpenProcess、进程线程操作函数的全部、钩子函数的全部、驱动服务操作函数的出现 1 次。

• 先读文件再复制内存再写文件交叉进行，10 分钟内 4 次以上。其表达式为：<(ReadFile,memcpy,WriteFile),10,4>!(IE,System,Explorer,inetinfo,svchost,services,winlogin,csrss,smss)

• 频繁的网络活动和文件操作，120 分钟文件操作 10 次以上，网络活动 50 次以上。

(下转第 134 页)

(上接第 117 页)

表达式的匹配过程并不复杂，需要注意的是：密匹配疏——表达式中的最长时间和最少次数组合起来表示在小于等于最长时间内，特征系统调用发生的大于等于最少次数；子匹配母——时间先后关系是并且关系的子集；多匹配少——匹配(WriteFile, 30,10)&(send,30,10)的内容肯定匹配(WriteFile,30,10)。

把指定进程的特征系统调用记录序列总结成可疑程序行为特征记录时，函数调用频率描述单元中的最长时间和最小次数也要进行一定比率的变化，最长时间等于所用时间的 4/3，最小次数等于调用次数的 2/3(用户可调整此数字)。整个总结过程分为 3 步：

- 找出重复的调用组合
- 累加所用时间和调用次数
- 换算最长时间和最小次数

3 结语

文中提出的软件提高了文件防篡改软件的主动防御能力，同时本软件也存在以下不足：程序行为监控模块对特征系统调用记录序列的分析方法和可疑程序行为特征库的设计比较粗糙，尚需改进。

参考文献

1 纪玉春.我国大陆地区九千多网站被篡改.信息网络安全, 2008,(5):72－73.
2 盖玲.防网页篡改技术比较分析.图书与情报, 2007,1:92－94.
3 王刚.您的网站守护神:鹰眼主页防篡改系统.信息安全与通信保密, 2005,(6):136－137.
4 唐三平.基于散列函数的数字签名.信息网络安全,2005,(2):75－76.
5 罗亚丽,周安民,吴少华,等.一种基于行为分析的程序异常检测方法.计算机应用,2008,28(10):2492－2494.