

一种对 Windows 中 PE 文件进行启发式病毒扫描的算法

何志永

(天津开发区职业技术学院, 天津 300457)

摘要: 计算机病毒给千万个用户带来了不可估量的损失, 杀毒软件可以对计算机进行保护, 有效地防止病毒的破坏, 但对未知病毒的防护能力较差。本文根据分析已知类病毒的行为和特征, 总结一些病毒的特点, 来启发未知病毒, 提出一种能够识别 PE 文件的启发式病毒扫描算法, 给出了设计的伪代码, 并模拟算法对一些文件进行了试验, 得到了良好效果。

关键词: 病毒; 规则; 算法; PE 文件; 启发式

中图分类号: TP311.56

文献标识码: A

文章编号: 1001-7119(2013)05-0147-04

A Heuristic Virus Scan Algorithm for Windows PE File

He Zhiyong

(TEDA Polytechnic, Tianjin 300457, China)

Abstract: A computer virus brings great loss to users. In order to prevent the destruction of the virus effectively, an antivirus software is often used for protecting, but the protection ability is not so efficient for the unknown virus. According to the analysis of known virus behavior and characteristics, this paper summarized the characteristics of some viruses to inspire unknown viruses. The paper puts forward a heuristic virus scanning algorithm to identify PE file and gives the design of the pseudo code. A series of experiments have been made to the simulation algorithm for some document, and good effect has been made.

Key words: virus; rule; algorithm; PE file; heuristic

0 引言

随着计算机信息技术的不断发展, 计算机已经成为人们的必需品, 各行各业都广泛使用, 自计算机病毒诞生以来, 给千千万万的用户带来了无法估计的损失, 如今病毒的肆虐已经达到了无孔不入的地步。Windows NT 3.1 引入了一种名为 PE 文件格式的新可执行文件格式, 此文件格式是跨 Windows 平台的, 即使 Windows 运行在非 Intel 的 CPU 上, PE 装载器都能识别并使用这种文件格式。例如, 我们平时在 Windows 下使用的 .exe 文件就是典型的 PE 类型文件, PE 文件是计

算机中非常重要, 最常用的文件, 到现在 Windows 平台下绝大部分可执行文件的结构都是 PE 文件格式。PE 文件被组织成一个线性的数据流, 这些文件的组织格式基本上是一样的, 包括文件标志、文件头和可选头部等^[1], 这对我们对病毒的分析 and 扫描很重要。

1 启发式病毒扫描特点

病毒和正常程序一样就是一段程序代码, 但是在很多方面是有区别的, 例如, 病毒具有自我复制的能力。而这些区别对于一个普通的用户来

收稿日期: 2013-01-10

作者简介: 何志永(1982-), 男, 天津宝坻人, 汉族, 硕士, 天津开发区职业技术学院计算机系, 讲师, 研究方向: 计算机软件编程、手机应用程序开发。

说是看不见的,但对于一个熟练的病毒分析程序员来说并不是很困难。启发式病毒扫描技术实际上就是程序员把自己的技术经验以及相关知识,运用到病毒分析软件中,使这种软件具有“根据现象和特征判断并发现病毒”的能力。实际上就是以特定方式实现的动态反编译器,它可以通过对程序相关的指令序列进行反编译,从而逐步分析并确定其真正的功能特征,并判断是否为病毒。传统的病毒扫描技术,是通过扫描磁盘中的每一个文件进行扫描,和病毒库中的数据文件进行比较,从而发现计算机中已感染的病毒,并作相关处理,这种技术对未知名的病毒毫无办法,虽然误报的可能性比较小,但漏报的可能性非常大。而启发式病毒扫描技术正是弥补了这一缺点,不至于使未知名的病毒毫无被发现的可能^[2]。

因此,启发式病毒扫描技术,采用的是主动防御技术,不仅可以检测出已知名的病毒,还可以检测出未知名的病毒。将传统的病毒扫描技术和启发式病毒扫描技术的优势相结合,是一种提高病毒扫描技术的选择。

2 算法描述

根据以上启发式病毒扫描的特点,结合 PE 文件的特点,提出以下针对 PE 文件的启发式病毒扫描算法,我们根据以下规则来实现算法:

规则 0:入口是否在最后一节。通常程序的入口点都不会是在最后,通常病毒会把自己附加在最后一节,如果一个程序的入口点在最后一节,它就很可能是病毒。

规则 1:是否有两个节名不是常见的。如果病毒附加到正常程序后添加了新的节,节的名字都比较特殊。

规则 2:最后一节不是常见的名字。如果没有名字的节是最后一节,那就很可能是病毒。

规则 3:已知的节存储了不是对应的节的内容。如果病毒做的不完善的话,它的节的信息和名字不相符,例如,代码节里放了数据节的内容。资源节里放了要执行的代码,这样可疑性就很大。

规则 4:引出表的里面的引入表非常少。这可能是程序被加壳造成的。

规则 5:隐藏了 API 的调用。病毒程序为了防

止被他人分析,通常会把自己调用的 API 给隐藏了。

规则 6:E8 00 00 00 00 在入口附近。病毒附加到正常程序后,必须重定位才能执行。E8 00 00 00 00 正好是 Call \$+5,用来重定位的。

规则 7:文件大小无缘无故的变大了。如果病毒没处理好的话,文件头存储的文件的大小和实际的大小会不一致。

规则 8:杂乱的内容或程序在头文件的头结构。有的病毒会把感染代码加到 section table, PE header。

规则 9:所有节的大小和 PE header 里记录的不一致。

规则 10:最后的节的大小和节表 SizeOfImage 域中的内容不匹配。一般病毒会把自己附加到最后一节,如果它没有改节表里的 SizeOfImage,那么就不会出现不一致。

规则 11:用数字调用 Kernal32 中的 API 函数。病毒程序为了防止被他人分析,用数字调用自己引用的 API。

规则 12:文件的校验和为 0。有时病毒用它来做感染标记。

规则 13:存在有危险倾向的 API 调用。一些 API 正常程序是不会调用的,例如:添加注册表启动项用的 RegSetValueExA, 创建远程线程用的 CreateRemoteThread, 添加服务用的 CreateServiceA, 打开文件镜像用的 CreateFileMappingA, 还有网络下载、创建互斥量、释放文件等。

以上每条规则可以根据它的重要性来给它一个权值,如果总的权值达到了警戒线,那它就是可疑文件了。

3 实现过程详细设计

以下对算法实现过程使用伪代码的方式进行简单描述。

3.1 判断 PE 文件结构模块详细设计

伪代码如下:

(1)先将文件的头结构读入缓冲区。

```
fread(x,sizeof(int),500,fp)
for(i=0;i<300;i++)
{
```

```
Convert32to8(x,y,i,i*4);
```

```
}
```

(2)判断文件开始是否有 DOS 头标志“MZ”。

```
if((y[0]!=0x4d)&&(y[1]!=0x5a))
```

(3)判断 PE header 开始是否有 PE 文件标志“PE”。

```
if((y[x[0x3c/4]]!=0x50)&&(y[x[0x3c/4]+1]!=0x45))
```

如果是 PE 文件则返回 1, 否则返回 0。

3.2 启发式规则扫描模块详细设计

这个模块是对以上启发式病毒扫描算法的具体实施过程描述, 它将具体判断文件的信息和属性是否符合该算法的各种规则, 其伪代码如下:

(1)将文件节表读入缓冲区, 然后取出节属性, 判断是否可写, 如果可写, 则把权值累加上。

```
if(writable != ((Characteristics>>31)&1))
```

```
test=1;
```

```
if((test !=1)&&(i!=(SectionNumber-2)))
```

```
score=score+5;
```

(2)将文件头读入缓冲区, 和各节的偏移比较, 看它是否在最后一节, 如果在最后一节, 就把它的权值累加上^[3]。

```
Deviation=AddressofEntryPoint—
```

```
NextVirtualAddress;
```

```
FileEntryPoint= Deviation+x [x [0x3c/4]/4+0xf8/4+0x28/4*(i+1)+0x14/4];
```

```
Characteristics=x [x [0x3c/4]+0xf8/4+0x28/4*(i+1)+0x24/4];
```

```
if(writable != ((Characteristics>>31)&1))
```

```
score=score+3;
```

(3)将入口附近代码读入缓冲区, 看 E8 00 00 00 00 是否在入口附近, 如果有此段代码, 就把它的权值累加上。

```
int value [5]= {0xE8, 0x 00, 0x 00, 0x 00, 0x 00};
```

```
for(i=0 ;i<400 ;i++)
```

```
{ for(j=i ,k=0 ;k<5 ;j++)
```

```
{ if(n[j]!=value[k++])
```

```
{ k--;
```

```
break;
```

```
}
```

```
}
```

```
if(k==5)
```

```
{ score=score+10;
```

```
break;
```

```
}
```

```
}
```

(4)取出存储校验和字节的内容, 校验和是否为零, 如果为零, 就把它的权值累加上。

```
int checksum=x[x[0x3c/4]/4+0x58/4];
```

```
if(checksum!=0)
```

```
score=score+3;
```

(5)读引入表到缓冲区, 读入 IMAGE_IMPORT_BY_NAME 的内容, 判断是否用数字从 KERNEL32.DLL 调用引入函数, 如果是, 就把它的权值累加上。

```
if(!strcmp(dllname,dll))
```

```
score=score+2;
```

(6)统计引入函数的个数, 判断文件是否加壳^[4], 如果是, 就把它的权值累加上。

```
if(u<5)
```

```
score=score+5;
```

(7)判断权值 score 是否超过了警戒线, 如果超过了, 就发出警报。

4 模拟软件实现过程描述

应用以上算法, 我们想分析一些病毒来检验算法, 在实现算法过程中, 我们模拟的步骤如下:

首先, 将检验的文件按顺序读入, 之后通过读文件头是否有“MZ”, 来判断是否是 PE 文件, 是则进行下一部分分析, 不是的不进行分析, 并记录 PE 文件是否分析准确。

之后, 若是 PE 文件则把文件读入缓冲区, 进行判断, 对于上述算法的每一个规则都定制一个权值, 若违反了规则, 则给加上一个权值。

最后, 把每个文件所得到的分权值和警戒线权值相比较, 若超过警戒线, 则说明是病毒。

现将对 500 个文件实验的结果统计如下:

表 1 试验结果统计表

Table 1 Experiment statistical table)

文件类别	个数	识别数	识别率/%
非 PE 文件	85	85	100
病毒	280	278	99.29

通过试验, 我们基本可以看出, 对与 PE 文

件,基本格式是稳定的,本文提出的病毒扫描算法识别率还是比较高的。

5 结论

尽管杀毒软件越来越先进,功能也越来越强大,但计算机病的种类越来越多,更新换代愈发频繁,自身的隐藏蒙蔽用户的手段也越来越多,不断有新的病毒出现,使我们防不胜防。除了平时要树立安全防范意识外,非常重要的一点就是主动防御病毒的入侵。通过研究发现,本文提出的启发式病毒扫描算法既可以识别已知病毒,也可以防御未知病毒。算法基本可靠,但若要变成病毒扫描软件,并要清除病毒,还需要很多技术,到推广还有很多难题要攻克。在病毒越来越猖獗的时代,我们需要好的算法,准确、主动地去防范病毒,尽量避免误报、虚报和谎报,这需要我们进一步了解病毒的行为特征^[9]。

启发式病毒扫描技术给我们提供了一套新的思路,它可以产生一种新的打破传统观念的病毒扫描软件,我们需要的是能够充分的结合病毒的各种行为和特征,减少虚报和谎报的概率的好的启发式病毒扫描算法,这样就不需要我们不断地去升级病毒库,这样既能够给用户减少不必要的麻烦,又能给整个社会来经济效益。

参考文献:

- [1] 纪宏亮, 浅析网络病毒的检测与防治 [J]. 农村电气化, 1999,07.
- [2] 鄢海霞,张西臣. 计算机病毒的防治[J].网络科技时代, 2005,8.
- [3] Carey Nachenberg.Computer Virus Coevolution[J].COMMUNICATIONS of the ACM, 1997,40.
- [4] Derek Uluski,Micha Moffie,David Kaeli. ACM SIGARCH [J]. Computer Architecture News, 2005,3.
- [5] 卓新建,郑康锋,辛阳.计算机病毒原理与防治[M]. 北京: 北京邮电大学出版社, 2007.

(上接第110页)

载荷冲击大,载荷性质一般属于交变重载,一般采用工业闭式齿轮油、硫-磷型极压工业齿轮油,在具体使用过程中要根据需要合理选择。

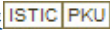
6 结语

通过对某矿的减速器齿轮在选材、齿轮参数设计、热处理工艺、机械加工、润滑等方面采取以上措施进行试验,使该矿减速器齿轮在有限的空间内,满足大工作载荷的要求,确保了减速器的正常工作,延长了减速器的使用寿命,提高了齿轮工作的可靠性,使其在强度、硬度、韧性、耐磨性和承载能力等方面充分满足煤矿采掘机减速器的要求。

参考文献:

- [1] 齿轮手册编委会.齿轮手册(上、下册)(第二版)[M].北京:机械工业出版社, 2000.
- [2] 赵建军.提高煤矿机械渐开线齿轮减速器综合性能探讨[J].山西煤炭管理干部学院学报, 2006,1:87-89.
- [3] 潘维忠,张秀清.提高采掘机械齿轮强度问题的探讨[J].煤炭技术, 2007,3:120-122.
- [4] Nobuyoshi Yoshida,Tokihiko Taki.Micropitting generation mechanism for gears [J].International Journal of Autonmation Technology[J].2008,5:341-347.
- [5] 胡延平. 煤矿机械传动齿轮失效形式分析及改进措施[J].江西煤炭科技, 2010,3:105-107.
- [6] 白树全,高美兰,王红.采煤机重载齿轮的制造及其热处理工艺[J].铸造技术, 2012,4:412-413.
- [7] 马跃林,一种新型输送机减速器[J].煤炭技术, 2012,6:12-13.

一种对Windows中PE文件进行启发式病毒扫描的算法

作者：[何志永](#), [He Zhiyong](#)
作者单位：[天津开发区职业技术学院, 天津, 300457](#)
刊名：[科技通报](#) 
英文刊名：[Bulletin of Science and Technology](#)
年, 卷(期): 2013, 29(5)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_kjtb201305035.aspx