一种数字光盘数据保密方法①

刘昌平, 刘 洋, 陈佳实

(广东宏景科技有限公司, 广州 510663)

摘 要: 提出了一种支持数据保密的光盘文件系统以及适用于数字光盘的数据保密方法. 改进了 ISO 9660 Level 2 标准的光盘文件系统, 划分为明文区、保密区和参数区三部分, 分别存储明文、密文和参数数据. 光盘制作模块将明文数据、密文数据及参数数据制作成光盘文件系统的镜像文件, 同时将光盘阅读模块注入到明文区. 光盘阅读模块专门存取保密区数据. 该方法能实现明文、密文数据及其解密工具并存于同一光盘, 无需安装额外的软件, 该文件系统兼容目前所有的光盘设备.

关键词: 加密; 光盘; 文件系统; 磁盘分区

New Method of Data Protection for Digital Optical Disc

LIU Chang-Ping, LIU Yang, CHEN Jia-Shi

(Gangdong Gloryview Technology Co. Ltd., Guangzhou 510663, China)

Abstract: A new file system of data protection for optical disc and its data protection method were proposed. This file system was divided into plaintext section, ciphertext section and parameter section. A producer of file system was suggested to pack plaintext, ciphertext, parameter in one ISO image file. A special reader for accessing ciphertext section was injected into plaintext section. This method bound plaintext, ciphertext and the special reader together, and needn’t any extra software. This file system is compatible with all optical disc devices.

Key words: encryption; optical disc; file system; disk partition

移动存储介质通常包括 SD 卡、Flash 存储器、小尺寸硬盘以及光盘等. 磁媒介存储介质支持读、写操作, 通常采用 FAT32、NTFS 或者 EXT2 文件系统. 光媒介存储介质, 例如 CD-ROM、DVD-ROM、CD-RW 等, 通常采用 ISO 9660[1]、Joliet 或者 UDF(Universal Disc Format)[2]文件系统. ISO-9660 目前有 Level1 和 Level2 两个标准[1], 其中 ISO 9660 Level2 支持长文件名, 是目前应用最广的光盘文件系统, 适用于各种类型的计算机.

因其可移动性、便携性和兼容性, 数字光盘存储介质得到广泛应用, 例如数据备份、文件归档、软件发布和音像制品等用途, 同时也造成了信息安全上的漏洞, 例如非法复制、未经授权的存取而导致的信息泄露, 对涉密数据的机密性构成安全威胁. 针对这个问题, 学术界和产业界开展了数字光盘数据保密研究及其产品开发. DVD 联盟按地理区域将全球划分为 6 个区[3], 光盘存取设备的区域码必须与数字光盘的区域码一致或者兼容才能读取光盘的数据, 主要用于保护光盘音像制品. 这种方法保护光盘数据的作用不大, 很多光盘存取设备支持全区解码, 而且在同一地理区域内, 便失去了数据保密的作用.

一种代表性方法是自定义光盘存储格式, 使之与专用设备相关[4]. 这种自定义光盘存储格式与 ISO 9660、UDF 等文件系统完全不兼容, 通用的光盘驱动器不能识别这种存储格式, 仅专用的存取设备或者装置才能解码光盘涉密数据. 这种方法对硬件设备的依赖性过高, 光盘数据保密依赖于硬件设备的保密, 硬件设备构成新的安全脆弱点, 数据保密成本较高, 适应于涉密级别高(例如政府、军队等场所)的应用, 通用性较差.

① 基金项目:广东省科技厅 2012 年度重大科技专项资助项目(2012A080102004)

收稿时间:2013-05-27;收到修改稿时间:2013-07-03

其他数字光盘数据保密的方法是加密数据文件, 以通用光盘文件系统的格式(例如 ISO 9660)写入光盘存储介质[5,6]. 这种方法通常仅能保密文件数据, 而目录数据以明文形式存储, 不在保密范围. 此外, 这种方法需要在计算机上安装专用的加、解密工具, 才能读取数字光盘的涉密数据, 工具缺失(例如病毒破坏、误删除)或者版本变动将直接导致无法读取光盘数据, 造成数据存取障碍. 肖飞等提出了一种光盘数据保密方法[7], 其原理是修改 ISO 9660 文件系统的部分标志位, 使得光盘驱动器无法读取受保护的文件, 属于信息隐藏范畴, 其本质与上述方法相同.

本文提出了一种支持数据保密的光盘文件系统及其数字光盘数据保密方法. 全文组织如下: 第 1 节阐述改进的光盘文件系统, 第 2 节阐述数据保密方法, 第 3 节是光盘文件系统及其方法的实验与分析, 最后第 4 节总结全文.

1 改进的光盘文件系统

本文在 ISO 9660 Level 2 文件系统的基础上加以改进, 提出了一种支持数据保密的光盘文件系统. 改进后的光盘文件系统由明文区、保密区和参数区构成, 其存储格式见图 1 所示.

图 1　光盘文件系统格式

1.1 明文区

图 1 所示的明文区用于存储明文文件和光盘阅读模块, 光盘阅读模块将在第 2 节加以阐述. 明文区存储格式采用标准的ISO 9660 Level 2 文件系统格式, 重定义了 ISO 9660 Level 2 的光盘容量标志, 阐述如下.

ISO 9660 Level 2 文件系统的主描述符 PVD (Primary Volume Descriptor)[9]位于光盘起始地址的第 16 扇区处, PVD 的偏移地址 0x50(光盘偏移地址 0x8050)处是光盘容量标志 VSS(Volume Space Size), 长度为 8 字节, 表示整个光盘文件系统的容量大小, 单位是扇区, 每扇区的大小是 2048 字节. 如果遵循 ISO 9660 标准, VSS 设置为图 1 的明文区容量大小, 那么光盘驱动器无法读取明文区以外的区域. 因此, 本文将 VSS 重定义为明文区、保密区和参数区的总容量大小, 使得光盘驱动器能够访问明文区以外的区域.

1.2 保密区

保密区用于存储涉密数据, 采用标准的 ISO 9660 Level 2 文件系统格式. 与明文区不同之处是, 保密区所有的数据均是密文, 包括保密区的 PVD 和所有数据空间. 包括 Windows、Linux、Unix 在内的所有操作系统及光盘读写软件都无法解析保密区的数据.

保密区数据采用 AES 加密算法, 明文区和保密区都不存储加密密钥. 用户在图 1 所示的光盘阅读模块中输入用户登录密码, 根据该登录密码采用 SHA1 算法生成 AES 加密算法的密钥.

1.3 参数区

参数区对用户是不可见的, 大小固定为 2048 字节, 存储光盘的说明性参数, 包括密码的 MD5 值、光盘序列号、光盘文件系统检测标记、保密区大小, 见图 2 所示, 各参数的含义说明如下.

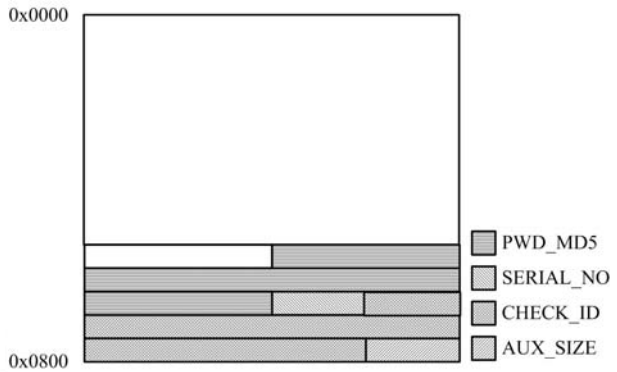

图 2　参数区格式

1) PWD_MD5: 密码的 MD5 值, 宽度为 256 位, 位于参数区偏移地址 1976 字节处, 根据用户设置的密码生成 MD5 值, 用于用户身份鉴别;

2) SERIAL_NO: 光盘序列号, 宽度为 32 位, 位于参数区偏移地址 2008 字节处, 光盘文件系统的唯一性标识符号;

3) CHECK_ID: 光盘文件系统检测标记, 宽度为 256 位, 位于参数区偏移地址 2012 字节处, 是光盘文件系统的特征码, 光盘阅读模块依据此参数判断光盘文件系统的合法性;

4) AUX_SIZE: 保密区模块大小, 宽度为 32 位, 位于参数区偏移地址 2044 字节处, 表示保密区模块的

容量大小, 单位是扇区.

2 数据保密方法

本文改进了ISO 9660 Level 2文件系统, 划分为明文区、保密区和参数区, 因此需要一个专用工具制作该文件系统的光盘镜像文件; 另外, Windows 等操作系统及光盘软件无法访问保密区, 因此还需要一个专用工具, 用于身份鉴别和解析保密区数据. 本节阐述改进的光盘文件系统用于数据保密的方法.

2.1 光盘制作模块

光盘制作模块是一个工具软件, 制作第 1 节所述光盘文件系统的镜像文件. 光盘制作模块用一个二叉树表示 ISO 9660 Level 2 目录树数据结构, 每个文件(或文件夹)用一个四元组表示, 见图 3 所示, 四元组含义说明如下.

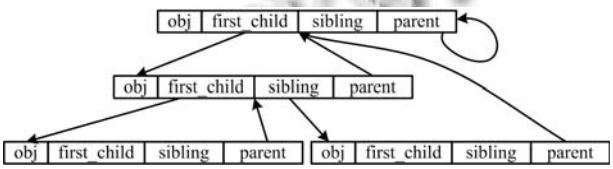

图 3 目录树

1) obj: 表示文件(或文件夹)对象, 包括名称、大小、创建日期、访问路径以及其他描述性属性;

2) first_child: 对于文件夹, 该元素表示文件夹的第一个孩子对象, 是该文件夹下的文件或者子文件夹;

3) sibling: 表示该对象的下一个兄弟结点, 是同一父文件夹下的文件或者子文件夹;

4) parent: 表示双亲对象, 对于根节点, 该元素指向自己, 见图 3 所示.

第 1 节所述光盘文件系统存在 2 个独立的分区, 即明文区和保密区, 分别为每个分区创建如图 3 所示的目录树.

根据 ISO 9660 标准, 存在 2 种途径定位一个文件或文件夹. 第一种是按照 ISO 9660 的数据存储结构, 从根目录开始逐级遍历每一级目录, 直至找到指定的文件或文件夹; 第二种途径是搜索 ISO 9660 的路径列表[9]数据块. 后者的对象定位速度快, 不需要遍历整个光盘文件系统, 但是按照 ISO 9660 标准, 路径列表不是必需的数据块.

相应地, 创建图 3 所示的目录树, 也存在上述 2 种途径, 可以根据实际情况选择其中一种. 很显然, 创建目录树是光盘制作模块中很费时、费内存空间的部分, 本文第 3 节将对这部分进行实验与分析.

2.2 光盘阅读模块

光盘阅读模块是一个工具软件, 对用户进行身份鉴别, 解密保密区的密文数据. 2.1 节所述光盘制作模块具有光盘镜像文件的编辑功能, 因此从本质上讲, 光盘阅读模块是光盘制作模块的一部分, 实现光盘制作模块的部分功能, 仍采用图 3 所示的目录树结构, 不再赘述. 该节阐述光盘阅读模块解密保密区涉密数据的主要流程, 见图 4 所示, 阐述如下.

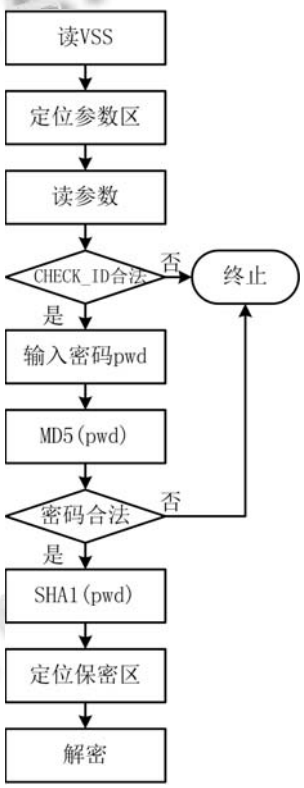

图 4 解密流程

1) 读取如图 1 所示的光盘容量标志 VSS;

2) 根据 VSS 定位参数区模块的偏移地址;

3) 读取参数区模块的说明性参数;

4) 检查光盘文件系统的特征码 CHECK_ID;

5) 如果 CHECK_ID 合法, 请求输入密码 pwd;

6) 依据用户的输入, 生成密码的 MD5 值, 即 MD5(pwd);

7) 与 PWD_MD5 对比, 鉴别用户的合法性;

8) 生成密码的 SHA1 值, 即 SHA1(pwd);

9) 定位保密区偏移地址, 保密区偏移地址计算公

式为: 保密区偏移地址=VSS-保密区大小-参数区大小;

10) 以 SHA1(pwd)为密钥解密保密区的密文数据.

2.3 数据保密方法

数字光盘数据保密方法见图 5 所示, 由光盘制作模块、光盘阅读模块和光盘文件系统构成. 光盘阅读模块首先内嵌在光盘制作模块中, 光盘制作模块将明文数据、涉密数据打包制作成光盘镜像文件的同时, 将光盘阅读模块注入到光盘文件系统的明文区, 见图 1 和图 5 所示. 生成明文区数据时, 按时 ISO 9660 Level2 标准将明文区数据写入到 ISO 文件中, 写入 VSS 标志时, 按照 1.1 节所述写入明文区、保密区和参数区的总容量大小; 生成保密区数据时, 按照 ISO 9660 Level2 标准写入 AES 算法加密后的密文数据; 最后将参数区数据以明文形式写在 ISO 文件的末尾.

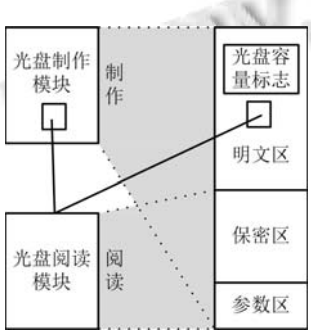

图 5　数据保密方法

当光盘镜像文件刻录到光盘介质后, 光盘阅读模块随之一并刻录进光盘介质, 与其他明文数据、密文数据并存于同一光盘介质, 且不可删除, 不可更改, 也不会感染病毒. 在计算机上运行光盘阅读模块, 即可按图 4 所示流程访问光盘的保密区数据.

本文方法的数据保密性在于保密区数据是加密后的密文数据, 并且光盘阅读模块是唯一的保密区访问工具, 光盘阅读模块解密密文数据之前进行用户身份验证(见图 4 所示), 仅合法用户才能访问保密数据.

3　实验与分析

本文在 Windows 7 平台上实现了光盘制作模块和光盘阅读模块, 软件开发语言是 Microso Visual C++ 2008 SP1. 首先编译生成光盘阅读模块, 将该模块的 EXE 文件作为光盘制作模块的资源文件, 共同编译生成光盘制作模块的 EXE 文件. 光盘制作模块生成 ISO 文件时, 向 ISO 文件释放出光盘阅读模块的 EXE 文件. 两个模块的人机交互界面见图 6 和图 7 所示.

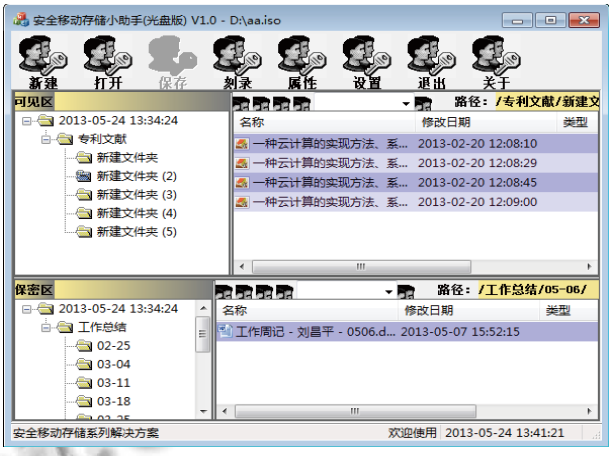

图 6　光盘制作模块

图 7　光盘阅读模块

按图 6 所示生成 ISO 文件并刻录成数字光盘后在 Windows 浏览器中打开光盘, 仅明文区(即图 6 所示的可见区)的文件是可直接访问的, 保密区的文件不可直接访问, 必须通过光盘阅读模块(即图 8 所示的 Reader.exe)才能访问.

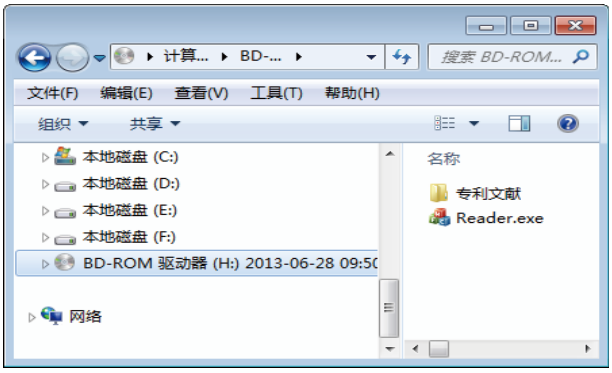

图 8　数据保密

创建图 3 所示目录树将增大整个方法的时间和空间复杂度, 尤其是在文件数量很大、单个文件较小的情况下. 本节选择了 2 个测试用例来测试该方法的时间和空间复杂度, 测试用例及实验结果见表 1 所示.

表 1 测试用例及结果

序号	文件个数	目录个数	时间(s)	空间(M)
1	53246	6713	55	30.4
2	8216	1563	7	5.0

测试用例 1 的文件和目录个数累计 59959 个对象, 累计大小是 2.33GB, 文件平均大小是 40.7KB, 创建图 3 所示的目录树, 耗时 55 秒, 空间消耗约 30MB. 对于容量为 4.7GB 的 DVD 光盘而言, 如果文件平均大小相近, 那么创建一张 DVD 光盘的目录树, 耗时约 110 秒, 空间消耗约 60MB. 如果文件平均大小增大, 那么性能消耗将降低, 例如测试用例 2, 累计 9779 个对象, 累计大小是 15.3G, 文件平均大小是 1.6MB, 时间和空间消耗明显降低.

4 结语

本文提出了一种改进的光盘文件系统以及数字光盘数据保密的方法, 将光盘的存储空间划分为明文区、保密区和参数区, 将保密区的阅读工具绑定在明文区. 该方法创建的数字光盘兼容所有的光盘驱动器, 与设备无关, 无需安装额外的工具软件, 使用简便, 保密成本低. 如何安全、便捷地打开保密区涉密文件是本文的后续研究工作.

参考文献

1 Tanenbaum AS. Modern Operating Systems. 3rd., New jersey: Prentice Hall, 2009. 175−178.

2 白兆华.基于 UDF 文件系统的蓝光媒体操作软件的研究与实现[硕士学位论文].西安:西安电子科技大学,2009.

3 苏斌.DVD 区域码及区域码管理.实用无线电,2000,5: 49−49.

4 张卫民.一种对光盘数据加密的方法及装置.发明专利, 201010154843.9,2010.

5 王年华.一种光盘数据加密方法.发明专利,200910194378.9,2009.

6 杨耀东.光盘授权播放内容加密算法研究[硕士学位论文].武汉:华中科技大学,2011.

7 肖飞,王运琼,李映松,李必谨.基于光盘映像文件的 CD-ROM 数据加密与解密方法.计算机科学,2009,36(5): 299−301.

(上接第 161 页)

6 结语

本文使用常用的聚类算法在校园一卡通的消费数据上进行了实验. 通过对多个指标的分析, 得出了 K-means 算法最合适在当前的数据上做聚类分析的结论. 该结论将对以后进一步挖掘和应用现有的一卡通消费数据具有十分重要的指导意义. 而文献[4]中使用 K-means 算法在一卡通数据上得出的良好效果也从另一个角度验证了结论的正确性.

参考文献

1 张兵兵,王建,张建威,等.数据挖掘在校园一卡通系统中的应用初探.数理医药学杂志,2009,22(5):572−575.

2 罗华群,易国平.校园一卡通数据的挖掘与应用.科技信息, 2010,(1X):41−41.

3 李珊娜.基于校园一卡通平台的数据挖掘应用研究.铁路计算机应用,2010,(6):55−58.

4 徐剑,陈劲舟.数据挖掘在校园一卡通数据的应用与研究.电脑知识与技术,2012,33.

5 陈慧萍,林莉莉,王建东,等.WEKA 数据挖掘平台及其二次开发.计算机工程与应用,2008,44(19):76−79.

6 魏丽.数据挖掘中聚类算法比较研究.电脑知识与技术, 2007,21.

7 Sharma N, Bajpai A, Litoriya MR. Comparison the various clustering algorithms of weka tools. facilities, 2012, 4: 7.

8 Krishna P, Senguttuvan A, Swarna T, Latha D. Clustering on Large Numeric Data Sets Using Hierarchical Approach: Birch. Global Journal of Computer Science and Technology, 2012, 12(12-C).