文章编号:1001 - 9081(2006)02 - 0338 - 03

随机花指令加密算法研究

王海平,曹 林,孙国梓,陈丹伟
(南京邮电大学 计算机学院,江苏 南京 210003)
(sun@njupt.edu.cn)

摘 要:在对已有的花指令加密策略及其存在的问题进行分析的基础上,提出了随机花指令加密算法。给出了该算法的 JMP 扩展和 JMC 变换两种基本策略,研究了随机花指令的算法的工作流程和算法描述,并采用自编写的 JCEE 软件对上述算法进行了具体实践。

关键词:花指令;加密算法;代码模糊变换;反汇编
中图分类号: TP309 **文献标识码**:A

Study of stochastic junk code encryption algorithm

WANG Hai-ping, CAO Lin, SUN Guo-zi, CHEN Dan-wei
(*College of Computer, Nanjing University of Posts & Telecommunications, Nanjing Jiangsu* 210003, *China*)

Abstract: Based on current junk code encryption strategies and the analysis of existent problems, the stochastic junk code encryption algorithm was presented. After presenting two basic strategies named as JMP expansion and JMC switching used in the algorithm, the workflow and algorithm description of the algorithm was studied. At last the JCEE tool, which was self-developed, was used to validate the algorithm.

Key words: junk code; encryption algorithm; obfuscating transformation; disassemble

0 引言

代码模糊变换策略[1,2]应用于恶意代码可以增强恶意代码的生存能力,因此成为对抗逆向工程分析和误用检测工具的有效手段。同样,模糊变换策略也可以实施于信息系统安全、软件加密和攻击欺骗等安全领域。

花指令加密即所谓的添加冗余指令,是代码模糊变换策略中的一种有效方法。花指令也叫做伪指令、垃圾指令,该技术的主要特点是通过在程序体中插入无效指令或不影响原始代码功能的无用指令来破坏反汇编的结果和增加代码分析的难度,以达到加密保护的目的。本文主要是研究新的花指令加密算法,使之更有效破坏反汇编器的工作,最大限度地影响反汇编的结果。

1 花指令加密方法及研究现状

1.1 花指令加密的工作原理

动态跟踪是软件破解的基本过程,如果程序代码具有反动态跟踪的能力,那软件被破解的可能性就变小了[3,4]。花指令加密可以应用在反跟踪和反静态分析中。

花指令加密方法需要在原始指令的适当位置插入一些无用的字节。这些被插入的字节不能被执行,否则原始指令的逻辑会遭到破坏,这是花指令变换的基本条件。根据反汇编的工作原理,只有当花指令的最后一个或两个字节同正常指令的开始几个字节被反汇编器识别成一条指令时,才能有效破坏反汇编的结果。插入的花指令应当是一些不完整的指令,被插入的不完整指令可以是随机选择的。

1.2 研究状况及存在的问题

目前采用比较多的一种花指令形式是无条件跳转指令 JMP 加花指令[5]。这种策略通过简单变换后的各种扩展形式,在一定程度上增强了花指令的生存能力,但是这些扩展形式还是存在特征码,用工具仍可以自动去除。

另外一种形式是直接在原代码中的 JMP 指令后添加花指令[6,7],缺点是受代码中 JMP 指令个数的限制,如果 JMP 指令过少,则代码变换效果不明显。

综合而言,要想提高模糊度(反汇编器产生错误结果的程度),就必须应用多个方法,并保证方法使用的随机性和方法本身的随机性,使产生的花指令没有固定的特征码。由此,我们提出随机花指令加密算法的目的就是保证花指令无特征码并且具有较高的模糊度。

2 随机花指令加密算法

2.1 算法思想

根据前文的分析,目前采用花指令进行加密的问题主要在于存在特征码,容易用工具自动去除,其次是模糊度不够,效果不明显。对于这两个问题,本文提出两种策略来加以解决。第一种策略是 JMP 加花指令的扩展形式,简称为“JMP 扩展”。这种变换保证其具有足够的随机度,即使存在特征值,但特征值的数量巨大,完全列举所有情况也需要很多工作量。第二种策略是将代码中的条件跳转指令变换为反条件跳转指令加 JMP 指令的形式,简称为“JMC 变换”。这个策略进一步增加 JMP 指令的数量以便可以添加更多花指令。

算法的具体思想是:首先将原始指令分成若干个候选块

收稿日期:2005 - 08 - 25;**修订日期**:2005 - 11 - 09
基金项目:国家科技攻关项目(2004BA811B04);江苏省高校自然科学研究计划项目(05KJD520150);南京邮电大学校基金项目(NY205043)
作者简介:王海平(1970 -),男,江西南昌人,讲师,硕士,主要研究方向:信息安全、计算机应用; 曹林(1980 -),男,山东滨州人,硕士研究生,主要研究方向:信息安全; 孙国梓(1972 -),安徽天长人,副教授,博士,主要研究方向:信息安全、计算机应用; 陈丹伟(1970 -),陕西商洛人,副教授,博士,主要研究方向:信息安全、计算机应用.

并确定一个候选块随机序列,然后对每一个候选块进行如下处理:应用“JMP 扩展”策略对候选块内指令进行变换,然后应用“JMC 变换”策略对候选块内条件跳转指令进行变换,最后在变换后的代码的 JMP 指令后随机添加花指令。

2.2 “JMP 扩展”算法描述

JMP 扩展是用三条 JMP 指令对一段正常的指令进行变换,具体方法如下:

```
        JMP Label1
Label3:Instruction1
       Instruction2
       …                  } 正常指令
       InstructionN
        JMP Label2
Label1:JMP Label3
Label2: …                  正常指令
```

跳过的正常指令个数是随机选取的,但是个数不能超过候选块的指令数目。图 1 给出了“JMP 扩展”的工作流程。

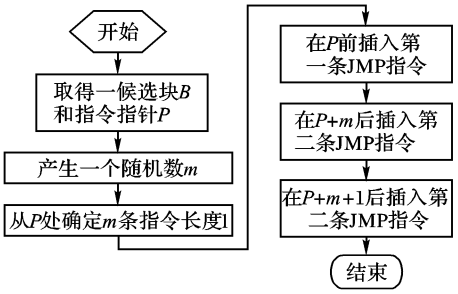

图 1　JMP 扩展工作流程

算法描述:

1) 取到一指令候选块 B,指令指针 P 指示候选块的第一条指令;

2) 产生一随机数 m,m 的大小在 B 的指令数目内;

3) 从指针 P 开始处向后扫描 m 条指令,确定这 m 条指令的长度 l;

4) 在 P 前插入第一条 JMP 指令 JMP + (l + 2),其中 2 是第二条 JMP 指令长度 2 字节;

5) 在 $P + m$ 后插入第二条 JMP 指令 JMP + 2,其中 2 是第三条 JMP 指令长度 2 字节;

6) 在 $P + m + 1$ 后插入第三条 JMP 指令 JMP - (l + 2),跳转到正常指令处。

2.3 “JMC 变换”算法描述

JMC 变换是将一条条件跳转指令变换为反条件跳转指令加 JMP 指令,具体方法如下:

```
JE Label1 ; 原始指令
        ↓ 变换
JNE Label2
JMP Label1;原 JE 跳转位置
Label2:…; 正常指令
```

所有的条件跳转指令都可以做这种变换。Intel x86 指令集中条件跳转指令编码方式有其特点,一条条件跳转指令的操作码和 0x01 异或就可以得到其反条件跳转指令。根据这个特点,“JMC 变换”就变得容易多了。图 2 给出了“JMC 变换”的工作流程。

算法描述:

1) 取到一指令候选块 B,指令指针 P 指示候选块的第一条指令;

2) 如果指令 P 为条件跳转指令则进行如下操作,否则跳转到 8);

3) 区分指令的操作码 opcode 和操作数 oprand;

4) 将 opcode 与 0x01 异或得到反条件跳转操作码 opcode′;

5) 根据 oprand 判断跳转类型是短跳转还是长跳转,得到 oprand′;

6) 将指令 P 变换为 opcode′ + oprand′,其中 oprand′是下条 JMP 指令的长度;

7) 在 P 后插入 JMP + oprand 指令;

8) 指令指针 P 指向下一条指令;

9) 如果 B 中指令全部处理完则结束,否则跳转到 2)。

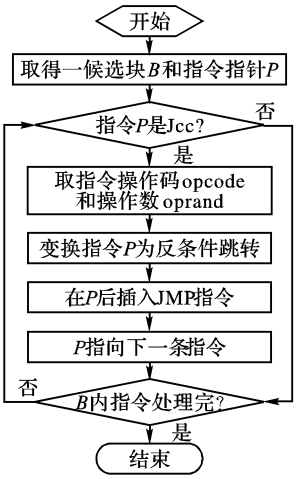

图 2　JMC 变换工作流程

2.4 算法描述

算法首先对原始指令进行分块操作,每个分块被称为候选块,分块操作的目的一是为了使“JMP 扩展”操作不会重叠影响,二是可以根据空闲空间的多少灵活调整分块大小。分块的大小可根据空闲空间多少灵活确定。确定一个分块的随机序列的目的是假如空闲空间不够,那么随机变换的候选块都不会一样。随机花指令加密算法会调用前面的两种策略对代码变换后,在 JMP 指令后添加花指令。图 3 给出了随机花指令加密算法的工作流程。

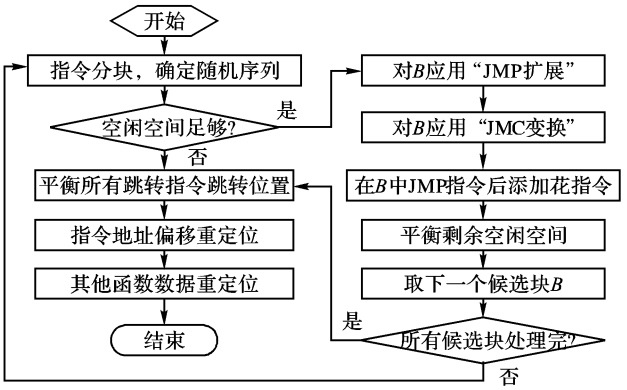

图 3　随机花指令加密算法工作流程

随机花指令加密算法描述:

首先对所有指令进行分块,对候选块产生一个随机序列 P,例如 P(4,1,3,7,2,5,6…),下面按照随机序列的顺序对所有分块如下处理:

1) 如果空间不够则退出,否则进行下面操作;

2) 对一个候选块进行一次“JMP 扩展”操作;

3) 对一个候选块中的条件跳转指令进行“JMC 变换”操作;

4) 在候选块中的 JMP 指令后随机添加花指令;

5) 平衡代码段剩余可用空间;

6) 如果所有候选块处理完毕则向下执行,否则取下一个候选块并跳转到 1);

7) 平衡所有跳转指令跳转位置;

8) 指令地址偏移重定位;

9) 其他数据、函数地址重定位。

3 随机花指令加密算法实践

3.1 随机花指令算法的实验步骤与结果

1）实验环境

硬件环境：celeron1.7GHz/256M Memory/20G HardDisk

软件环境：操作系统 Windows XP + SP2/JCEE 变换器/W32Dasm8.93/实验用可执行程序 test.exe。其中 JCEE 变换器是为实现变换而编制的测试程序，W32Dasm 是应用广泛的反汇编器调试器，程序 test.exe 是用来进行变换的程序。

2）实验步骤

a）对实验用原程序进行反汇编，得到汇编伪码 P；

b）用 JCEE 变换器对原程序进行变换，得到变换后的程序；

c）用反汇编器对变换后的程序进行反汇编，得到汇编伪码 P'；

d）比较 P 和 P' 得到结论 R。

3）实验结果

候选块的大小选为 10 条指令，每条 JMP 指令后最大添加 8 个字节花指令。test.exe 程序经变换后得到的数据如表 1 所示。

表 1 随机花指令变换结果相关数据（单位：bytes）

	代码长度	空闲空间	指令数目	JMP 数目	Jcc 数目	JunkCode
P	528	499	151	22	40	
P′	858	169	200	61	4232	
P′ - P	330	-330	49	49	0232	

经验证，采用本随机花指令加密方法对 test.exe 文件进行花指令变换，采用花指令去除工具对变换后的结果进行花指令搜索，均未能发现相关的花指令特征。

3.2 实验分析

经过“JMP 扩展”变换后的代码要跳过一定数量的正常指令，然后在三条 JMP 指令后添加花指令，总共存在的特征值是 8 * 8 * 8 * 10 = 5 120 种，如果基数扩大，特征值数量更多。因此想要定义所有的特征不太实际。在 JMP 后添加了花指令后不能自动去除还有一个原因是 JMP 指令作为原始代码的组成部分是不能被去除的，否则程序逻辑遭到破坏。

4 结语

文章提出的随机花指令加密算法所产生的花指令具有无特征码的特点，不能用工具自动去除；并且对原始指令变换后可以添加更多的花指令，达到较高的模糊度，起到了软件保护的目的。本文以实例的方式对随机花指令加密算法的有效性进行了验证，今后需要更为深入地研究程序逻辑一致性的形式化证明方法，由此，才能更为有效地保证该算法的有效性。

参考文献：

[1] Hume. 病毒和网络攻击中的多态、变形技术原理分析及对策[R/OL]. http://www.xfocus.net/pro jects/Xcon/2003/Xcon2003_hume.pdf, 2003, 12.

[2] 卿斯汉. 恶意代码机理[Z]. 北京：北京大学软件学院, 2004.

[3] 段钢. 加密与解密[M]. 第 2 版. 北京：电子工业出版社, 2003.

[4] 于淼, 孙强. 对加壳技术的改进：超粒度混杂技术[J]. 计算机应用, 2004, 24(8)：137 - 139.

[5] 林宣雄, 李怀祖, 张文修. 扰码机制在反静态分析中的应用[J]. 微电子学与计算机, 1996,(1)：16 - 19.

[6] LINN C, DEBRAY S. Obfuscation of Executable Code to Improve Resistance to Static Disassembly[A]. Proceedings of the 10th ACM conference on Computer and communications security[C]. 2003. 290 - 299.

[7] GRIFFITHS AL. Binary protection schemes [J]. CodeBreakers - Journal, 2005, 2(1)：1 - 91.

（上接第 317 页）

录服务器仅仅接受授权用户的写请求，和接受拥有 userCertificate 属性的实体的写请求，当然授权用户只能将解密后的证书写入属于自己实体的目录下。另外，ACLs 列表必须经过某种策略配置，配置后，用户密码不能被其他的用户看见。这样就会减少其他终端用户进行字典攻击的可能性。

拥有加密私钥的终端用户必须下载对应更新后的证书，本文提供的 Schema 有一个可选变量，这个可选变量可以将身份表示和 POP 信息更紧密地联系在一起，并且保证用户下载的是对应的证书。实现过程如下：用户在 LDAP 目录上使用的加密密码仅仅是用户密码的一半，另外一半在用户申请更新的过程中提供，这部分密码作为 CA 和终端用户的秘密共享，用户必须把两半密码合并后，才能得到目录的授权和下载对应的更新后被加密的证书。

CA 可以基于不同的策略来释放 Schema，例如，如果三天后在目录里证书仍然是加密状态，证书将被删除掉，让用户不能到目录上来使用证书。另外，证书解密后，用户在 LDAP 上的密码也被删除掉，这样就可以使用户没有权限访问其他的目录。

3.3 可用性问题

用 LDAP 管理 PKI 过程的应用提高了 PKI 的应用性，对于用户来说，解密一个数据包，比让自己的私钥暴露在通信过程中更安全。解密一个象证书和密码的数据包是解密密钥的功能，这些已经在很多客户端软件中使用了。另外很多被 LDAP 支持的客户端软件能够拿到其他的证书，这些特点能够扩展和联合，自动地从目录中得到加密数据，将它解密并且将证书存回目录中，最后的这个过程对于终端实体来说是透明的。

4 结语

本文在首先对几种传统的加密证书更新时的 POP 方法的优缺点进行分析，提出一种基于 LDAP 的 Schema 来解决加密证书更新时的 POP 管理方法。实践表明，本方法过程简单，私钥不暴露，安全性高，还可以用于加密证书的撤销等过程，提高了 PKI 的应用性。

参考文献：

[1] BICAKCI K, BAYKAL N. A New Design of Privilege Management Infrastructure with Binding Signature Semantics[A]. EuroPKI 2004[C]. LNCS 3093, 2004. 306 - 313.

[2] KARATISLIS V, LIPPERT M, WIESMAIER A. Using LDAP Directories for Management of PKI Processes[A]. EuroPKI 2004[C]. LNCS 3093, 2004. 126 - 134.

[3] 王春耕. 朱建涛. 大规模机群系统中基于 LDAP 的用户管理[J]. 计算机工程与应用, 2004, 40(18)：47 - 49.

[4] 赵妍, 袁野, 刘冰. 基于 LDAP 协议与 Kerberos 认证机制的统一认证[J]. 信息技术, 2004, 12(28)：46 - 49.

[5] ITU-T X.509[DB/OL]. http://houmb.qlsc.sdu.edu.cn/ebook/is/X509/X509_4thEditionDraftV8.pdf, 2001 - 05 - 03.