数字版权管理中 eBook 安全机制研究*)

邓 珂 邢春晓 周立柱

(清华大学计算机科学与技术系 北京100084)

摘 要 DRM 系统的安全性依赖于其信任组件的安全性,该信任组件处于非安全的环境,其安全模型与基于互联网的系统的安全模型有着本质的区别。本文讨论了当前 DRM 系统的信任组件所采用的安全技术,并着重分析了基于 DRM 的电子书管理策略和信息组件模型的开发,提出了一个基于机器硬件指纹识别的 DRM 技术方案,该模型的实现将对数字图书馆的构建具有重要的实际意义。

关键词 DRM,安全模型,信任组件,硬件指纹识别

Study on Security Management Mechanism of eBook in DRM

DENG Ke XING Chun-Xiao ZHOU Li-Zhu

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract The security of the DRM system is depended on the security of its trust components, which is run in unsafe circumstance. Its security model is essentially different from the system based on Internet. We discuss the technology used by the confidential component of current DRM system, and emphasize on analyzing the development of eBook manage strategy and confidential components, which are based on DRM. This paper proposes a new method for DRM Trust Component based on the algorithm of hardware fingerprint recognition. The implement of this method will pave the way for the construct of the digital library.

Keywords DRM, E-book, Security model, Trust component, Hardware fingerprint recongnition

1 引言

20世纪90年代末,数字化技术的飞速发展,使得大量的书籍资料转化为电子书的形式成为可能;而互联网的普及也为电子书的发展起了巨大的推进作用,它使得电子书的复制和传播变得越来越方便和迅速。但同时也带来相应的版权保护问题,即如何防止电子书的非授权复制和传播,这是每个出版商所必须解决的问题。提供这种对数字内容进行版权管理控制的技术就是我们通常所说的数字版权管理技术(Digital Rights Management, DRM)[1]。

在数字图书馆系统的构建过程中,我们需要解决下面两个重要问题:

·如何将电子书安全地交到客户手中。

·如何防止客户得到的电子书被非法(未经授权)复制和传播。

第一个问题,是怎样安全传递数字内容的问题。第二个问题,是怎样解决远端控制数字内容的问题。这里的远端控制,不是指网络上的远程控制,而是指在没有服务器参与的情况下,在客户端实行的控制。对于第一个问题,已经存在很多理论成熟、算法健壮的解决方案,但对于第二个问题,至今还没有理论上可行的解决方案。这主要是因为用户机器的软硬件环境的复杂性造成的。由于电子书的用户绝大多数是 PC 用户,因此我们的分析以客户环境 PC 机为主。众所周知,PC 机是一个开放的体系,从 CPU、内存到存储设备(如硬盘)等都是公开的,任何一个具备一定知识的人都可以利用某种工具软件察看内存中的数据,硬盘上的文件,和当前 CPU 的状态。在这种运行环境下,为了保护数字内容,就必须采用加密技术。这里需要指出的是,仅仅采用加密技术还远远提供不了足够的保护。虽然一个健壮的加密算法能够做到在只知道密文的情况下,即使知道算法的细节,攻击者也不可能有效地推算出明文(至少在一段时间内),但对于 DRM 的情形来说,如图1所示,黑框里的部分都会受到攻击者的攻击,而加密算法本身却不解决解密密钥的保护问题[2]。可见 DRM 技术不仅仅是一种加密技术。

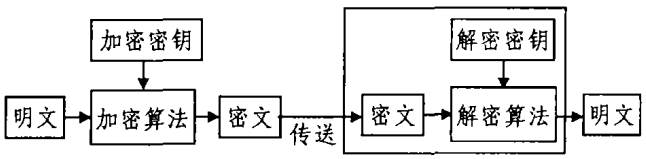

图1 加密机制

DRM 系统保护的一般是知识产权比较敏感的东西,DRM 系统的破解将会给数字内容的所有者带来重大损失。这几年黑客技术和逆向软件工程的飞速发展,使得我们意识到,一个有效的 DRM 方案必须基于如下假设,那就是攻击者了解有关该系统加密算法的一切知识,除了某些密钥。因此说,如何保证解密密钥的安全问题,对一个 DRM 系统来说是至关重要的。

2 DRM 安全模型的介绍

到现在为止,很多研究机构和公司都提出了各自的 DRM 方案,但是他们的方案大多都是针对系统整体框架设计和用户权益语言(如 Xrml 和 ODRL)的定义。DRM 系统的重要功能是实现数字对象权益的安全和管理,在具体实现上,大致上分为两种情形,一种是在 DRM 系统中实现专门的组件

*)本课题研究得到了"973"国家重点基础研究项目(No. G1999032704)的资助,同时得到国家自然科学基金基金(No. 60221120146)的资助。邓 可 硕士研究生,主要研究方向:数字版权管理,电子书的安全技术等;邢春晓 博士,副教授,主要研究方向:数据库技术、海量信息管理、数字图书馆等。

Privacy Engineering[2]来处理权益，另一种是将数字对象的权益管理从 DRM 系统中分离出来，实现专门 PRM(Privacy Rights Management)[3]，通过 FPM Server 来处理。不足之处是这些方案的权益管理控制都是在服务器端实现的，并没有考虑在客户端实现权益管理的情形。这两种情形下，DRM 系统的工作模式有所不同，我们称之为系留模式和非系留模式，下面我们将给出具体说明。

2.1 系留模式和非系留模式

系留模式(Tethered)和非系留模式(Untethered)是 DRM 系统的两种模式。在系留模式下，解密密钥是保存在服务器上，并不与数字内容存放在一起。只有当数字内容被解析时，解密密钥才从服务器传递到客户机上。而在非系留模式下，解密密钥不是从服务器端获得的，而是在用户端获得。这样，用户可以随时访问数字内容而不需要访问远端服务器。

大多数 DRM 系统都工作在系留模式下，身份认证是该模式下解决安全问题的有效手段，而且也有很多成熟的有关身份认证的解决方案[4]。相对于系留模式，非系留模式下的安全问题就比较复杂，至今还没有人提出理论上的解决方案。虽然现在不少 DRM 公司的产品都宣称已经支持非系留模式，而且各个公司都有其独特的技术方案，然而这些公司在谈到他们的技术方案时都比较隐讳，那么这些公司的技术究竟怎样保证 DRM 技术的安全呢?下面先介绍一下非系留模式下的 DRM 系统的安全模型。

2.2 非系留模式下的 DRM 系统的安全模型

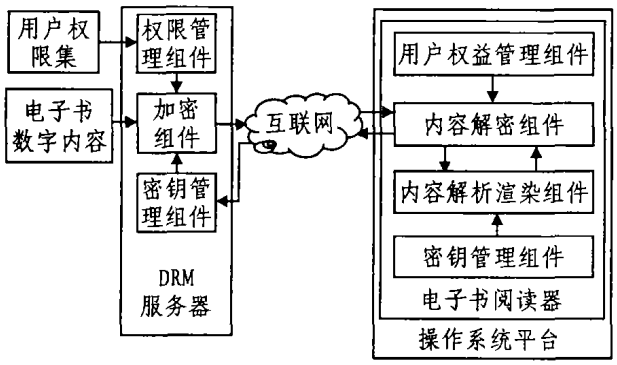

图2 基于安全组件的 eBook 管理

DRM 的安全模型与基于互联网(Internt)的系统的安全模型有着本质的区别。基于互联网的系统的安全性在于构成它的部分组件是在一个安全的环境下运行的。举例来说，构成电子商务系统的重要组件，如 Web 服务器和相关服务器程序，是在安全的环境下运行的。在这种安全的环境下，只用受信任的人才能接触到重要组件的程序。这样，怀有不良企图的客户，由于接触不到重要组件的代码，从而无法有效地攻击整个系统。

对于 DRM 系统，情形却恰恰相反。数字内容的提供商要将数字内容传送到一个不能保证安全的用户环境。为了能在这种不可信任的环境下保证数字内容的安全，DRM 系统只能依靠在用户环境下运行的组件来实现。该组件将确保数字内容的拥有者的权益得到保障，并使得用户所赋予的权利得以充分实现。在我们的管理电子书的 DRM 系统中，在客户端运行的组件是电子书阅读器。这里电子书阅读器不仅仅用来显示电子书，而且还负责保障电子书的安全和用户权益的实现。

3 基于 DRM 的 eBook 管理框架

我们的基于 eBook 的 DRM 系统在结构上采用的是 C/S 模式，如图2所示。服务器端是 DRM 服务器，主要负责权益的创建管理和 eBook 内容的加密。客户端是 eBook 阅读器，其完成 eBook 内容的解密，相关权限的验证和内容的解析展示。

用户的权益在 DRM 服务器端被加密包装，以授权证书的形式下载到客户端。授权证书的内容由用户的权益描述和相关电子书的解密密钥组成，并被加密封装。在客户端，eBook 阅读器对证书解密，验证其中的用户权限是否合法，如果合法，则对电子书进行解密，用户才能阅读。

3.1 eBook 阅读器的实现

目前，国内和国外开发的电子书阅读器种类比较多，国内比较有影响的有超星阅读器和北大方正的 Apabi[5]，国外的有 Adobe 公司的 eBook Reader[6]，Microsoft 公司的 Microsoft Reader[7]。我们将从以下三个方面对各个公司的电子书阅读器进行比较。

·电子书格式 超星的电子书格式采用的是一种图像编码格式，占用的磁盘存储空间比较多。其他三家阅读器也都有各自专门的电子书格式，Apabi 的是 CEB 格式，Adobe 的是 PDF 格式，Microsoft 的是 LIT 格式，他们共同点都是采用数字编码来保存原始文字信息，因此能够取得比较大的压缩比，占用的磁盘存储空间也比较小。由于这几家公司的格式标准互不相同，因此不能有效地实现电子书之间的共享。

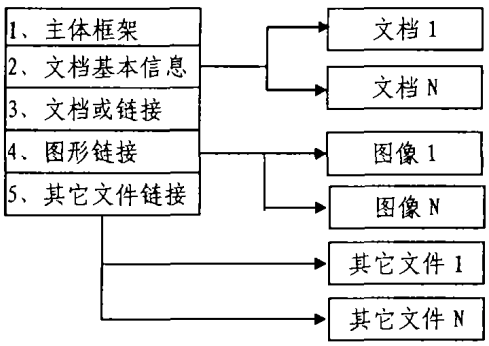

图3 OEB 文档结构格式

·电子书的显示 由于超星的电子书的格式采用的是图像格式，在显示上能较好地反映原始印刷品的原貌。Adobe 公司的显示技术是 Cool Type 技术，Microsoft 公司的技术是 Clear type 技术，这两种技术都能很好地使得在显示器上的阅读达到像阅读实体书的感觉。Apabi 采用的也是类似的技术。

·权益的认证保护 超星的电子认证方式通过读书卡的方式来实现，这种方式存在如下问题，攻击者可能通过对超星阅读器的代码进行篡改，从而绕过安全认证部分的检查而直接到服务器端免费下载电子书。Apabi，Adobe，Microsoft 公司的安全认证都采用电子认证证书的方式来保护电子书的安全。通过电子认证证书把电子书和用户身份信息、用户权益相绑定，安全性要高得多。

下面将从上面所说的三个方面介绍一下我们开发的电子书阅读器。

在电子书阅读器的格式上，我们采用了国际上通用标准 OEB 规范1.1(Open eBook formula Publication Structure)作为电子书的内容格式标准。由于 OEB 的技术是开放的，对电子书的兼容性和以后的发展都有极大的好处。

由于 OEB 格式包含的文件是以链接的形式存在，比较分散，如图3所示。我们把这些文件进行编码合并到一起，形成一

个统一的文件，这样不仅便于进行存储压缩，而且更有利于实现版权保护。

在显示技术上我们采用了子像素(Subpixel)技术。Subpixel 技术是类似 Mircsoft 公司的 ClearType 的一种技术。它通过将单个像素能量扩展到邻近像素的方法，来弥补显示设备能力不足而对图像显示造成的粗糙，从而给读者在阅读上提供一种舒适的感觉。在功能上，我们的电子书阅读器可以选择拷贝文字，对段落进行标注，放大和缩小字体。对于电子书的安全认证，将在下面介绍。

3.2 基于硬件指纹识别的电子书认证算法

我们的 DRM 系统提供给每个用户的电子书都有相应的电子证书。电子证书由数字签名和用户权益信息组成。数字签名用来验证电子证书的合法性。电子书的内容和用户权益信息都被加密保存。

(1)定义

1)p 和 q 是 RSA[6]的两个大素数，g 是 p 或 q 的本原元，H 是一个单向函数；

2)计算 $n=p*q$，选择随机选择加密密钥 e，利用 Euclid 算法计算解密密钥 d，满足 $e*d=l(\mathrm{mod}(p-1)*(q-1))$[4]；

3)hkey 是客户机的机器硬件指纹，它由客户机的 CPU 序列号、主板信息、硬盘序列号、网卡序列号等硬件信息组成，是一个长度约为300多位的大整数；

4)m 是用户的身份信息，b 是电子书的序列号，m 和 b 都是一个32位整数；

5)g，n 和 d 是公开的，p 和 q 丢弃，e 保密。

(2)生成电子证书数字签名的基本算法

1)第一步：电子书阅读器将 hkey、m、b 提交到 DRM 服务器，DRM 服务器进行如下运算

$$ID=H(hkey,m,b) \tag{1}$$

ID 就是电子证书的标识号。这个标识号是用户身份和机器硬件信息、电子书序列号的绑定，且是唯一的。

2)第二步：计算 $S=ID^e \bmod n$，S 是保密的，因为它和加密密钥 e 相关。选择一个随机数 R，这里我们确保 R 是随机的，并且相对每一个电子证书来说，这个 R 都是不同的。计算：

$$X=g^{d*R} \bmod n \tag{2}$$

$$Y=S*g^R \bmod n \tag{3}$$

则 X，Y 就是电子证书的签名。

(3)验证签名　在客户机端，电子书阅读器将本机的 hkey，以及用户信息 m，电子书序列号 b 做如下运算：ID'=H(hkey,m,b)，由于

$$Y^d/X \bmod n=(S*g^R)^d/g^{d*R} \bmod n=S^d \bmod n$$
$$=ID^{e*d} \bmod n=ID \tag{4}$$

因此如果 ID'=ID 这证明了该电子书签名确实是与由本机硬件信息产生的。

(4)电子书的内容和用户权益信息的解密　电子书阅读器在验证了电子证书签名的合法性后，计算

$$K=ID^d*Y/ID \bmod n=ID^d*S*g^R/ID \bmod n$$
$$=ID^d*ID^e*g^R/ID \bmod n=g^R \bmod n \tag{5}$$

K 是与 R 有关的，由于对于每一个电子证书来说，R 都是不同的，因此我们可以把 K 看作是电子证书特有的标识号。它将作为计算解密密钥的入口参数，这样由每个电子证书计算出来的解密密钥都是唯一的。具体的解密过程是在证书管理组件中进行的。在该组件中，我们采用了的是一种混合编码(Scrambling)技术，它是很多已知加密算法及其变形的集合。该技术的目的不是用来提供额外的加密强度，而是防止攻击者比较容易发现解密密钥。攻击者要想得到真正的解秘密钥，就不得不对众多的加密算法进行分析，如果攻击者知难而退，也就达到我们的设计目的了。

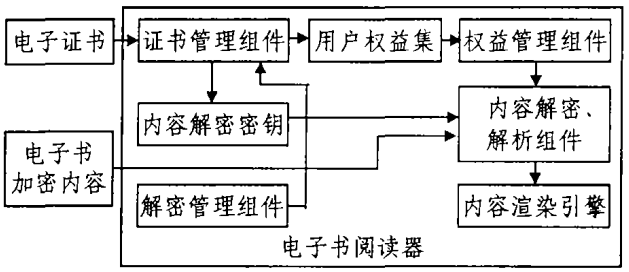

图4　电子书阅读器内部结构图

(5)安全性分析　由 X，Y 计算 R 和 S 要面临着计算 g 的离散对数问题，这是个 NP 问题，是非常困难的。而由 ID 计算加密密钥 e 是 RSA 的安全问题，只要素数 p 和 q 选得足够大，就能保证足够的安全。在实际应用中，我们的密钥的位数取在1024位以上，R 的位数在512位左右，就能够保证足够的安全。由于电子证书的验证和解密密钥的获得都与机器的硬件指纹相关，当电子证书拷贝到其他机器上，就会验证失败并且计算出来的解密密钥是错误的。这就有效地防止了电子书的非法复制和二次传播。

电子书的内部功能框图如图4所示。密钥管理组件用来生成本机的机器理件指纹。证书管理组件完成证书验证、计算解密密钥和解密用户权益集。用户权益管理组件用来验证用户权益的合法性并确保其得到贯彻。内容解密、解析组件用来解密电子书内容并进行语法解析。内容渲染引擎用来将解析后的电子书内容显示到输出设备上。从软件安全的角度，我们还从阅读器软件整体上采用反跟踪、代码反篡改技术，进一步提高了电子收的安全。

结束语　在数字图书馆的构建中 DRM 系统的安全技术是一个十分重要的研究课题，许多研究机构和公司都在加大对该领域的研究和开发。在电子书的安全管理方面，我们认为一个有效的 DRM 系统应该充分进行安全性和实用性的综合考虑，不要企图实现一种绝对安全的 DRM 系统，而是建立一种使破解者需付高昂代价破解的技术方案。本文提出的电子书管理框架和基于机器硬件指纹识别的 DRM 技术就是折中考虑上述的问题，试验表明，该原型系统具有较好的安全性和较高的实用性。未来的主要研究方向是，进一步完善 DRM 服务器中的各种功能，并将其推广到数字图书馆中的其他数字媒体的安全管理中。

参 考 文 献

1 Davis D M. Impact of digital rights management on access and distribution of intellectual property in libraries and information centers. Technicalities, 2000, 20(4): 5～7

2 Feigenbaum J, Freedman M J, Sander T, et al. Privacy Engineering for Digital Rights Management Systems. Berlin: Springer-Verlag, 2001. 76～105

3 Kenny S, Korba L. Applying digital rights management systems to privacy rights management. Computers and Security, 2002, 21(7): 648～664

4 Kwok S H, Cheung S C, Wong K C, et al. Integration of digital rights management into the Internet Open Trading Protocol. Decision Support Systems, 2003, 34(4): 413～425

5 方正 Apabi 重要文档安全保护解决方案. http://www.apabi.com/project/pro-1/pro-c1.htm

6 Adobe eBook Reader. http://www.adobe.com/products/ebookreader/main.html

7 Microsoft Reader. http://www.microsoft.com/reader/default.asp

8 Schneier B. Applied Cryptography: Protocols, Algorithms and Source Code in C. John Wiley & Sons, New York, 1994