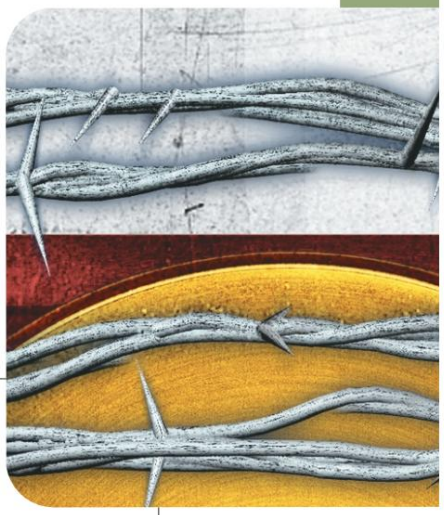
记忆通过的数据：磁盘清理研究实践



西蒙·L。加芬克尔和 ABHI 谢拉特
马萨诸塞州研究所技术

保密和可恢复，作为作者自己的实验显示。这些信息的可用性很少公开，但对它的认识肯定会传播开来。

许多废弃的硬盘驱动器包含的信息既

信息安全的一个基本目标是设计防止未经授权的披露的计算机系统的机密信息。有很多方法可以保证此信息隐私。最古老和最常见的技术之一是物理隔离：保持机密只有经过授权的个人才能访问的计算机上的数据。例如，大多数单用户个人计算机，包含对该用户保密的信息。

具有不同授权级别的人使用的计算机系统通常采用身份验证、访问控制列表和特权操作系统来维护信息隐私。过去 30 年的大部分信息安全研究都集中在改进验证技术和开发方法，以确保计算机系统正确实施这些访问控制规则。

密码学是另一种可以确保信息隐私的工具。用户可以在发送数据时对其进行加密，并例如，使用安全套接字层 (SSL) 加密协议在预期目的地对其进行解密。他们还可以对存储在计算机磁盘上的信息进行加密，以便只有以下人员可以访问该信息那些具有适当解密密钥的人。加密图形文件系统 1-3要求输入密码或密钥启动，之后它们会自动将数据加密为它被写入磁盘并在读取数据时解密数据；如果磁盘被盗，小偷将无法访问数据。然而，尽管存在加密文件系统，一般公众似乎很少使用它们。

在没有加密文件系统的情况下，当所有者不正确地更换磁盘驱动器时，机密信息很容易获得。例如，2002 年 8 月，美国退伍军人管理局医疗中心印第安纳波利斯淘汰了 139 台计算机。其中一些系统一些被捐赠给学校，而其他的则在公开市场上出售，并在至少三个人最终在一家旧货店一位记者购买了它们。不幸的是，VA 忽略了清理计算机的硬盘驱动器 也就是说，它未能删除驱动器的机密信息。许多计算机后来被发现包含敏感的医疗信息，包括患有艾滋病和心理健康问题的退伍军人的姓名。新主人还发现了44个信用卡号码印第安纳波利斯工厂使用的。4

弗吉尼亚州的惨败只是美国众多著名案例之一一个受托处理机密信息的组织之前忽略了对硬盘进行适当的清理计算机的处置。其他情况包括：

- 2002 年春天，宾夕法尼亚州劳工和工业部将一系列计算机出售给当地经销商。这些计算机包含“数千有关国家雇员的信息文件”，该部门未能删除。 5
- 2001 年 8 月，Dovebid 拍卖了 100 多个 Viant旧金山办公室的电脑咨询公司。硬盘驱动器包含机密 Viant 未能删除的客户信息。6
- 普渡大学的一名学生在学校的剩余设备交换中心购买了一台二手的 Macin tosh 电脑，却发现这台电脑的硬驱动器包含一个 FileMaker 数据库，其中包含超过 100 人的姓名和人口统计信息申请学校昆虫学系。
- 1998 年 8 月，其中一位作者购买了 10 本地计算机商店的计算机系统。这电脑，其中大部分是三到五年的，

表 1. 每年出货的 Tbytes
全球硬盘市场。
（由 IDC 研究提供）

年	已发货 TB
1992	7,900
1993	16,900
1994	33,000
1995	77,800
1996	155,900
1997	344,700
1998	698,600
1999	1,500,000
2000	3,200,000
2001	5,200,000
2002	8,500,000

包含其前所有者的所有数据。一通
puter 曾是一家律师事务所的文件服务器并包含
特权客户-律师信息。另一台计算机有一个社区组织使用的
数据库
提供心理健康服务。其他磁盘包含大量个人文件。

· 1997 年 4 月,内华达州帕朗的一名妇女购买了
一台 159 美元的二手 IBM 电脑,发现它
包含 2,000 名患者的处方记录
在亚利桑那州坦佩的 Smitty s Supermarket 药房配药。
包括患者的
姓名、地址和社会安全号码以及一份清单
他们购买的所有药物中。记录包括患有艾滋病、酗酒和抑郁
症的人。 7

这些轶事报道很有趣,因为
它们的相似性和相对稀缺性。显然,机密信息已多次通过在二
级市场上出售的计算机被披露。

那么,为什么很少有关于意外事件的报告
披露?我们提出三个假设:

· 此类披露极为罕见
· 退休人员经常披露机密信息
此类事件根本没有新闻价值的系统
· 二手设备充斥着机密信息
化,但没有人在寻找它 否则有
人们在寻找,但他们并没有公开这一事实

为了进一步调查问题,我们购买了更多
二级市场上超过 150 个硬盘。我们的目标
是确定它们包含哪些信息,以及
如果有的话,前业主用来清洁的方法是什么?
在他们丢弃驱动器之前。在这里,我们介绍
我们的发现,以及我们用于描述从打捞驱动器中恢复或可恢
复的编队的分类法。

硬盘市场

每个人都知道,有一个戏剧性的增长
磁盘驱动器容量和相应的质量减少 -
近年来的仓储成本。尽管如此,很少有人意识到如何
真正惊人的数字实际上是。根据
市场研究公司Dataquest,近1.5亿磁盘
驱动器将在 2002 年淘汰 高于 2002 年的 1.3 亿个
2001. 尽管许多此类驱动器被销毁,但仍有大量驱动器被重
新用于二级市场。 （这
市场正在迅速增长,甚至成为主流企业的供应来源,10 月 15
日的封面就证明了这一点
CIO 杂志中的故事，“便宜的好东西:如何使用
二级市场对您的企业有利。”8)

根据市场研究公司 IDC 的数据,全球磁盘驱动器行业的出
货量将在 210 到 215 之间
2002 年百万磁盘驱动器;这些磁盘的总存储量
驱动器将是 850 万 TB（8,500 PB,或 8.5
× 1018 字节）。虽然摩尔定律规定了翻倍
每 18 个月集成电路晶体管,硬盘
存储容量和运送的总字节数是
以更快的速度翻倍。表 1 显示了 TB
过去十年在全球硬盘市场出货。

不可能知道任何磁盘驱动器会持续多久
继续服役； IDC 估计典型驱动器的使用寿命为五年。如表 2
所示,Dataquest 估计
每 10 个磁盘驱动器,人们将淘汰 7 个磁盘驱动器
2002年出货;这高于退休率
1997 年 10 比 3（见图 1）。作为 VA 医院的
经验表明,许多被一个组织 “再累”的磁盘驱动器可能会出现
在其他地方。除非

表 2. 全球硬盘市场。 （由 Dataquest 提供）

年	发货单位（以千计）	每兆字节成本给最终用户	退休（以千计）	退休率*（百分比）
1997	128,331	0.1060	40,151	31.2
1998	143,927	0.0483	59,131	41.0
1999	174,455	0.0236	75,412	43.2
2000	199,590	0.0111	109,852	55.0
2001年	195,601	0.0052	130,013	66.4
2002年	212,507	0.0025	149,313	70.2

* 每年淘汰的驱动器与出货的驱动器的比率

退役驱动器被物理损坏,信息不佳
安全实践可能会危害信息隐私。

硬盘无处不在

与其他海量存储介质相比,硬盘
在确保长期数据机密性方面存在特殊和重大问题。一个原因是身体和
其他大容量存储设备的电子标准有
多年来发展迅速而不兼容,而
集成驱动电子/先进技术
附件 (IDE/ATA) 和小型计算机系统接口 (SCSI) 接口都保持了前向
和向后兼容性。人们使用硬盘驱动器
已经 10 岁了,拥有现代消费电脑
只需将它们插入即可:物理、电气和逻辑标准非常稳定。

这种前所未有的兼容性水平维持了正规和非正规二级市场
用过的硬盘。对于磁带、光盘、闪存和其他形式的大容量存储来说,情况并非如此,
那里的多样性要多得多。随着电流
设备,人们通常不能使用旧媒体,因为
格式更改(例如,数字音频磁带 IV 驱动器,
无法读取 DAT I 磁带,也无法读取 3.5 英寸磁盘驱动器读取 8 英寸软盘。)

导致维护数据机密性问题的第二个因素是数据的长期一致性。
文件系统。今天的 Windows、Macintosh 和 Unix 操作系统可以透明地使用 FAT16 和 FAT32
微软在 1980 年代普及的文件系统和
1990 年代。(正如我们在“通过擦除进行消毒”中讨论的那样部分,FAT代表文件分配表,是一个链接
DOS 用来管理随机访问设备上的空间的磁盘集群列表;16 或 32 是指扇区号
位长度。)因此,使用了 10 年的硬盘驱动器不仅在机械和电气上与当今的计算机兼容,而且它们包含的数据无需专用工具即可轻松访问。旧磁带并非如此,

通常使用专有备份系统编写,可能使用专有压缩和/或

加密算法也是如此。

对盒式磁带进行消毒的常用方法是使用
大容量磁带橡皮擦,成本不到 40 美元,可以擦除
只需几秒钟即可完成一整盘磁带。大容量橡皮擦可以擦除
几乎市场上的任何磁带。一旦擦除,磁带可以
像新的一样重复使用。然而,大容量橡皮擦很少
与硬盘一起工作,创造了使数据机密性复杂化的第三个因素。在某些情况下,商业
可用的大容量橡皮擦根本不能产生足够的
强磁场影响磁盘表面。当他们
确实,它们几乎总是使磁盘无法使用:此外
擦除用户数据,批量擦除器删除低级轨道和
格式化信息。虽然可能
使用供应商特定的命令恢复这些格式代码,这些命令通常对用户不可用。

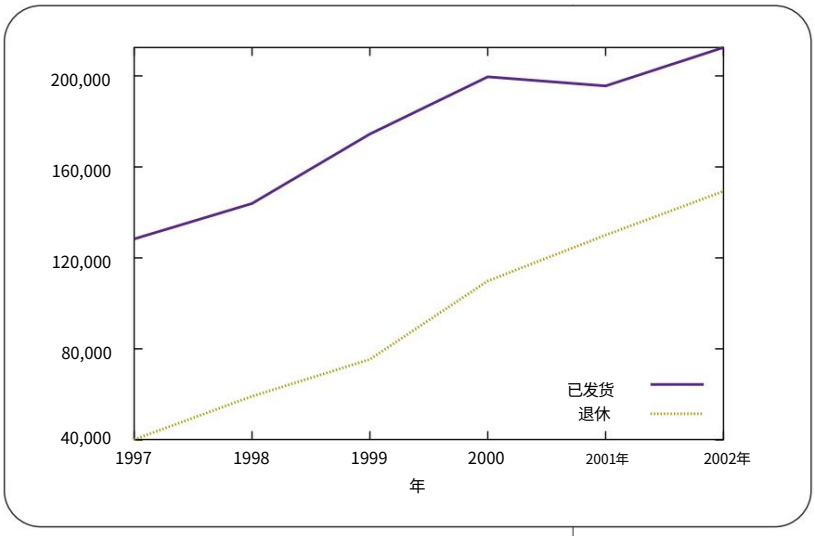


图1。全世界硬盘单位市场发货与每个退休年。(礼貌数据查询)

消毒问题

当数据存储设备在二级市场上销售时,人们用来确保信息隐私的大多数技术都失败了。例如,计算机操作系统提供的任何保护都会在某人丢失时丢失

从计算机中取出硬盘驱动器并将其安装到
可以读取磁盘格式的第二个系统,但是
不尊重访问控制列表。这种脆弱性
留在信息系统上的机密信息已
自 1960 年代以来得到认可 9

确保数据机密性的法律保护是
同样无效。在加利福尼亚诉格林伍德案中,美国最高法院
法院裁定弃置物不存在隐私权
材料10 同样,个人或
公司可以声称在他们自己出售的系统中拥有隐私或商业秘密利益。
经验表明,人们经常捡拾
来自废物流和再利用的电子元件
他们在原主人不知情的情况下。

因此,为了保护他们的隐私,个人和组织必须从磁盘中删除机密信息
驱动器重新使用、报废或将其作为
完整的单元 也就是说,他们必须清理他们的驱动器。

正确消毒的最常用技术
硬盘包括

- 物理损坏驱动器,使其无法使用
- 对驱动器进行消磁以随机化磁域 很可能导致驱动器在
  过程
- 覆盖驱动器的数据,使其无法恢复

消毒因社会规范而变得复杂。显然,
确保驱动器信息受到保护的最佳方法是
物理销毁驱动器。但是很多人觉得
当 IT 设备被丢弃和
被摧毁而不是重新定向到学校、社区

nity 组织、宗教团体或欠发达的
其他人可能会从使用该设备中受益的国家　即使该设备已经过时几年。

通过擦除进行消毒

许多人认为,当他们删除计算机文件时,他们实际上是在破坏信息。大多数情况下,
然而,删除或擦除命令实际上并没有
从硬盘中删除文件的信息。铝
尽管“擦除”的确切概念取决于文件
在大多数情况下,使用的系统最常删除文件
仅重写指向文件的元数据,但
保留包含文件内容的磁盘块。

考虑 FAT 系统,它是主要文件
我们研究中使用的格式。有四种略有不同
此文件系统的版本:FAT12、FAT16、VFAT 和
FAT32。硬盘总是按 512 字节寻址
部门。 FAT 文件系统进一步将数据扇区分组为
簇,由2i个扇区组成,其中i是参数集
格式化驱动器时。每个硬盘集群都有一个
FAT 中描述其状态的条目。集群是

- 文件的一部分,并指向该文件的下一个簇
- 文件中的最后一个簇,因此拥有一个特殊的文件结束 (EOF) 值
- 免费,因此为零
- 标记有缺陷

本质上,FAT 是一个与文件相对应的簇的链表。 （对于更全面的概述
FAT 文件系统,请参阅 Microsoft 的规范。11)

当操作系统擦除一个 FAT 文件时,两个
事情发生。一、系统修改文件名的第一个
文件目录条目中的字符,表示该文件
已被删除并且目录条目可以被回收。其次,系统将文件的所有
FAT 簇移动到硬盘驱动器的空闲簇列表中。实际文件

数据永远不会被触及。确实有很多节目
正如我们稍后讨论的那样,可以恢复已删除的文件。

尽管我们的“擦除”语义概念暗示
删除数据后,FAT 文件系统（以及许多其他现代文件系统)不符合我们的期望。

通过覆盖进行消毒因为物理破坏相对复杂且

不满意,并且因为使用操作系统
擦除文件并不能有效地清理它们,许多人更喜欢通过故意用
其他数据覆盖该数据来清理硬盘驱动器信息,以便

无法恢复原始数据。虽然覆盖
比较容易理解和验证,可以
在实践中有些复杂。

覆盖硬盘的一种方法是用 ASCII NUL 字节（零)填充每个可寻址块。如果
磁盘驱动器正常工作,然后这些中的每一个
blocks 报告在回读时填充了 NUL 的块。
我们在实践中观察到了这种行为:对于大多数家庭和业务应用程序,只需用
ASCII NUL 字节提供了足够的清理。

一个组织解决了以下问题
清理存储介质的是美国国防部,它创建了一个“清理和清理矩阵”12 ,为国防部承包商提供了三种政府认可的硬盘清理技术:

- 使用 I 型或 II 型消磁器消磁
- 通过分解、焚烧、粉碎、切碎或融化
- 用随机字符覆盖所有可寻址位置

演员,用字符的补码覆盖,然后验证。 （但是,作为指南

声明　全部大写　这种方法不是
批准用于对包含绝密的媒体进行消毒
信息。）

国防部的覆盖策略很奇怪,因为它不建议编写改变模式,

并且因为该方法特别未被批准用于
绝密信息。这种遗漏和限制是
几乎可以肯定是故意的。彼得·古特曼,计算机
奥克兰大学的安全研究
研究了这个问题,指出:“官方的……问题
数据销毁标准是其中的信息
可能部分不准确以试图欺骗对方
情报机构（这可能就是为什么很多
媒体消毒指南属于机密）。”13

事实上,一些研究人员一再断言,
简单的覆盖不足以保护数据免受确定的攻击者的攻击。
在 1996 年一篇极具影响力的文章中,
Gutmann 认为理论上可以检索
写入任何磁记录设备的信息会导致磁盘盘片的低级磁场模式

写入和覆盖数据的函数。作为
Gutmann 解释说,当计算机试图编写一个
1 或 0 到磁盘,媒体将其记录为这样,但
实际效果更接近于1覆盖1时获得1.05和1覆盖0时获得0.95。

尽管普通磁盘电路会将这两个值都读取为
那些,“使用专门的电路可以解决
以前的“层”包含什么。”13 Gutmann 声称
“高质量的数字采样示波器”或磁性
力显微镜 (MFM) 可用于检索
覆盖的数据。我们将此类技术称为外来技术,因为它们不依赖于标准硬盘接口。

Gutmann 提出了 22 种不同的模式,你可以
可以按顺序写入磁盘驱动器以最大程度地减少数据恢复。文章发表八年来,
一些卫生工具开发人员（例如
WIPE 项目,例如 14)采用了这些“Gutmann
模式”作为福音,并已将他们的工具编程为

表 3. 清理分类法。

等级	在哪里找到	描述
0级	常规文件	文件系统中包含的信息。包括文件名、文件属性和文件内容。经过定义,不尝试清理 0 级文件信息。 0 级还包括作为任何清理尝试的一部分写入磁盘的信息。例如,如果一个副本 Windows 95 已安装在硬盘驱动器上,试图清理驱动器,然后文件安装到 C:\WINDOWS 目录将被视为 0 级文件。没有特殊的工具需要检索级别 0 数据。
1级	临时文件	临时文件,包括打印后台处理程序文件、浏览器缓存文件、“帮助”应用程序的文件,以及回收站文件。大多数用户要么期望系统自动删除这些数据,要么甚至没有意识到它的存在。注意:0 级文件是 1 级文件的子集。经验表明,区分这个子集很有用,因为许多天真的用户在浏览计算机硬盘驱动器以查看其是否包含敏感信息时会忽略级别 1 文件。无需特殊工具即可检索级别 1 数据,尽管需要特殊培训来教操作员查看位置。
2级	已删除的文件	从文件系统中删除文件时,大多数操作系统不会覆盖其上的块写入文件的硬盘。相反,他们只是从包含目录。然后将文件的块放置在空闲列表中。这些文件可以恢复使用传统的“取消删除”工具,例如 Norton Utilities。
3级	保留的数据块	可以从磁盘中恢复但显然不属于命名文件的数据。第 3 级数据包括松弛空间、虚拟内存的后备存储和第 2 级中的信息已被部分覆盖的数据,因此无法恢复整个文件。普通的3 级数据的来源是已使用 Windows Format 命令格式化的磁盘或 Unix newfs 命令。尽管这些命令的输出可能暗示它们会覆盖整个硬盘驱动器,但实际上它们不会,并且绝大多数格式化磁盘的信息都可以通过适当的工具恢复。可以使用高级数据恢复 3 级数据 可以“取消格式化”磁盘驱动器或专用取证工具的恢复工具。
4级	供应商隐藏数据	此级别由只能使用供应商特定命令访问的数据块组成。这级别包括驱动器的控制程序和用于坏块管理的块。
5级	覆盖数据	许多人坚持认为,即使在硬盘驱动器被删除之后,也可以从硬盘驱动器中恢复信息。覆盖。我们为此类信息保留第 5 级。

煞费苦心地使用每个经过清理的磁盘上的每个模式。此外,其他组织警告说,未能使用这些模式或采取其他预防措施,例如物理销毁磁盘驱动器,意味着“某人技术知识和使用专业设备可能能够从已删除的文件中恢复数据。”15

但事实上,鉴于当前这一代的高密度磁盘驱动器,这些都可能不会覆盖模式是必要的 Gutmann 本人的观点承认。较旧的磁盘驱动器之间留有一些空间轨道;写入轨道的数据有时可以使用特殊仪器从该轨道间区域恢复。今天的磁盘驱动器有一个显着的写磁头大于读取头:因此磁道重叠,并且之间不再有任何可恢复的数据轨道。此外,今天的驱动器在正常运行时严重依赖信号处理。简单地覆盖带有一两次随机数据的用户数据可能是足以使覆盖的信息无法恢复erable Gutmann 在文章的更新版本中提出的观点,该文章出现在他的网站上（www.cryptoapps.com/~peter/usenix01.pdf）。

事实上,研究人员之间有一些共识,对于许多应用程序,用几次随机通道覆盖磁盘将足以对其进行清理。世界上最大的磁盘驱动器供应商之一 Max Tor 的工程师最近告诉我们,将覆盖的数据恢复为类似的东西“到 UFO 体验。我相信这可能是可能的……但它不会是国家安全局以外的任何人都能轻易获得的东西。”

消毒分类法

现代计算机硬盘驱动器包含各种各样的数据,包括存储在文件中的操作系统、应用程序和用户数据。驱动器还包含虚拟内存和操作系统的后备存储元信息,例如目录、文件属性和分配表。逐块磁盘驱动器检查还揭示了以前文件的残余

已删除但未完全覆盖。这些残留物有时称为可用空间,并在末尾包含字节部分填充的目录块（有时称为slack空间）,不严格属于操作系统一部分的启动软件（例如引导块）,以及在工厂初始化但从未写入。最后,驱动器还包含无法通过标准 IDE/ATA 或 SCSI 接口,包括用于坏块管理的内部驱动器块和用于容纳驱动器自己的嵌入式软件。

描述在恢复的磁盘驱动器上找到的数据和促进对消毒做法和法医的讨论分析,我们创建了一个消毒分类法（见表 3）。

表 4. 免费和市售的消毒工具样本。

程序	成本	平台	评论
高压灭菌器 http://staff.washington。edu/jdlarios/高压灭菌器	自由	自启动 电脑磁盘	只写零、DoD 规范或 Gutmann 模式。非常方便且易于使用。擦除整个磁盘，包括所有松弛和交换空间。
CyberScrub www.cyberscrub.com	39.95 美元	视窗	擦除文件、文件夹、cookie 或整个驱动器。实现古特曼模式。
DataScrubber www.datadev.com/ds100.html	1,695 美元	Windows、Unix 处理 SCSI 重新映射和交换区域。声明将在 与美国空军信息福利中心合作。	
DataGone www.powerquest.com	90 美元	视窗	擦除硬盘和可移动媒体中的数据。支持多种覆盖模式。
橡皮擦 www.heidi.ie/eraser	自由	视窗	擦除目录元数据。运行时清理 Windows 交换文件 DOS。通过创建巨大的临时文件来清理闲置空间。
OnTrack DataEraser www.ontrack.com/dataeraser	$30–$500	自启动 电脑磁盘	擦除分区、目录、引导记录等。仅包括专业版的 DoD 规格。
SecureClean www.lat.com	49.95 美元	视窗	安全地擦除单个文件、临时文件、空闲空间等。
Unishred Pro www.accessdata.com	450 美元	Unix 和 电脑硬件	了解一些用于 SCSI 驱动器坏块管理的供应商特定命令。可选地验证写入。 实施所有相关的 DoD 标准并允许自定义模式。
擦除 http://wipe.sourceforge.net	自由	Linux	使用 Gutmann 的擦除模式。擦除单个文件和随附的元数据或整个磁盘。
WipeDrive www.accessdata.com	39.95 美元	可启动的 PC 磁盘	安全擦除 IDE 和 SCSI 驱动器。
Wiperaser XP www.liveye.com/wiperaser	24.95 美元	视窗	擦除 cookie、历史记录、缓存、临时文件等。图形用户界面。

消毒工具

许多现有的程序声称可以正确消毒硬
驱动器，包括 1,695 美元的商业产品
政府认证，超过 50 种工具获得许可
对于单个计算机系统，以及似乎提供大致相同功能的免费软件/
开源产品。一般来说，有两种清理程序可用：磁盘清理器和解密
器，以及

闲置空间消毒剂。

磁盘清洁剂和解密器旨在擦除所有用户数据
在组织中处置或重新利用之前从磁盘中提取。因为覆盖操作系
统的引导
磁盘信息通常会导致计算机崩溃，
磁盘清洁剂很少在现代的启动磁盘上运行
操作系统。相反，它们通常在不受保护的操作系统（例
如 DOS）下运行，或者作为独立操作系统运行
应用程序直接从可启动媒体（软盘
光盘或 CD-ROM）。（对硬块进行消毒相对容易
不是引导盘的磁盘。例如，对于 Unix，您
可以使用设备/dev/hda清理硬盘
命令dd if=/dev/zero of=/dev/hda。）
使用我们的分类法，磁盘清理器试图擦除所有
驱动器的 1、2、3 和 5 级信息。消毒剂
具备供应商特定磁盘驱动器的知识
命令也可以擦除 4 级信息。

松弛空间清理器清理磁盘块（和部分
磁盘块）不属于任何文件且不包含
有效的文件系统元信息。例如，如果 512-
字节块保存文件的最后 100 个字节，仅此而已，
slack-space sanitizer 读取块，保留字节 1-100 未触及，并将
字节 101-512 归零。空间消毒剂
还压缩目录（删除忽略的条目），和
覆盖空闲列表上的块。其中许多程序
还删除临时文件、历史文件、浏览器 cookie、
删除的电子邮件，等等。使用我们的分类法，松弛空间
消毒剂试图擦除所有 1 级到 4 级驱动器信息，同时保持 0 级信
息完整。

表 4 提供了一些免费和商业的示例
可用的卫生工具；完整列表可在
www.fortunecity.com/skyscraper/true/882/Comparison
_碎纸机.htm。

取证工具

清理工具的另一面是取证分析工具，
用于恢复硬盘信息。取证工具比清理工具更难编写，而不是

令人惊讶的是，可用的这些工具越来越少。许多
确实存在的软件包是为执法机构量身定制的。表 5 显示了取证
工具的部分列表。

几乎所有的取证工具都可以让用户分析硬盘或

表 5. 取证程序。

程序	成本	平台	评论
DriveSpy www.digitalintel.com	$200–$250	DOS/Windows	检查闲置空间和删除的文件元数据。
EnCase www.guidancesoftware.com	2,495 美元	视窗	具有复杂的驱动器映像和预览模式、错误检查和验证，以及搜索、浏览、时间线和注册表查看器。图形用户界面。包括用于分类已知文件的哈希分析。
取证工具包 www.accessdata.com	595 美元	视窗	取证信息的图形搜索和预览，包括搜索 JPEG 图像和 Internet 文本。
ILook www.ilook-forensics.org	不适用	视窗	处理数十个文件系统。已删除文件的资源管理器界面。生成文件的哈希值。过滤功能。此工具仅适用于美国政府和执法机构。
诺顿实用程序 www.symantec.com	49.95 美元	视窗	包含用于恢复已删除文件和逐个扇区检查计算机硬盘的工具。
验尸官工具包 www.porcupine.org/forensicsl/tct.htm	自由	Unix	一组用于在入侵后对 Unix 磁盘进行事后取证分析的程序。
任务 http://atstake.com/research/tools/task	自由	Unix	对使用 dd 创建的磁盘映像进行操作。处理 FAT、FAT32、工具包。分析已删除的文件和空闲空间，包括时间线 NTFS、Novel、Unix 和其他磁盘格式。建立在验尸官的工具包。

来自各种不同操作系统的硬盘映像，并提供 Explorer 风格的界面，因此您可以阅读文件。工具当然受限于原版计算机的操作系统，因为不同的系统在写入时会覆盖不同数量的数据或元数据删除文件或格式化磁盘。尽管如此，其中许多取证工具可以找到“未删除”的文件（2 级数据）和显示不再关联的硬盘信息带有特定文件（3 级数据）。大多数工具还提供不同的搜索功能。因此，运营商可以搜索关键字或模式的整个磁盘映像，然后显示包含搜索模式。

为执法量身定制的程序还提供记录操作员在硬盘驱动器期间的每一次击键检查过程。这个特性据说可以防止证据篡改。

哦，消毒，你在哪里？

尽管有现成可用的清理工具以及提供取证分析的工具带来的明显威胁，但仍有不断有报告称，一些包含机密信息的系统正在二级市场上出售。

我们为这种状态提出了几种可能的解释事务：

- 缺乏知识。处理设备的个人（或组织）根本没有考虑问题（例如，他们可能缺乏培训或时间）。
- 对问题缺乏关注。个人认为问题，但实际上并不认为该设备包含机密信息。
- 缺乏对数据的关注。个人意识到问题 驱动器可能包含机密信息 但不关心数据是否泄露。
- 未能正确估计风险。个人知道的问题，但不相信设备的未来所有者会泄露信息（即，个人假设设备的新所有者将使用开车存储信息，不会到处翻找寻找前任主人留下的东西）。
- 绝望。个人意识到问题，但不认为可以解决。
- 缺乏工具。个人意识到问题，但没有正确清理设备的工具。
- 缺乏培训或无能。个人尝试对设备进行消毒，但尝试无效。
- 工具错误。个人使用工具，但它没有表现如宣传的那样。（例如，Linux擦除命令的早期版本有许多错误，导致数据实际上没有被覆盖。版本例如，0.13 没有擦除文件中一半的数据，因为一个错误；见 http://packages.debian.org/unstable/utils/擦除.html）
- 硬件故障。装有硬盘的计算机可能会损坏，因此无法消毒硬盘驱动器，而无需将其卸下并将其安装到其他计算机中，这是一个耗时的过程。或者，计算机故障可能会使硬盘驱动器也发生了故障，实际上它没有。

在非专业用户中,尤其是那些使用 DOS 或 Windows 操作系统的用户中,缺乏培训可能是消毒措施不佳的主要因素。

在专家用户中,我们提出了不同的解释:他们知道 Windows 格式化命令实际上并没有覆盖磁盘的内容。矛盾的是,媒体对奇异的数据恢复方法的迷恋可能会因为看起来过于繁重而降低了这些用户的净化程度。在反复采访中,用户经常会说:“FBI 或 NSA 总是可以根据需要取回数据,那为什么还要先清理磁盘呢?”由于这些未经证实的恐惧,有些人甚至没有采用基本的消毒措施。当然,这种推理是有缺陷的,因为大多数用户应该关心保护他们的数据免受更多的行人攻击,而不是美国执法和情报机构的攻击。即使这些组织确实对某些用户构成了威胁,但当今随时可用的清理工具仍然可以保护他们的数据免受其他可信威胁。

无论它们多么有趣,非正式的采访和偶尔的媒体报道都不足以衡量当前的消毒做法。为此,我们必须购买大量磁盘驱动器,并实际查看其前所有者留下的数据。

我们的实验在 2000 年 11 月
和 2002 年 8 月期间,我们在二级市场上购买了 158 个硬盘驱动器。我们从多个来源购买驱动器:专门销售二手商品的电脑商店、销售 2 到 5 个驱动器的小企业,以及销售 10 到 5 个驱动器的整合商。 20 个驱动器。

我们通过在 eBay 在线拍卖服务中赢得拍卖购买了大部分大容量硬盘。

与二级市场设备的情况一样,驱动器的制造商、尺寸、制造日期和状况各不相同。很大一部分驱动器受到物理损坏、包含无法读取的扇区或完全无法操作。

因为我们对每个驱动器的数据感兴趣,而不是它的物理恶化,所以我们的目标是尽可能减少驱动器的处理。收到后,我们将每个驱动器的物理特性和来源记录在数据库中。然后我们将驱动器连接到工作站

运行 FreeBSD 4.4 操作系统,然后使用原始 ATA 设备中的 Unix dd命令将驱动器的内容逐块复制到

我们称之为“镜像文件”的磁盘文件。完成此映像操作后,我们尝试使用多个文件系统安装每个驱动器:FreeBSD、MS DOS、Windows NT 文件系统、Unix 文件系统和 Novell 文件系统。如果我们成功挂载了驱动器,我们使用 Unix tar命令遍历整个文件系统层次结构并将文件复制到压缩的 tar 文件中。

这些文件与我们的分类标准的 0 级和 1 级文件完全相同。

然后,我们使用专门为此项目编写的各种工具分析数据。特别是,我们将每个级别 0 和级别 1 文件的完整路径名、长度和 MD5 加密校验和存储在数据库中。 (MD5 是一种单向函数,可将数据块缩减为可用于验证文件完整性的 128 位电子“指纹”。)我们可以针对该数据库运行查询以报告这些文件的发生率。未来,我们计划通过查找 MD5 冲突以及将我们的数据库与美国国立卫生研究院商业软件的 MD5 代码数据库进行比较来识别文件的唯一性。

标准和技术正在组装。16为了便于分析,我们还

创建了“取证文件系统”,这是 Gifford 及其同事首先提出的一种语义文件系统。17 FFS 让我们可以使用传统的 Unix 文件查看取证信息并对其采取行动系统工具,例如 ls、more、grep 和 strings。例如,在 FFS 中,目录列表同时显示正常文件和已删除文件;它修改已删除的文件名以防止名称冲突并指示文件的内容是否不是

可恢复、部分可恢复或完全可恢复。
(取证分析的难度很大程度上取决于用于创建目标文件系统的操作系统;特别是,在 FAT 格式的格式化磁盘上恢复文件比在大多数 Unix 文件系统上要容易得多。)

初步发现我们总共获得了
75 GB 的数据,其中包括 71 GB 的未压缩磁盘映像和 3.7 GB 的压缩 tar 文件。

从一开始,这个项目最有趣的一个方面就是磁盘驱动器的变化。
当我们向人们介绍我们最初的项目计划时,许多人回应说他们肯定收集到的绝大多数驱动器将是X,并且 X 的值因扬声器而异。例如,有些人“肯定”所有恢复的驱动器都包含活动文件系统,而另一些人则确信所有驱动器都将被重新格式化。有些是

肯定我们会找到数据,但它太旧而没有意义,而其他人则确信几乎所有的驱动器都会得到适当的清理,“因为没有人会愚蠢到丢弃包含活动数据的驱动器。”

文件系统分析即使是这种有
限的初步分析结果也表明行业中没有标准做法。在我们成功成像的 129 个驱动器中,只有 12 个 (9%)通过用零填充块完全覆盖它们的扇区进行了适当的清理; 83 个驱动器 (64%) 包含可挂载的 FAT16 或 FAT32 文件系统。 (我们收集的所有驱动器都有 ei

表 6. 按类型划分的可恢复级别 0 和 1 文件。

文件类型	找到号码	在驱动器上	每个驱动器的最大文件数
微软 Word (DOC)	675	23	183
展望（太平洋标准时间）	20	6	12
微软 PowerPoint (PPT)	566	14	196
微软写 (WRI)	99	21	19
微软工厂 (WKS)	68	1	68
微软 Excel (XLS)	274	18	67

其他 FAT16 或 FAT32 文件系统。)另外 46 个驱动器
没有可挂载的文件系统。

在具有可挂载文件系统的 83 个驱动器中,有 51 个似乎是新格式化的 也就是说,它们要么没有文件,要么文件是由

DOS格式 c:/s命令;另外六个驱动器是
格式化并安装了 DOS 或 Windows 3.1 的副本。在这 51 个驱动器中,19 个具有可恢复的 3 级
数据 表示驱动器已格式化
在它们被用于另一个应用程序之后。

在我们无法安装的 46 个驱动器中,有 30 个有更多
超过一千个可恢复的 3 级信息扇区
化。其中许多驱动器具有可恢复的 FAT 目录
条目也是如此。

文档文件分析

我们对可挂载文件进行了有限的分析
系统来确定留在文件的类型
驱动器。表 6 总结了这些结果。

总体而言,具有活动文件系统的 28 个驱动器包含的文档文件相对较少 远少于
我们希望在经常使用的个人电脑上找到。
我们认为这是因为驱动器的先前所有者
故意删除这些文件,试图至少
在丢弃之前对驱动器进行部分消毒。

为了测试这个理论,我们编写了一个程序,让我们扫描 FAT16 和 FAT32 图像以查找已删除的文件和目录。使用这个程序,我们可以扫描磁盘
可能被驱动器的原始数据删除的数据
处置驱动器之前的所有者。结果很亮眼:除了已清除的磁盘（所有

块归零）,几乎每个磁盘都有显着
已恢复的已删除目录和文件的数量
有能力的。即使是包含许多未删除文件的 28 个磁盘也包含大量已删除但可恢复的文件
目录和文件也是如此。仔细检查
已删除的文件表示,一般来说,用户删除了数据文件,但保持应用程序文件完整。

恢复的数据

目前,我们可以使用tar文件来恢复 Level 0 和 1 级文件。我们在这些中找到的一些信息
文件包括:

- 关于人事问题的公司备忘录
- 一封致 7 岁儿童医生的信
  孩子的父亲抱怨治疗
  孩子的癌症不令人满意
- 加利福尼亚儿童医院的传真模板（我们
  预计对该驱动器的额外分析将产生
  医学敏感信息）
- 情书
- 色情

使用稍微复杂的技术,我们
编写了一个扫描信用卡号码的程序。这
程序搜索数字字符串（可能
空格和破折号分隔符)通过了所有信用卡号所需的 mod-10 校验位测试,并且还
属于信用卡号的可行数值范围内
范围。例如,没有主要信用卡号开头
与八。

在我们的研究中,42 个驱动器的数字通过了这些
测试。确定一个数字是否真的有效
信用卡号需要尝试交易
信用卡网络。我们没有这样做,而是检查了数字的上下文。包含两个驱动器
一致的财务风格日志文件。这些驱动器之一
(#134) 包含 2,868 个日志格式的数字。之上
进一步检查,这个硬盘似乎是
最有可能用于伊利诺伊州的 ATM 机,并且
没有努力消除任何驱动器的财务
信息。日志包含帐号、日期
访问权限和帐户余额。此外,硬
驱动器拥有所有 ATM 机软件。虽然
该驱动器还包含程序和软件
更改 ATM 的 DES 密钥（这可能会确保
ATM与金融网之间的交易
工作）,实际的 DES 密钥显然存储在 ATM 机的硬件芯片中。

另一个驱动器 (#21) 包含 3,722 张信用卡
不同类型日志中的数字（其中一些重复）

表 7. 磁盘格式化结果。

磁盘大小	积木	由 WINDOWS 98 更改的块	由 WINDOWS 98 更改的块
		磁盘命令	格式化命令
10 GB	20,044,160 2.563 (0.01%)		21,541 (0.11%)

格式。此驱动器上的文件似乎已
擦除,并且驱动器被格式化。另一个驱动器
(#105) 在数据库文件中包含 39 个信用卡号
其中包括正确类型的信用卡,还有一个（#133)在缓存的 Web 中有信用卡号
页面网址。 URL 是一个“GET”类型的 HTTP 表单,
被提交到电子商务网站;它包含所有
执行电子商务交易所需的地址和到期信息。最后,另一个驱动器

(#40) 在一个文件中有 21 个信用卡号。

我们还编写了一个搜索 RFC 邮件的程序
标题。在分析的 129 个驱动器中,66 个驱动器具有更多
超过五封电子邮件。我们使用这个阈值是因为
一些程序,例如 Netscape Navigator,包括一个
安装时很少有欢迎电子邮件。在我们的一个驱动器
批次包含近 9,500 封电子邮件,日期为
从 1999 年到 2001 年。总共有 17 个驱动器有超过 100 个
电子邮件和大约 20 个驱动器之间有 20 个
和 100 封电子邮件。在此分析过程中,我们仅调查了邮件的主题标头;内容
似乎从典型的垃圾邮件到抱怨
追溯工资。

了解 DOS 格式

不清楚这52个格式化驱动器是否被格式化
清理数据,或者它们是否被格式化以确定它们的状况和在 sec 上的销售价值
市场。

在许多采访中,用户表示他们相信 DOS
和 Windows格式命令将正确删除所有硬盘驱动器数据。这种信念似乎是合理的,因为
DOS 和 Windows格式命令
警告用户“所有数据都不可移除
磁盘驱动器 C:将丢失”当计算机
从软盘启动,用户尝试使用C 格式:
命令。这个警告可能被正确地视为一个承诺
使用format命令实际上会删除所有
磁盘驱动器的数据。

当我们告诉许多用户时,他们感到很惊讶
格式化命令不会擦除所有磁盘
形成。正如我们的分类所示,大多数运营
系统格式命令只写一个最小的磁盘文件
系统;他们不会重写整个磁盘。为了显示
这个断言,我们拿了一个 10-Gbyte 的硬盘,装满了
每个具有已知模式的块。然后我们初始化了一个
使用 Windows 98 FDISK命令进行磁盘分区
并使用format命令格式化磁盘。后
每一步,我们检查磁盘以确定数量
已写入的块。表 7 显示了结果。

用户可能会发现这些数字令人沮丧:尽管
来自操作系统的警告相反,
format命令仅覆盖磁盘数据的 0.1% 以上。然而,该命令需要

在 10-Gbyte 上完成工作需要 8 多分钟
磁盘　给人的印象是计算机实际上是
覆盖数据。事实上,计算机正在尝试
读取驱动器的所有数据,以便构建坏块
桌子。期间实际写入的唯一块
格式化过程是那些对应于引导
块、根目录、文件分配表和
很少有测试扇区散布在整个驱动器的表面。

数量与磁盘驱动器的质量相比

虽然 150 个磁盘驱动器看起来很多,但它很小
每年都会被出售、重新利用和丢弃。因此,
我们的发现和统计数据必然是定性的,而不是
定量的。不过,我们可以得出一些结论。

首先,人们可以删除机密信息
在丢弃、重新利用或出售之前从磁盘驱动器中提取
他们在二级市场上。此外,免费提供
工具使磁盘清理变得容易。

二、目前对“病历”的定义
可能不够广泛,无法涵盖家庭和工作环境中的医疗敏感信息范围。例如,我们在计算机上发现了包含医学敏感信息的个人信件

以前属于一家软件公司。许多
常规电子邮件还包含医学敏感信息
不应披露的信息。如果员工
给老板发信息说他会错过会议
因为他有一个需要去看医生的具体问题,
例如,他在公司电子邮件系统中创建了他的医疗状况记录。

第三,我们的研究表明,二级硬盘市场几乎肯定充斥着以下信息:

既敏感又保密。

根据我们的调查结果,我们提出以下建议:

- 用户必须接受有关正确技术的教育
  清理磁盘驱动器。
- 组织必须采取适当清理计算机系统和存储介质上的驱动器的策略,

  被出售、销毁或重新利用。
- 操作系统供应商应包括系统工具

安全地删除文件,并清除闲置空间和整个磁盘驱动器。

· 未来的操作系统应该能够自动清理已删除的文件。它们还应该配备自动清理操作系统当前未使用的磁盘扇区的后台进程。 · 供应商应鼓励使用加密文件系统,以尽量减少数据清理问题。 · 磁盘驱动器供应商应为其驱动器配备工具,以快速甚至即时删除所有磁盘驱动器信息。例如,他们可以为磁盘驱动器配备一个加密子系统,该子系统在写入块时自动加密每个磁盘块,并在读取块时解密块。然后,用户可以通过安全擦除密钥使驱动器的内容变得难以理解。 18

通过几个月的工作和相对较少的财务支出,我们能够检索到数以千计的信用卡号码和许多人的非常个人信息。我们认为,缺乏有关此问题的媒体报道仅仅是因为在这一点上,很少有人希望重新利用硬盘驱动器来获取机密材料。如果清理实践没有得到显着改善,那么重新利用硬盘驱动器上的机密信息被个人和组织利用对我们造成伤害只是时间问题。

□

致谢许多麻省理工学院的学生和教职员工对这个项目提供了有用的评论和见解。我们特别感谢 David Clark 和 Ron Rivest 教授对本文先前草稿的持续支持、建议和评论。 Hal Abelson 和 Charles Leiserson 教授也一直是鼓励和精神支持的源泉。我们收到了 Brian Carrier、Peter Gutmann、Rich Mahn、Eric Thompson 和维策维尼玛。

参考

1. Network Associates， PGP Windows 95、98 和 NT 用户指南,版本 6.0。 1998; 6.02 版包括 pgpdisk 加密文件系统,可从 www.pgpi.org/products/pgpdisk 下载。

2. M. Blaze，“适用于 Unix 的加密文件系统”,第一届 ACM 会议。通讯。和计算安全， ACM 出版社,纽约,1993 年,第 9-16 页。

3. 微软，“Windows 2000 的加密文件系统”,www.microsoft.com/windows2000/techinfo/howitworks/安全/加密.asp。

4. J. Hasson，“VA 在 PC 处理失误后加强安全性”,《联邦计算机周刊》， 2002 年 8 月 26 日； www.fcw.com/fcw/articles/2002/0826/news-va-08-26-02.asp.

5. M. Villano，“硬盘魔术:让数据消失”永远，”纽约时报， 2002 年 5 月 2 日。

6. J. Lyman，“麻烦的网络公司可能会暴露机密客户数据”， NewsFactor Network， 2001 年 8 月 8 日;万维网。newsfactor.com/perl/story/12612.html。

7. J. Markoff，“使用过的电脑中出现患者文件”，纽约时报， 1997 年 4 月 4 日。

8. S. Berinato，“物美价廉”， CIO， 2002 年 10 月 15 日,第 53-59。

9. 国家计算机安全中心，“A Guide to Understanding Dataremanence in Automated Information System”,图书馆 No. 5-236,082,1991,NCSC-TG-025； www.radium.ncsc.mil/tpep/library/rainbow/NCSC-TG 028.ps

10.加利福尼亚诉格林伍德案， 486 US 35,1988 年 5 月 16 日。

11. Microsoft，“Microsoft Extensible Firmware Initiative FAT32 File System Specification”,2000 年 12 月 6 日;万维网。microsoft.com/hwdev/download/hardware/fatgen103.pdf。

12. 美国国防部，“清洁和消毒矩阵”,DOS 5220.22-M,华盛顿特区，1995 年； www.dss.mil/isec/nispom_0195.htm。

13.P。 Gutmann，“从磁性和固态存储器中安全删除数据”， Proc。第六次 Usenix 安全 Symp.， Usenix Assoc.,1996； www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html。

14.T。 Vier，“擦除 2.1.0”,2002 年 8 月 14 日； http://sourceforge。净/项目/擦除。

15.D。 Millar，“在出售/捐赠之前清理旧电脑”,1997 年 6 月； www.upenn.edu/computing/security/advisories/oldcomputers.html。

16.美国国家标准与技术研究院,《国家软件参考库参考数据集》； www.nsrl.nist.gov。

17.DK Gifford 等人，“语义文件系统”， Proc。第 13 届 ACM 研讨会。关于操作系统原理， ACM 出版社,1991 年,第 16-25 页。

18.G. Di Crescenzo 等人，“如何忘记秘密” ,计算机科学理论方面的专题讨论会(STACS 99),计算机科学讲义,Springer-Verlag,柏林，1999 年,第 500-509 页。

Simson L. Garfinkel是 MIT 计算机科学实验室密码学和信息安全组以及高级网络架构组的研究生。 Garfinkel 是许多关于计算机安全和策略的书籍的作者,包括Database Nation: the Death of Privacy in the 21st Century (O Reilly, 2000)和Practical UNIX and Internet Security (O Reilly, 2003)的合著者。他目前的研究兴趣集中在安全技术和可用性的交叉领域。通过 simsong@lcs.mit.edu 联系他； http://simson.net。

Abhi Shelat是麻省理工学院计算理论组的研究生。他的研究兴趣包括计算机安全、算法和数据压缩。他还喜欢拍照和制作家具。通过 abhi@lcs.mit.edu 联系他； http://理论。 lcs.mit.edu/~abhi。