反计算机取证技术研究

张有东[1,2],王建东[1],朱梧槚[1]

(1.南京航空航天大学计算机科学与工程系,江苏 南京 210016;2.淮阴工学院计算机工程系,江苏 淮安 223003)

摘要:分析了反计算机取证的基本概念和方法,比较了反计算机取证所采用的数据擦除、数据加密、数据隐藏、数据混淆和数据转换等主要技术,并提出了一种安全、高效的基于 m 序列的数据擦除方法.

关键词:反取证 数据擦除 数据隐藏 数据混淆 数据转换

中图分类号:TP311 **文献标识码**:A **文章编号**:1000-1980(2007)01-0104-04

目前,在计算机网络犯罪手段与网络安全防御技术之间的对抗不断升级的形势下,计算机取证作为一门交叉学科,越来越受到计算机安全和法律专家的重视.与此同时,一种新的反计算机取证(Anti-Computer Forensic,以下简称反取证)技术也正悄然兴起.国内外关于这种技术公开的、系统的报道还不多.然而,通过分析研究反取证技术以保证计算机取证的科学性和有效性是非常重要的[1].本文将提出反取证技术的定义与方法,分析比较反取证采用的主要技术,并提出一种基于 m 序列的反取证算法.

1 反取证定义

黑客技术、反侦查技术和入侵检测逃避的主要目的是逃避系统的安全检测以成功地侵入系统,但同时也会在系统中留下大量的电子证据.随着取证技术的发展,高明的计算机犯罪分子在作案前开始周密计划和做反跟踪准备,作案后更改、删除目标机的信息,并清理自己的工具机,这是反取证技术兴起的最初动因.另外,反取证技术的发展还来源于保护个人隐私的需要,即保护合法公民或公司的正当法律权利不受侵犯.

早期的反取证技术可定义为"删除或破坏电子证据,使得取证方法无效".随着计算机取证的程序化及其技术的完善,反取证发展为针对计算机取证过程的各个阶段及其形成证据链(Chain of Custody)的条件,破坏电子证据的调查、保护、收集、分析和法庭诉讼,减少被获取证据数量,降低被获取证据质量,从而尽量隐藏或不在对方系统甚至自己系统中留下法律意义上的证据.

2 反取证基本方法分析

传统的反取证方法有改变文件扩展名、使用 Swap 空间等,现代计算机取证工具已经能很好地处理它们.由于计算机取证调查受技术和非技术因素的影响,反取证的基本方法同样有策略上的考虑和技术上的方法 2 个方面.

计算机取证需要建立有效的证据链,否则将产生法律疑点(Legal Doubt),也就是使得控方律师在被告有罪的推理中出现疑点,从而不被法庭采纳.由于 Internet 的特点,匿名访问、匿名账号、匿名存储以及利用 Root Kit 绕开系统日志等,都将使取证调查人员在确定罪犯在特定计算机上的活动时产生困难,罪犯能够匿名的越多,调查中的疑点就越多,甚至罪犯会"黑客"自己的计算机,以造成他们的计算机是被其他罪犯利用来犯罪的假象.

在技术层面上,目前的反取证技术主要是针对证据的收集和分析.罪犯利用取证调查开始时影响判断的因素来干扰、阻挠取证,导致调查偏离犯罪活动场所,这些因素包括数据集的信息异常、失效或不产生指定的结果.证据收集阶段的反取证技术主要有摧毁、隐藏、假造信息以及防止证据的创建等.

收稿日期:2006-03-12

基金项目:国家重点基础研究发展规划(973)项目(G1999032701)

作者简介:张有东(1967—),男,江苏淮安人,副教授,博士研究生,主要从事数据挖掘、入侵检测研究.

在取证分析阶段,取证人员已获得并保护了证据,并且将要去理解这些证据,针对这个阶段的反取证技术是通过数据加密阻止分析人员对所获取的证据的理解,甚至通过某种方法歪曲取证人员可能会感兴趣的数据[2].

另外,由于目前的计算机取证主要由一些取证工具软件完成,反取证的一种实用的方法是针对取证工具软件的缺点进行攻击,如对 Unix 下的取证工具 TCT 的攻击等.

3　反取证主要技术分析

3.1　数据摧毁技术

摧毁证据是阻止取证的最有效的方法,主要攻击取证的收集阶段.从隐私与敏感数据保护角度,Peter Gutmann 在 1996 年就提出了数据摧毁的问题[2].从存储原理上讲,摧毁证据可以有 3 种选择:摧毁实际数据,摧毁元数据(Meta-Data),或者是两者的结合.对于磁存储介质上的数据摧毁,有物理摧毁和逻辑摧毁 2 种方法,物理摧毁包括消磁(Degaussing)或暴力(Brute Force)破坏,逻辑摧毁则包括改变驻留媒体的数据组成、移走相关数据的踪迹(Traces)等.进一步地,反取证不仅摧毁实际的数据,还可能根除摧毁痕迹,这极大地降低了取证人员在计算机系统上发现证据的能力.

值得注意的是,一般的文件删除甚至删除后的多次覆盖,并不意味着数据的摧毁,事实上,诸如 MFM(Magnetic Force Microscopy)和 STM(Scanning Tunneling Microscopy)技术可以从磁介质中找到被完全重写的数据.内存芯片也有一种未公开的诊断模式,允许访问比 1 比特位小的信息,通过修改电路,类似于信号调制,可以检索旧的信息[3].

3.2　数据加密技术

虽然摧毁所有潜在的证据很有效,但罪犯可能想保存一些有用的数据.为了使这些数据不被取证人员所理解,罪犯通常会使用加密技术.现代加密技术的发展使得这种方法很容易被犯罪分子所利用,当然也有专门研究用于反取证的加密技术,文献[4]就介绍了一种基于两层 PBKDF2 的反取证数据存储方法.

3.3　数据隐藏技术

数据隐藏(Data Hiding)是一种有效的对付取证收集的技术.数据隐藏包括隐秘术(Steganography)和数字水印 2 个主要研究方向.在反取证中应用的主要是隐秘术,它将秘密信息嵌入到看上去普通的信息中进行传送,以防止第三方的检测.也就是说,信息隐秘是通过建立传送秘密信息的秘密信道来实现的.可以对隐秘系统的隐秘性和鲁棒性这 2 个基本要求进行定量分析:

已知消息 M,载体 C 可看作一列随机变量,隐藏信息 K 通常假定为独立、均匀分布的随机变量,对给定的 K,由编码算法 f_k 产生的隐秘消息 $S = f_k(C, M)$ 仍为一随机变量,隐秘性要求 S 与消息源中的典型消息不可区分,这一条件可用编码约束 $Ed(C, S) <= \delta$ 来表示.其中 d 为失真度,E 是关于 C 和 S 的联合概率分布的期望算子.显然,界 δ 要足够小才能保证达到隐秘性.鲁棒性要求在各种扰动和攻击后,失真后的信息 S^1 仍然具有合理的质量,可用通道约束表示为 $Ed(S, S^1) <= \delta$.

对反取证而言,加密的缺点是直接表明通信是秘密的,而隐秘术的缺点是需要大量的开销来隐藏相对少的信息比特,即 K 和 S 的比值很小,效率很低,另外,一旦该系统被发现,就会变得完全没有价值.由此,先加密信息,再用隐秘术来隐藏,将使取证更加困难.文献[5]还指出了与新的加密技术相联系的更有效的使用 Slack 空间隐藏数据的方法.

3.4　数据转换技术

罪犯一旦攻陷了一个远程系统的组件,其主要目标就是尽可能长地保持对新的已获取资源的控制.对于管理和调查人员而言,在证实有未授权对象存在时会认为有网络入侵者侵入系统,这些对象包括文件、目录、进程、任务或网络连接等.为此,罪犯会采取数据转换(Data Transformation)技术将上述对象从常规的操作中隐藏,并保持或重建一个调查人员的信任,使调查人员迷失方向或产生错误的判别[6].

3.5　数据混淆技术

混淆(Obfuscation)通常描述故意使某事物难以理解的行为.数据混淆则是一种逃避技术,主要指用特殊符号(如 Backspace,Insert 和 Delete 等控制符)来隐藏攻击.这种技术要求用十六进制形式表示字符以逃避检测.另外,使用 Unicode 表示法也是一个有效的逃避检测的方法.数据混淆使得罪犯能够在对程序和数据进行

访问的同时,限制调查人员对证据的识别和收集.

3.6 防止数据创建技术

防止数据创建(Data Creation Prevention)是一种有效的攻击取证收集的手段.在一个犯罪分子和信息系统的常规事务中,大量潜在的证据被建立,如果罪犯能够在调查人员识别、收集之前防止相关数据的创建,将增加他们成功避开被检测和取证的机会.直接的防止技术包括使用 Root Kits 或修改系统二进制(System Binaries)[7],但是,这种方法在限制数据创建的同时也限制了罪犯的攻击能力.

4 数据擦除反取证应用分析与设计

4.1 数据擦除技术

数据擦除(Data Wiping)用来定位用户磁存储介质上存储的活动记录,搜寻并不可逆地删除它们,是一种逻辑摧毁数据的方法.数据擦除起源于对用户隐私的保护,但目前已被犯罪分子用来进行反取证.

数据擦除的基本思想是用某种覆盖方案反复重写磁介质.典型的覆盖方案有3种:美国国家安全局推荐的7次随机覆盖方案[8];美国国防部 DOD 5220.22-M(C and E)标准推荐的3次0,1交替覆盖方案;Gutmann 方案[2],它需要进行35次覆盖,被认为是最安全的擦除算法[9].前两者较简单,后者是针对当时的3种磁盘编码 MFM,PPM 和 PWM 提出的,数据擦除彻底,但缺点是速度太慢,且仅对 MFM 等3种编码效果较好.目前的一些数据擦除软件主要是基于简单的0,1交替覆盖方案,并不能完全清除所有数据[10].

4.2 基于 m 序列的数据擦除覆盖方案设计

现代编码技术可以用五元组$(d,k;m,n;r)$来表示不同的 RLL 码,其中,d,k 分别表示记录序列中相邻两个1之间至少有 d 个零、最多不得超过 k 个零,只要正确设计 d,k 值,即可获得优良的 RLL 编码性能,在近几年发展的高密度磁盘中,不同的 RLL 码得到了广泛应用.由于编码表示的统一,基于 Gutmann 方案的思想,笔者设计了基于 m 序列的通用覆盖方案.

m 序列是一种二进制伪随机序列,通常由线性反馈移位寄存器生成,设一线型反馈移位寄存器为 r 级,其序列多项式 $G(x)$可表示为

$$G(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + \cdots = \sum_{n=0}^{\infty} a_nx^n \tag{1}$$

其中 a_n 是一个二元有限域 $\{0,1\}$的元素,运算为模2和.显然,序列 $\{a_n\}$与 $G(x)$一一对应.再根据 Caley-Hamiton 定理,可推出其特征多项式

$$f(x) = \sum_{i=0}^{r} c_ix^i \tag{2}$$

其中 c_i 为反馈系数,其取值也是一个二元有限域 $\{0,1\}$的元素.通过变换,可推出序列多项式与特征多项式之间的关系为 $G(x) = x/f(x)$.由此可得到产生序列 $\{a_n\}$的方法,即对 $f(x)$进行长除,得到序列多项式 $G(x)$,$G(x)$的系数就是所求序列.当该线性反馈移位寄存器的抽头系数取自一本原多项式时,所得序列即为一 m 序列.

根据 m 序列的性质,其周期 $L = 2^r - 1$,考虑到1个扇区的字节数为512,若 $L+1$ 对应于一个磁盘扇区字节数,则可得 r 为9,其本原多项式共有 $\Phi(2^9-1)/9 = 48$ 条,可从中选择反馈系数(1000010001),即得本原三项式

$$f(x) = x^9 + x^4 + 1 \tag{3}$$

它产生的 m 序列长度为511位,在最后加上与其前一位反转的值,就得到一个512位的序列 m_1.

m_1 可以用软件方法产生,根据 RLL(d,k)码的性质,在产生 m_1 时对2个1之间零的个数进行限制,即2个1之间至少有 d 个、最多不得超过 k 个零.这只需要在算法中增加0计数器 Count 0,当产生1个1时 Count 0 开始对0计数,当下一个1产生时,若 Count 0 < d,则将该1反转,同时将 Count 0 加1,当下一个1产生前,若 Count 0 > k,则将该0反转,并重置 Count 0.由此可产生一个针对不同 RLL 码的以扇区为单位的磁翻转序列 m_2,而且,不同于流加密,这样产生的序列是不可逆的.一组 $m_2^i(i=1,\cdots,n)$即可组成一个覆盖方案,其覆盖次数 n 可根据用户对删除数据的重要性来确定.

显然,该方案有3个性质:(a)该方案是一种基于 RLL 码原理的通用覆盖方案;(b)覆盖是按扇区进行的;

(c)覆盖序列具有密码学性质.由于数据擦除的程度与覆盖次数及磁盘记录编码有关,因此本文所提出的方案,对于相同的覆盖次数而言,据性质(b)和(c),其安全性能和速度高于 NSA 方案 如果覆盖次数相同,据性质(a)和(b),其速度高于 Gutmann 方案且具有通用性.

5 结 束 语

目前的反取证技术还处于起步阶段,越来越多的安全技术与理论,尤其是为保护个人稳私所采取的加密、隐藏、数据擦除等技术,已经被应用于计算机犯罪和反取证.面对反取证技术的发展,计算机取证系统的设计需要进一步加强现有软件系统的防范措施.可以预见,取证与反取证的对抗必将越来越激烈,只有对反取证技术深入了解,才能不断提高取证软件的可靠性、可信任性.在某种意义上,保护隐私以外的反取证是一种犯罪行为,对反取证技术的研究将成为信息安全与取证技术研究的重要方向.

参考文献:

[1] JOHANSSON C. Forensic and anti-forensic computing[EB/OL]. 2002[2006-01-16]. http://www.fukt.bth.se/~uncle/papers.

[2] GUTMANN P. Secure deletion of data from magnetic and solid-state memory[C]//The Sixth USENIX Security Symposium. San Jose:[s.n.], 1996:23-35.

[3] FARMER D, VENERMA W. Forensic discovery[M]. Boston:Amazon, 2004:146-147.

[4] CLEMENS F. TKS1—an anti-forensic, two level and iterated key setup scheme[EB/OL]. 2004[2006-01-16]. http://clemens.endorphin.org.

[5] RIPE P. Advanced anti forensics[EB/OL]. 2005[2006-01-16]. http://www.phrack.org/show.php.

[6] CHRISTIAN S J, MICHAEL L. Digital anti-forensics: emerging trends in data transformation techniques[EB/OL]. [2006-01-16]. http://www. seccuris.com/documents/papers.

[7] CALOYANNIDES J, MICHAEL A. Privacy protection and computer forensics[M]. Boston:Artech House, 2004:223-225.

[8] U S DoD. Standard 5220.22-M National industrial security program operating manual(NISPOM)[S]. Washington: GPO, 1995:800-831.

[9] PETER B. Data security for surplus or end of life hardware[C]//2004 NYS Cyber Security Conference. New York:[s.n.], 2004:21-33.

[10] MATTHEW G. Evaluating commercial counter-forensic tools[C]//DWFS 2005. New Orleans: LA, 2005:106-118.

Research on computer anti-forensics

ZHANG You-dong[1,2], WANG Jian-dong[1], ZHU Wu-jia[1]

(1. *Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing* 210016, *China*;
2. *Department of Computer Engineering, Huaiyin Institute of Technology, Huaian* 223003, *China*)

Abstract: The concept and methodology about computer anti-forensics were analyzed. The main techniques of anti-forensics, such as data wiping, data hiding, data obfuscation, and data transformation, were compared. Then, a safe and high-efficient data wiping algorithm based on sequence *m* was proposed.

Key words: anti-forensics; data wiping; data hiding; data obfuscation; data transformation