EPUB 电子书版权保护技术研究和实现

张伟，宋美娜，朱先忠

（北京邮电大学计算机学院 PCN&CAD 中心，北京 100876）

摘要：随着移动互联网的蓬勃发展，电子书已成为信息时代图书发展趋势，EPUB 作为一个开放的电子书标准，正逐步成为终端电子书阅读的主流格式。本文将通过研究 EPUB 文档格式和相关的信息加密算法，利用对称加密 DES 和非对称加密 RSA 相结合，提出一种新的 EPUB 格式电子书内容版权保护机制(EEA)，并进行了实验和算法分析。

关键词：版权保护；EPUB；内容加密

中图分类号：TP309.7

Investigation and implementation of copyright protection for EPUB eBook

Zhang Wei, Song Meina, Zhu Xianzhong

(PCN&CAD Center, Beijing University of Posts and Telecommunications, Beijing, 100876)

Abstract: With the rapid development of mobile Internet, e-books have become the trend.As for EPUB is an open e-book standard, it is gradually becoming a mainstream e-book format for mobile terminal. This paper will firstly investigate EPUB document format and some related encryption algorithm. Using DES and RSA, this paper will then propose a new EPUB format e-book and content copyright protection mechanism(EEA) based on this new format. At the end, experiments and analysis of the proposed algorithm are given.

Key words: Copyright Protection;EPUB; Encryption

0 引言

自进入数字时代以来，人们的生活被无处不在的数字产品包围着，EPUB 电子书正逐步成为终端主流电子阅览格式，网络时代所特有的知识产权保护问题也日益突出。EPUB 是一种非加密、非固定版式的、基于 XML 的电子图书文档，“非加密”（unencrypted）指未对文档进行数字加密，任何人均可以打开（unzip）和编辑 EPUB 格式文档。由于未加密的 EPUB 资源可以进行未经授权的无线自由传播，EPUB 电子资源的知识版权保护成为一个值得研究和探讨的课题。

对于数字版权管理的定义，国际数据中心（Internet Data Center，IDC）的描述是：运用一系列的软硬件技术限制已授权用户对数字内容的使用并保护数字内容在其生命周期内获得合法利用。

本文对各类数字资源版权保护技术和加密技术做了分析比较，并在深入研究 EPUB 内部结构的基础上，提出了 EEA（EPUB Encryption Algorithm）算法：在创新性修改 EPUB 内部结构的同时，利用对称加密和非对称加密相结合的方式，最终实现了 EPUB 电子书的版权保护功能。

基金项目：国家图书馆横向委托项目--数字家庭互动媒体服务系统；国家科技支撑计划课题（2012BAH01F02）

作者简介：张伟（1990-），男，硕士，现代服务业及云计算

通信联系人：宋美娜，女，北京邮电大学计算机专业教授，未来通信、电信网管、现代服务业. E-mail: mnsong@bupt.edu.cn

1 EPUB 简介及内部结构

1.1 EPUB 简介

EPUB 是一种完全开放和免费的电子书标准，属于一种可以“自动重新编排”的内容：文字内容可以根据阅读设备的特性，以最适于阅读的方式显示。它的组成是：电子书内容都是 XHTML 的文件，描述都是 XML，并且是一个包含上述文件的 ZIP 文件。即 EPUB 元数据是 XML，EPUB 内容是 XHTML。由于其文字内容可以根据阅读设备的特性（屏幕尺寸、不同平台），以最适于阅读的方式显示，EPUB 也逐渐成为各终端类型（phone，pad 等）的主流电子书阅读格式。

EPUB 资源格式包括三项主要规格[1]：

OPS（Open Publication Structure）2.0：定义内容的版面；

OPF（Open Packaging Format）2.0：定义以 XML 为基础的.epub 档案结构；

OCF（OEBPS Container Format）1.0：将所有相关文件收集至 ZIP 压缩档案之中。

1.2 EPUB 电子书结构

EPUB 作为一个压缩格式，其内部结构如图 1 所示：

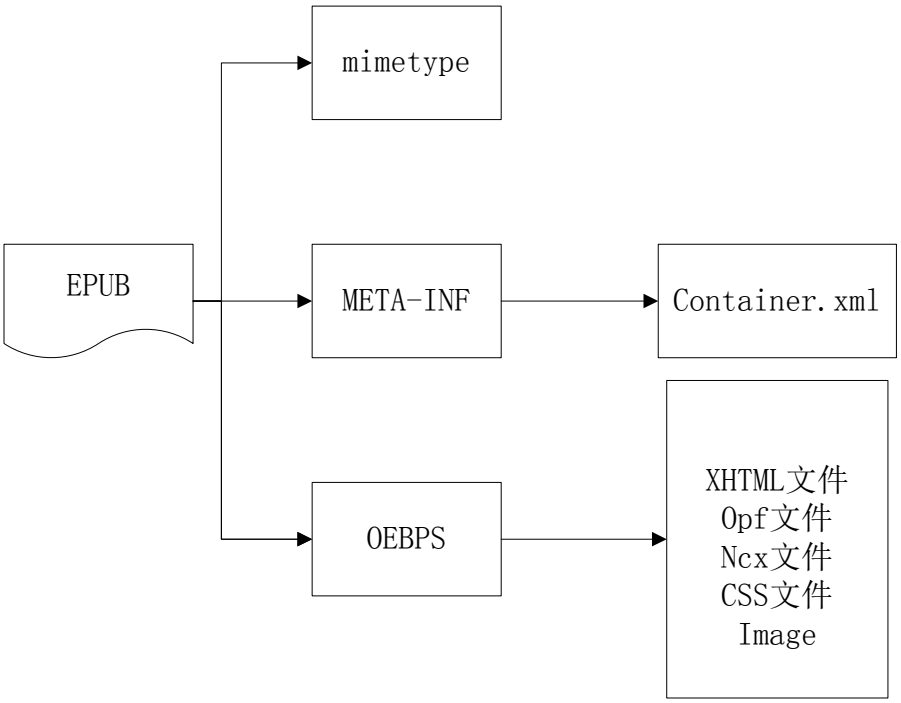

图 1 EPUB 电子书结构

Fig. 1 EPUB structure

1.2.1 mimetype

这个文件非常简单，必须命名为 mimetype，用来说明文件格式，文件内容如下：

application/epub+zip

1.2.2 META-INF

META-INF 是一个目录文件：其中主要包含一个 container.xml 文件，EPUB 阅读系统首先解析和查看该文件，该文件包含电子书的元数据文件位置和打开方式，因此本文件也是在解析 EPUB 电子书时首先要查找的文件。

1.2.3 OEBPS

OEBPS 同样是一个目录文件：它用于存放电子书内容 XHTML、OPF 文档、NCX 文档、相关图片及 CSS 等信息。元数据文件 OPF，文件名没有特殊要求，它指定了电子书中所有内容的位置。NCX 为电子书的逻辑目录， 定义了电子书的目录表。

2 EPUB 版权保护机制

数字版权保护技术是指在网络及数字化环境下，以一定的方法，如加密技术、数字水印权利描述、身份认证等，使数字内容的权利主体获得对其客体的控制权，从而防止非授权使用，保护权利所有人利益的一种综合性技术体制。

随着移动终端上面使用 EPUB 格式的电子书进行阅读的人越来越多，关于 EPUB 格式电子书的版权保护变得愈来愈急迫。由于 EPUB 本身内容是利用 XHTML 文本进行明文存储的，所以用户可以随意进行文本内容的解析，达不到版权保护的需求。本文将提出一种对称加密算法和非对称加密算法相结合的 EPUB 加密算法：EEA（EPUB Encryption Algorithm）算法。

2.1 DRM 及加密技术

在讲述本文的 EEA 算法之前，首先研究了解相关的数字版权保护技术：DRM 技术和信息隐藏算法[2]。

2.1.1 DRM 相关技术

① 数据加密技术：以数据加密和防拷贝为核心的 DRM 技术，是针对数字内容进行加密，只有授权用户才能解密使用。数据加密技术分为两类，即对称加密和非对称加密。对称加密：信息加密和解密使用相同的密钥，由于其速度快，对称性加密通常在消息发送方需要加密大量数据时使用，常用算法有 DES AES 等。非对称加密：算法需要公开密钥（publickey）和私有密钥（privatekey）两个密钥对应使用，目前主流的算法有 RSA ECC 等。

② 数字水印技术：数字水印技术的基本思想是在数字内容中嵌入数字水印， 它隐藏在数字内容的原始数据中，这种水印可以是注释，标识，序列号，检索信息等。数字水印方法，只能在发现盗版后用于取证或追踪，不能在事前防止盗版 [3]。

③ 身份认证：网络环境下用户的身份认证实质就是检验用户是否具有其所出示的对某种资源的使用和访问权利。一个完整的 DRM 系统的身份认证应该做到交易双方的准确辨认，不仅系统要认证用户，用户还应该认证系统[4]。

2.1.2 信息隐藏算法

通过研究 EPUB 文档格式和基于 XML 的信息隐藏方法，从而研究出的一种 EPUB 电子文档信息隐藏技术[5]。该算法通过修改 XHTML 元素置标尺寸进行同义词替换，使用不同的段落格式来表示不同的信息，从而实现了信息隐藏。

信息隐藏算法大致可分为三个步骤：

隐藏信息加密和变换

根据给定的密钥，将待隐藏的信息使用加密算法进行处理，然后将加密后的信息进行编码，使得加密后的信息转化为二进制信息。

解析文件

在这一步中，将 EPUB 文件使用 unzip 解压，将里面的文件分为两类：1.XML 类型的

文件和类似 XML 格式的文件；2.不属于上面两种的文件。将符合条件的文件编入文件集 FILE_SET，形成一个已排序的文件串 FILE_STRING。

隐藏算法

针对文件串 FILE_STRING 的文件，采用顺序的操作，根据二进制的隐藏信息，使用前面提到的利用置标尺寸信息隐藏算法对文件进行处理。

2.2 EEA 版权保护

EEA（EPUB Encryption Algorithm）加密：即对 EPUB 进行内容加密实现版权保护，主要包括数字内容加密和内容解密两部分。

2.2.1 电子书加密过程

众所周知，对称密钥加密算法非常快，适合于加密大量数据，但安全性较低。从安全性角度看，非对称加密算法更有优越性，其加密算法几乎不可能被破解，但加密大量数据时效率较低。

在本文的 EEA 加密过程中，将结合非对称加密算法和对称密钥算法的优点：首先利用对称加密算法 DES 密钥（DES key）加密数据内容本身，然后用非对称加密算法 RSA 中的密钥（Private key）加密对称算法的密钥（DES key），生成 Result 存入 EPUB license 文件中，其中 license 文件是在 EPUB 结构基础上添加的新文件。EEA 加密过程如图 2 所示。

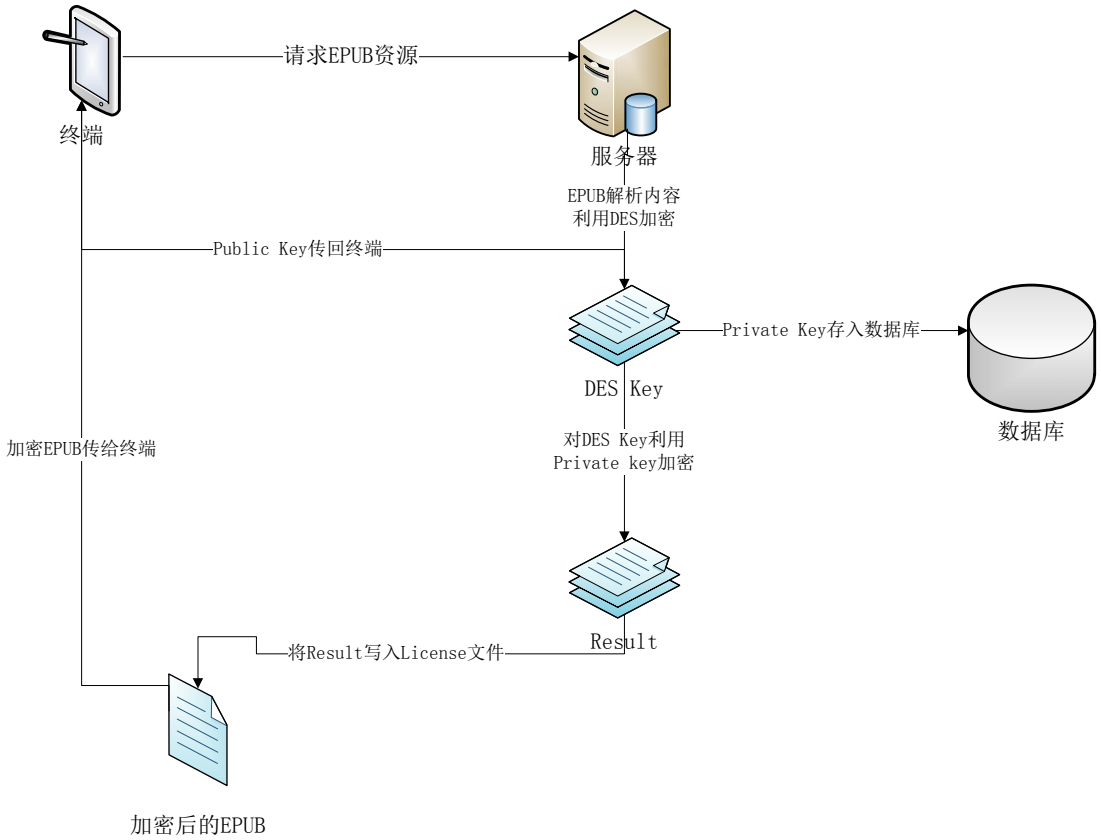

图 2 EEA 加密过程

Fig.2 EEA Encryption Process

EEA 加密过程大致分为四个步骤：

加密资源内容

用户请求相关的 EPUB 文件，服务器端解析 EPUB 文件，读取 EPUB 内容，利用 DES 对称加密算法对内容加密，加密后的内容重新写入对应的 XHTML 内容文件中，加密密钥 DES Key 作为第二步的输入。

加密对称秘钥

每个终端用户在请求 EPUB 数字资源时，都会由服务器产生一对对应的私钥（Private Key）和公钥（Public Key），公钥传给用户终端进行保存用于电子书的解密，私钥存于服务器端数据库用于电子书的加密工作。针对第一步产生的 DES Key 利用存放的 Private Key 进行非对称加密操作，产生 Result。

生成新的 EPUB

首先改进 EPUB 内部结构，增加一个 License 文件，然后将第二步产生的 Result 存入 License 文件中。改进后的 EPUB 结构如图 3 所示。

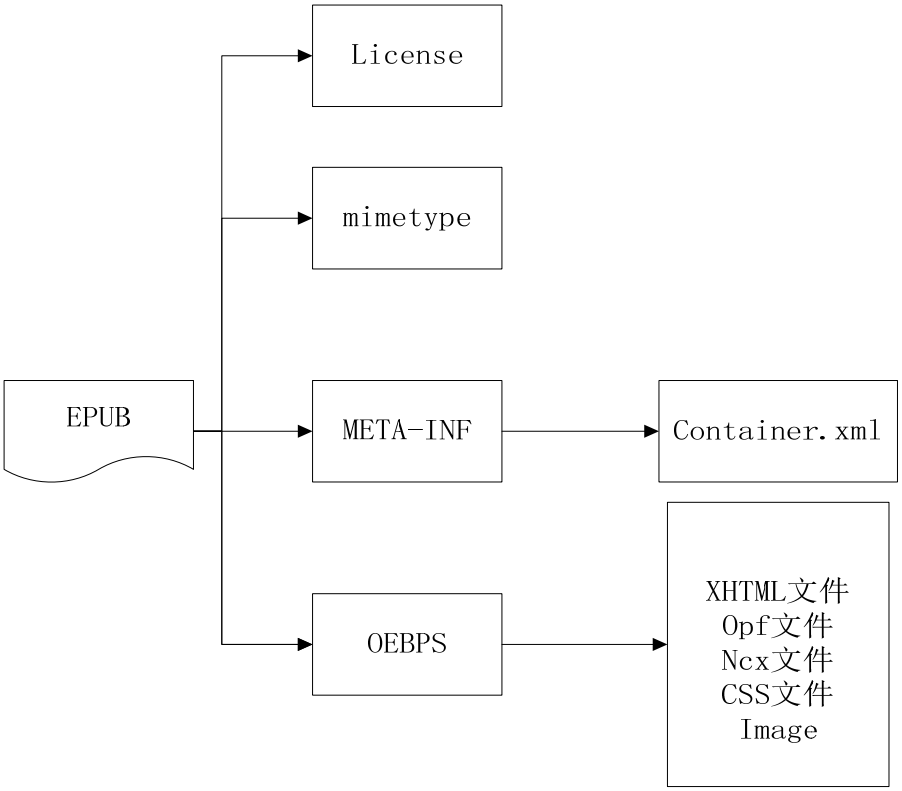

图 3 加入 License 文件的 EPUB 结构

Fig.3 new EPUB structure with added licence

EPUB 资源发布

将新生成的加密后的 EPUB 电子书发布，发送给终端用户，此时资源已具备版权保护功能，因为资源内容信息是加密后的暗文，必须利用对应的 Public Key 解析 EPUB 中的 License 文件内容才能进行查看和阅读。

2.2.2 电子书解密过程

用户终端根据自己的公钥对获得的加密 EPUB 电子书进行解密，解密过程如图 4 所示。

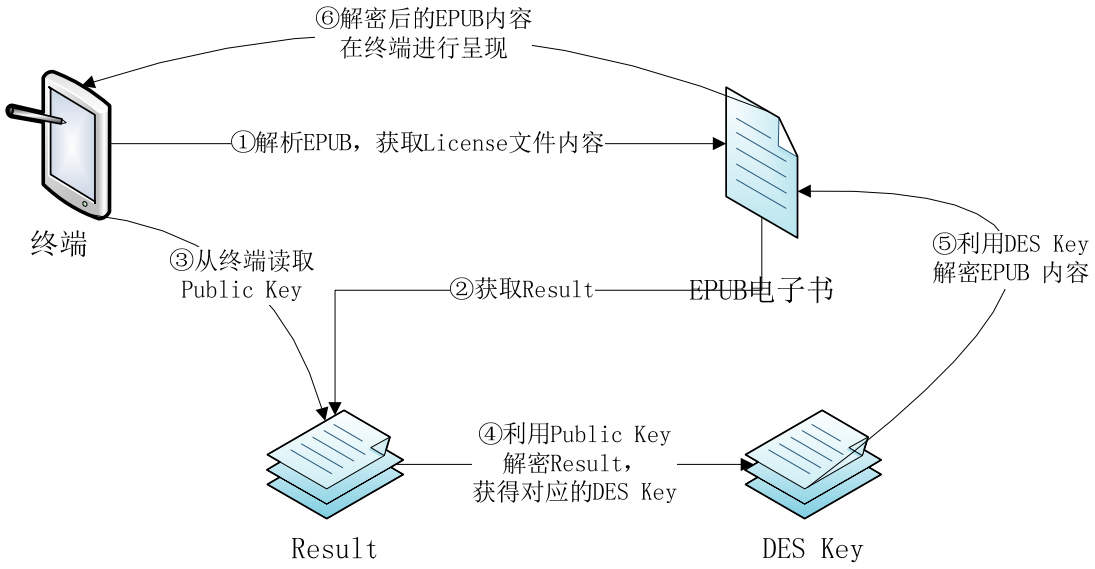

图 4 EEA 解密过程

Fig.4 EEA Decryption Process

EEA 加密过程大致分为以下三个步骤：

解析 license 文件

根据相应的程序代码解析 License 文件，获取 License 内容 Result。

获取对称秘钥

利用第一步产生的 Result 以及终端中保存的相应的 public Key 进行解密操作，解密后的结果就是加密内容的 DES Key。

获取资源内容

根据 DES Key 对 EPUB 内容进行对称解密操作，获取到的结果就是 EPUB 原来的电子书内容，解密完毕，EPUB 内容最终可以在终端进行呈现。

3 EEA 算法实现和分析

3.1 算法实现

EPUB 电子书加密过程大致如图 5 所示：

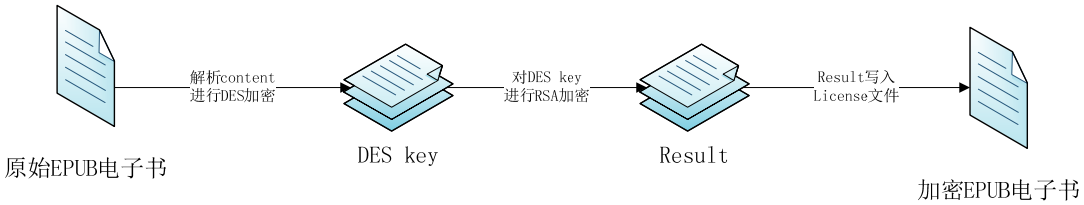

图 5 加密过程

Fig.5 Encryption Process

为了验证 EEA 算法在 EPUB 资源版权保护方面的功能，本文基于 DES 算法和 RSA 算法实现一个简易 EPUB 加密系统。

第一步，EPUB 内容解析，根据 container.xml 定位 opf 文件位置，然后读取 opf 内容，最终找到对应的内容文件存放地址，读取 EPUB 资源内容 content。

第二步，对 content 采用 DES 对称加密，加密完成后将加密后的 Econtent 重新写入对应

位置，DES 加密代码实现如下：

```
    private static String contentEncrypt(String content) throws Exception {
        Cipher cipher =Cipher.getInstance("DES");
        //得到加密的钥匙
        SecretKey key =KeyGenerator.getInstance("AES").generateKey();
        //初始化加密操作,传递加密的钥匙
        cipher.init(Cipher.ENCRYPT_MODE,key);
        //将加密的内容传递进去，返回加密后的二进制数据
        byte [] results =cipher.doFinal(content.getBytes());
        //将加密后的二进制数据写入到对应的内容文件中
        FileOutputStream fosData=new FileOutputStream("content.xhtml");
        fosData.write(results);
        fosData.close();
        return key.toString();//返回 DES key
    }
```

第三步，对 DES key 进行 RSA 非对称加密，加密完成后将此 Result 写入 license 文件中，RSA 加密如下：

```
    private static void RSAEnrypt(String content)throws Exception {
        Cipher cipher =Cipher.getInstance("RSA");
        //实例化 Key
        KeyPairGenerator keyPairGenerator=KeyPairGenerator.getInstance("RSA");
        //获取一对钥匙
        KeyPair keyPair=keyPairGenerator.generateKeyPair();
        //获得公钥，发送给终端用户
        Key publicKey=keyPair.getPublic();
        //获得私钥，并存放于服务器端
        Key privateKey=keyPair.getPrivate();
        saveKey(privateKey,"zxx_private.key");
        //用私钥加密
        cipher.init(Cipher.ENCRYPT_MODE,privateKey);
        //获取需要加密的 DES key
        String DESKEY=contentEncrypt(content);
        byte [] result=cipher.doFinal(DESKEY.getBytes("UTF-8"));
        //加密后的数据写入到 license 文件
        saveData(result,"license");
    }
```

最终，EPUB 内容加密完毕。

3.2 算法分析

本算法利用了 DES 对称加密和 RSA 非对称加密算法相结合的方式完成了 EPUB 内容的加密工作，并对 EPUB 内部结构做出了创新性的改进，添加了 License 文件作为版权保护工作的核心文件。

如果用户没有 EPUB 文件对应的 Public Key 即为资源非法拥有用户，或者 License 文件遭到了人为的修改和破坏，那么终端用户将无法对 EPUB 电子书进行解析和播放，实现了 EPUB 的版权保护功能，防止了对 EPUB 电子书的无限非授权传播。

4 结论

本文提出了一种基于 EPUB 电子资源的版权保护机制，并进行了具体的实验分析，为当前网络上电子资源的保护提供了一种可靠易行的解决方案。在综合分析各种 DRM 相关技术及信息隐藏算法的基础上[6]，EEA 算法利用了对称加密算法和非对称加密算法相结合的加密方式，并对 EPUB 内部格式进行了创新性修改。

综上所述，EEA 算法是维护 EPUB 电子资源知识产权的有效措施，并将为数字内容版权的保护提供新的角度。

[参考文献] (References)

[1] IDPF, EPUB 3 Specification[S]. Seattle, WA: International Digital Publishing Forum
[2] 蔡林峻.电子书 DRM 应用现况[J].电子与电脑,2010(4):27-29.
[3] 冯雪,朱新山,汤帜.多媒体数字水印技术研究进展[J].计算机工程与应用,2007(13):1-6.
[4] 吕格莉,邵自然.网络环境下身份认证技术探析[J].现代计算机,2006(11):53-56.
[5] 吴桐,郭燕慧.一种 EPUB 文档信息隐藏技术的研究[OL].中国科技论文在线,2011(11):4-6.
[6] 王法涛.数字权限管理技术（DRM）应用研究[D].长春:吉林大学.2006.