基于特征码技术的攻防策略①

Tactics of Attack and Defence Based on Characteristic Code Technology

张迎春　(上海电子信息学院 计算机应用系 上海 201411)

摘　要：本文深入讨论了特征码技术在病毒攻与防中的运用策略及其发展趋势。实验中针对不同类型的病毒特征码进行手动定位、分析和验证，并通过修改、加壳和加密等技术手段达到免杀效果。如果将单一的特征码技术和主动防御相结合，就能在攻与防的对抗中取得更大的突破。

关键词：病毒 特征码 定位 多态

当前单纯意义上的病毒已逐步被木马、蠕虫所代替。操作系统的升级为后门病毒大规模的破坏提供了便利，并且从自启动发展为注册系统服务，从单进程发展为守护进程和远程线程注入，甚至采用驱动技术来隐藏后门程序，让用户手动查找愈加困难。在这种局势下，各种新技术和杀软被不断开发出来，和病毒进行着没有硝烟的较量。

1　引言

所谓特征码，就是防毒软件从病毒样本中提取的不超过 64 字节且能独一无二地代表病毒特征的十六进制代码。主要有单一特征码、多重特征码和复合特征码这三种类型。特征码提取的思路是：首先获取一个病毒程序的长度，根据样本长度可将文件分为若干份（分段的方法在很大程度上避免了采用单一特征码误报病毒现象的发生，也可以避免特征码过于集中造成的误报），每份选取 16B 或 32B 的特征串，若该信息是通用信息或者全零字节则舍弃，认为或随机调整偏移量后重新选取。最后，将选取出来的几段特征码及它们的偏移量存入病毒库，标示出病毒的名称即可。根据这个思路可编写出特征码提取程序实现自动提取，并保存病毒记录。在扫描病毒时，防毒软件将目标文件通过模式匹配算法与病毒库中的特征码进行比对，以确定是否染毒[1]。

2　特征码的检测与处理

2.1 特征码的定位

单一特征码扫描，就是从病毒样本中提取连续的能标示此病毒的若干个字节。其好处在于开销小，便于升级和维护病毒库。但这种技术容易导致误查误杀，目前已较少使用。

对于多重特征码，可在单一特征码扫描的基础上进一步提取不连续的若干段特征码，仅当待检测文件完全符合这多段特征码时才报警。这样可以减少误杀率，提高查杀的准确度，因此成为多数防毒软件的首选技术。用“逐字节替换法”可手动定位多重特征码。把目标木马服务端或病毒逐字节地替换为 00h 或 ffh（其他亦可），每替换一次存为一个文件，然后对生成的几份文件杀毒，未被删除的就是被修改了特征码的文件。汇总被修改的字节就得到了杀软对该木马或病毒所定义的“特征码”。然而，手工操作量和占用空间都过于庞大，可用分段法加以改进，即逐步缩小特征码所在范围。实验中选取某防毒软件对黑客工具进行特征码定位。首先以 128B 为替换单位，从查杀后的文件可知特征码的偏移和范围，之后还原代码，再以 32B 为单位对该范围进行替换并查杀，最后使用逐字节替换定位出连续的特征码字节。这样每次仅需几兆的空间，且速度很快。整个由粗略到精细的定位结果如图 1~图 3 所示。至此，多重特征码定位成功，只需任选一段特征码来定位，修改后就可以逃过查杀。

① 收稿时间:2008-08-28

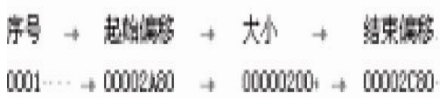

图 1　128B 的粗略定位

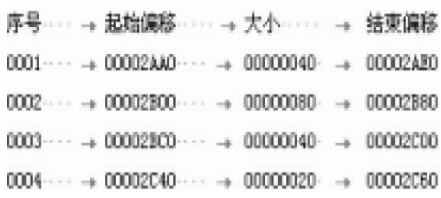

图 2　32B 较粗略定位 2A80-2C80 区段

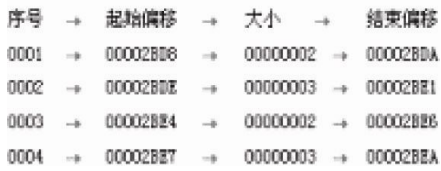

图 3　1B 精细定位 2BC0-2C00 区段

扫描复合特征码时，首先仍从病毒样本中提取多段特征码(假定为 N 段)[2]。与多重特征码不同的是，它不需要完全符合所有多段特征码才查杀病毒，而是选取一个种子 M(一般取 M=2)，当符合这多段中的任意 M 段才查杀病毒。假设对某样本抽取了如图 4 的 a、b、c、d、e 五个特征码片段，其中任何一个都无法独立标识特征码。针对这种特殊情况，考虑到 PE 头的重要性，可从尾端开始往前盖 0。直到如图 5 所示时，杀软才不能识别(b 片段被破坏了一个字节)，如此可定位出 b 的尾端。接着在 b 段中如图 6 所示从尾至头移 0，a、b 两段又完整了，符合复合特征码，可被查杀。定位出 b 段后先用 0 填充为新的样本，如果仍被杀就重复上述步骤定出 c 段。同样的方法继续定 c、d、e 段，直到不再被识别出来。

图 4　对某样本抽取的五个特征码片段

图 5　定位出 b 片段尾端

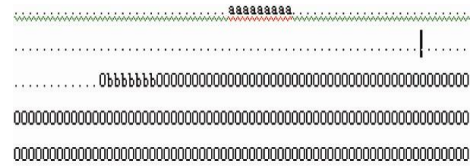

图 6　b 段已被定位出

以上思路可通过编程自动实现。为提高效率，考虑先用二分法粗定，到范围缩小时再逐字节替换。实验中考虑到 PE 头部的重要性，因此选择从尾端开始以 2n(2n<文件尺寸)的大小往前盖，一次生成约 20 个样本。检测后发现，盖了 128B 的还能被识别(如图 7)，而盖了 256B 的则不被识别了，这说明距文件末尾 256—128B 之间有个特征码片段。接下来就以图 7 为样本，以图 8 中经 X 标记的区域为定位范围。如此重复，当范围缩至<=32B 时再改用逐字节替换法，一次生成最多 32 个文件。

图 7　128B

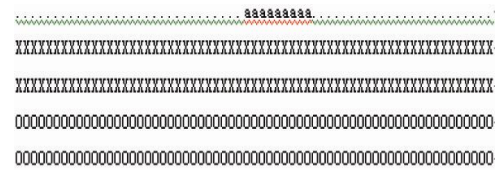

图 8　只对 X 区域中的特征码进行定位

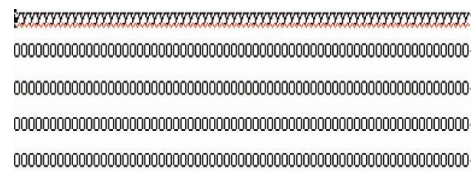

图 9　只对 Y 区域中的特征码进行定位

存在另一种可能：某次用二分法生成的所有文件都被认出，说明特征码集中在最大盖 0 范围前，这时可以图 5 为样本定位图 9 中的 y 区域。定出尾端后同样地前移 0，一次生成 32 个文件，若全不被杀就取这 32B 为定位结果。对于这种大片段无须完全定位出头端，因为一个片段中只要修改一个字节就足够。

在进行复合特征码扫描时，若提取 N 段特征码，

那么将会排列组合成(N-1)！种方案。种子 M 的精度越高，则误报率越低，反之，可提高对变种病毒的识别能力。因此，这种新的特征码扫描技术比较灵活，已被当前的防毒软件所采用。但是该技术将会增大病毒库的容量。

2.2 特征码的修改

对定位后的特征码进行修改便能逃过查杀。对可执行文件，要根据汇编代码来修改特征码，首先进行反汇编，使用调试器(如 Ollydbg)调试程序，并根据特征码的文件偏移地址转换成的虚拟地址找到汇编指令。修改方法主要有：修改字符串大小写法、等价替换法、指令顺序调换法、通用跳转法。现在许多病毒采用自动变形技术来逃避特征码检测，即所谓的多动态病毒，它在外观形态上没有固定的特征码[1]。

病毒的多态致使对其代码段的加密能完全改变原有特征码，因此需在零区域加入解密代码来解密，然后使用 JMP 指令跳回原指令代码执行，由于对某一字节执行 XOR 两次后可还原代码，因此可以手动加密代码，下次病毒执行时，可以解密代码并执行。致使特征码完全变样，达到免杀[4]。代码如下所示：

```
mov eax,AAAAAAAA        /*加密代码新入口点，AAAAAAAA 为代码段首地址*/
xor byte ptr ds:[eax],XX    /*XX 为随机字节，简单 XOR 加密*/
inc eax
```

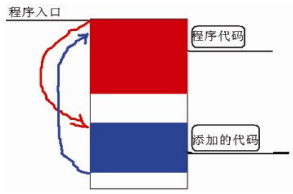

图 10　手动加密病毒代码段

```
/*首地址加 1*/
cmp eax,BBBBBBBB        /*比对代码段的末地址 BBBBBBBB*/
jle (XOR byte ptr ds:[eax],XX)/*没有到末地址，跳回继续加密*/
jmp 病毒原入口点        /*跳转到原地址执行代码*/
```

2.3 文件的加壳处理

前面讨论了三种普通的文件特征码的扫描技术，然而，当某些木马/后门程序被运行后，病毒被提前加载到内存，此时防毒软件检测到了位于磁盘上的病毒文件，却由于 Windows 不允许删除运行中的病毒文件而无法清除。另外，有些病毒使用了加壳技术，在执行时，壳先还原加壳过程中可能被压缩、加密的原始程序代码，再把执行权还给原始代码。这样使得文件特征码解毒彻底失效。但是病毒运行后，其代码在内存中一定会被还原，因此，好的防毒软件除了能精确定位文件特征码，还需引入针对内存特征码的扫描和具备压缩还原技术，能自动还原 Petite、UPX、Exepack、ASPACK、FSG 等数百种加壳软件，彻底解除隐藏较深的病毒，提高查杀的准确性，也为那些修改特征码来逃避杀毒软件的人带来了困难。

3　策略改进与技术展望

特征码技术具有低误报率、高准确性、高可靠等不可比拟的优势，其技术机理和执行流程也非常成熟。为了弥补特征码技术的被动性，建议辅以如下几种防病毒新技术：

(1)输入表关联特征码

病毒运行时需要调用存在于输入表中的 API 函数，如果将特征码锁定在可执行文件的“敏感”区域——输入表中，由于输入表位置固定，因此不能用通用跳转法来修改特征码，这样能有效地保护特征码。

(2)伪特征码

防病毒软件可以检测某一段自己的特征码，如果发现它被填充为 0，那么就激活原先设置的随机效用的伪特征码并报警，就算能找到这些特征码，对于查杀没有影响。

(3)广谱特征串过滤技术

为应对不断变化和未知的病毒，启发式扫描方式出现了。启发式扫描是通过分析指令出现的顺序，或特定组合情况等常见病毒的标准特征来决定文件是否感染未知病毒。因为病毒要达到感染和破坏的目的，通常的行为都会有一定的特征，例如非常规读写文件，终结自身，复制自身到系统目录，修改注册表某一键值，调用特定的 API 函数等等。所以可以根据扫描特定的行为或多种行为的组合来判断一个程序是否为病毒。这种启发式扫描比起静态的特征码扫描要先进的多，可以达到一定的未知病毒处理能力，但仍会有不准确的时候。特别是因为无法确定一定是病毒，而不

可能做未知病毒杀毒。

4 结语

自从上世纪八十年代病毒出现以来，特征码扫描就一直作为国际上反病毒公司普遍采用的查毒技术。其核心是从病毒体中提取病毒特征值构成病毒特征库，以供杀毒软件进行分析和判断。随着当前的病毒数量和技术呈现出前所未有的爆炸式增长，日益严峻的安全形势导致反病毒力量愈发疲于应对，单一的特征码扫描技术正逐步丧失和病毒对抗的能力，为了弥补传统反病毒技术的局限性，人们提出了启发式扫描技术和主动防御技术的新概念[5]。然而，目前没有一种技术是万能的，即使是基于“主动防御”的优秀思想，其发现病毒的成功率也仅有 60%-80%，如果结合传统的“特征码技术”，事实上几乎可以发现 100% 的恶意程序。因此，只有做到攻守俱备（既有强大的杀毒能力，完善的病毒库和较小的资源占用，又能及时安全地监控病毒和启发式扫描），将单一的防御向“主动防御”+“特征码技术”过渡，从“补丁式”杀毒转变为主动杀毒，才是现阶段安全的反病毒技术的必然发展趋势。

参考文献

1 王利林,许榕生.基于主动防御的陷阱网络系统.计算机工程与应用,2002,38(17):34－35.

2 陈伟,孙勇,杨义先,钮心忻.面向特征的信息隐藏检测研究.计算机系统应用, 2006,15(3):32－35.

3 Kreibich C, Honeycomb CJ. Creating Intrusion Detection Signatures Using Honeypots.Computer Communication Review, 2004,1:51－56.

4 王振海,王海峰.针对多态病毒的反病毒检测引擎的研究.微计算机信息, 2006,22(9-3):134－136.

5 张森强,郭兴阳,唐朝京.检测多态计算机病毒的数学模型.计算机工程, 2004,30(17):24－25,162.