

基于 Jacob 的文档安全管理的实现

梁志罡

(华南理工大学 计算机科学与工程学院, 广东 广州 510006)

摘 要:网络的安全性问题是一个比较热点的问题。出于对教师的知识产权的保护,要求学生用户端不具备复制与下载 ppt 和 word 等文档的权限,而教师用户端则需要具备在线实时更改文档内容的权限。通过分析当前市场上存在的文档系统的基本特点和安全性要求,并结合自己的创新,采用 Jacob 以及 Webdav 等关键技术实现了一个新型的在线文档管理系统。系统实验结果表明本系统能够很好地实现文档的安全性保护以及文档在线编辑功能,同时本系统在性能上也能够应对比较大的并发访问。

关键词:文档安全;在线文档管理系统;Jacob;Webdav

中图分类号:TP309.5

文献标识码:A

文章编号:1673-629X(2011)02-0185-04

Implementation of Document Security Management Based on Jacob

LIANG Zhi-gang

(College of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China)

Abstract: The security of network is a hot problem now. Introduce a online-document management system, its goal is to protect the teacher's intellectual property, it requires that the student role does not have permission to download or copy the document which are uploaded by teacher, and the teacher role could edit the documents with a browser on-line. Through analyzing the base characteristic and security demand of the document management system in the market, combine with my innovation and then construct a new-style document management system with Jacob and Webdav techniques. The implementation of this system indicates that it can do well in the function of document security and on-line editing, in addition, the system also can deal with a fairly request simultaneously.

Key words: document security; online-document management system; Jacob; Webdav

0 引言

作为如今的大学,单单是课堂上的授课学习已经满足不了新时代学生的需求,建立基于网络的学习系统正是解决这一问题的最佳途径。但是问题出现了,可能会有一些同学把老师辛苦做出来的文档下载下来从而把这些文档发布到其他学校去,这样其他学校的老师就可以不废任何力气就做出自己的教学作品。因此,在这个信息化的时代,网络信息安全是非常值得关注的。文中将通过目前最为流行的 J2EE 框架 SSH 来实现一个 BS 架构的在线文档管理系统,首次以 Jacob 与 JSP 技术相结合的形式来实现文档的防复制与下载功能,很好地保证了文档的安全性。需要说明的是本文档管理模块不单只是简单的文档的上传与下载,还包括文档的在线修改、全文检索、文档备份等,使得文档的管理工作能够有很高的效率。

1 研究现状

就目前笔者查到的信息而言,J2EE 都受到了很高的青睐,大多数的企业都是用 J2EE 相关的框架来开发自己的产品。但是,就前面提到的文档的安全性这个问题来看,还没有发现哪个网站是搞得很好的。对于如今国外对 office 文档保护(防文档的复制下载)方面用得最多的就是 Microsoft Office 2003 或 2007 中的信息权限管理(Information Rights Management,IRM)功能,但是这项功能比较麻烦,需要知道所有要查看某份文档的用户的 .Net Passport 账户以赋予多个人打开文档的权限,这不是很符合实际。因此,笔者经过查阅相关的资料发现,如今对文档的安全性保护可以用以下的方式去实现:

1)通过脚本语言和标记语言封锁网页复制、网页另存为等功能;

2)可以考虑通过控件把文档保存在缓存,而不是硬盘上;

3)把文件进行流式的传送,类似与流媒体的形式,这也符合局域网速度快的特点;

4)通过相关的工具把文档进行格式转换之后再

收稿日期:2010-06-18;修回日期:2010-09-25

作者简介:梁志罡(1985-),男,广东茂名,人,硕士研究生,研究方向为计算机网络安全与应用。

上传。

2 关键技术简介

2.1 Jacob

Jacob, 它的官方解释为 Java COM Bridge, 即 Java 和 com 组件间的桥梁。com 一般表现为 dll 或 exe 等二进制文件^[1]。大家都知道, Java 不是微软的产品, 肯定没有 .NET, C++ 等对 com 的访问操作那么简单, com 可以直接操作 Windows 平台上各种软件, 如果 Java 想访问 com 组件的话, 就要通过 JNI 的技术去访问 com 组件, 但是如果单纯用这个技术的话实现起来是比较麻烦的, 也比较难看得懂。但是如果是 Jacob 的话就明显容易许多, 因为 Jacob 的实现方式就是使用 jacob.dll 动态链接库通过 JNI 的方式实现 Sun Java 平台上的程序对 com 的调用。因此, 只要知道 Jacob 的 API (封装在 jacob.jar 中), 就可以用它去进行一系列的 com 组件访问了^[2], 下面介绍的这个系统中用到的是通过 Jacob 去调用 Office 的宏, 并实现各种格式的转换^[3]。

2.2 WebDAV 协议

WebDAV (Web-based Distributed Authoring and Versioning) 是基于 HTTP 1.1 的一个通信协议^[4]。基于 WebDAV 可以实现一个功能强大的内容管理系统或者配置管理系统。通过 WebDAV 协议实现了用户直接编辑服务器上的文档, 并实现保存功能, 这一点是很重要的, 但对于不通的语言, 如 JSP 与 ASP.NET, 对 WebDAV 协议的使用又是不同的, ASP.NET 在对 WebDAV 的使用上是很方便的, 毕竟是微软的产品。而对于 JSP 而言就需要费比较大的力气去查相关的资料了, 毕竟网上基本上没有多少人愿意共享这些相关的技术。经过一段时间的研究发现通过服务器 Tomcat 的协助可以实现这一个功能, 这个问题在后面将会详细介绍。

3 设计方案

经过对以上所有知识的了解, 个人认为用 J2EE 来开发这个系统是很合适的: Java 是开源的, 并且它的开发工具大多是免费的, 开发出来的产品能够在不同的平台上使用, Java 产品的一次编译, 到处运行这句话深入人心。而 J2EE 开发的系统虽然说比较复杂, 特别是用到了框架, 但如果比较熟练后你会发现用 J2EE 的各种框架来开发会使得整个系统的框架结构非常的清晰, 各自功能分配很具体, 各个模块层之间的耦合度很少, 非常适合大型系统的开发, 并且在维护上非常方便^[5]。以上就是笔者选择 J2EE 的 SSH 框架来开发这个系统的主要原因。对于本系统的文档处理部分, 文档安全性保护方面笔者首先想到的是加密方法^[6], 但

话又说回来, 如果文档加密后学生端又怎样打开文档浏览呢? 不可取! 之后还想到了用流媒体技术去实现, 但是基于网站是采用 BS 架构, 最后决定采用文档转换的方法。通过查找大量的资料可以发现在 Java 中用 Jacob 技术可以帮助实现 Office 的文档转换, 这里笔者想到的大致方法是把 ppt 和 doc 文档转为相应的 html 文档, 之后再通过相应的处理实现这些 html 文档的防复制功能。如果可以实现 ppt 和 doc 转 html, 就肯定能够加快用户的访问, 因为访问 html 页面的速度肯定快过在客户端打开 Office 来浏览文档的速度。

4 文档处理模块的研制

4.1 SSH 开发框架的搭建

使用框架的好处是方便日后的维护, 虽然开发的时候有一定的麻烦, 但是整个系统的架构 (四层架构) 可以非常清晰地表现出来。四层体系架构是对三层 MVC 架构的扩展, 即分为: 表示层、控制层、服务层和数据持久层。开发的时候开发顺序就是先做好 JSP 页面表示层, 然后对之前设计好的数据库进行持久化对象处理 (Hibernate), 有了持久化对象之后, 就可以编写 DAO (数据访问对象) 来对数据进行插入、更新或删除了。最后就是写 Action (控制层), JSP 页面与 DAO 对象的连接, 最终体现为 Struts 负责控制层, Hibernate 负责持久层, 而 Spring 的作用贯穿了整个中间层, 将 Web 层、Service 层、DAO 层及 PO 无缝整合, 其数据服务层用来存放数据。这里想补充说明的是 Spring 的使用可以大大降低系统组件之间的耦合度问题。这样设计出来的一个系统, 虽然比较麻烦与复杂, 但是提高了该系统的可维护性和可扩展性^[7]。通过 SSH 框架的使用, 可以有效地整合整个系统, 使得系统能够更高效地运行。

4.2 文档模块的实现

本模块的实现包括多个用于显示信息的 JSP 页面, 如上传页面、文件列表页面、文件检索页面等。同时还有各个页面相应的控制层 java 文件, 形如 XXXAction.java, 接着就是数据对象访问层的 java 文件, 形如 XXXDAO.java, 然后就是数据对象持久层的 java 文件, 如 XXX.java, 表示数据库 XXX 表的对象。同时还有一些工具类, 如 Amendfile.java, WordtoHtml.java, PPTtoHtml.java, WordtoPDF.java 等。其中, 页面表示层与控制层是通过 struts-config.xml 联系起来的, 控制层与数据访问层以及数据持久层是通过 applicationContext.xml 文件联系起来的, 这就是 Spring 的依赖注入的管理形式的体现, 很好地降低了各层之间的耦合度。

本系统的文档处理模块与常见的简单的文档的上

传功能不尽相同,即在文档上传的过程中需要对文档进行一些处理:

(1) 对文档进行格式转换,其中格式转换又分为两种情况,即:当教师希望学生不能下载该文档时,系统会把文档统一转换为 html 文档格式;而对于一些文档教师是允许学生对其进行下载的,那系统就会把该文档转为 pdf 文档格式;

(2) 当文档上传成功后,日后为了方便对其搜索,本系统会采用 Lucene 引擎对文档进行分类存储,并建立详细的索引信息,方便今后对这些文档进行全文检索;

(3) 把文档拷贝一份到一个固定的文件夹,以做文档备份。

需要补充的是在 Java 中要用到 Jacob 开发包来实现把 Word, PPT 文档转为 html 文档,具体的转换方式读者可以参考 Jacob 的官方网站。但是由于存在许多版本冲突问题,笔者经过多次试验后得出以下的系统配置: Jacob 用 Jacob1.9, 平台可以是 Windows XP 也可以是 Windows 2003, 同时 Office 版本可以是 2003, jdk 用 1.5.09, 还有一个很重要的问题就是一定要把 Jacob 的动态链接库文件放到 System32 文件夹中去^[8]。当使用 Jacob 来处理 Word, PPT 转为 PDF 时,需要安装 Adobe Acrobat 7.0 和 gs811w32.rar (PDF 转换时所需要的脚本 ps) 以及 postscript.rar (PDF 虚拟打印机的驱动), 实现的原理是: doc -> ps -> pdf -> office 2003 -> gs811w32 -> Adobe Acrobat 7 -> postscript -> 打印机, 从而输出 pdf 文件。还有就是对于用 Lucene 来处理文档时采用的是 POI 的方法来处理 Word, PPT 和 PDF 文档。

处理完上面的步骤之后,对于教师与管理员用户而言,已经可以在线观看与下载文档了,但不能对文档进行在线修改,需要用到 Webdav 协议并结合 Tomcat 的技术来实现: Tomcat 应该是 4.1 以上的版本中都自带了 Webdav 协议^[9], 可以在 Tomcat 5.0\server\lib\servlets-webdav.jar 中找到相应的信息, 并且 Tomcat 的 Webapp 目录中本身就自带一个 Webdav 文件夹^[10]。通过查找相关的资料可以发现保存在这个文件夹中的 Office 文档可以实现在线编辑。因此,笔者通过对 Webdav 的研究以及查看一些利用 Office2003 或者 OfficeXP 的一个 ActiveX 控件 owssupp.dll 来实现客户端对服务器端文档的编辑的功能^[11]。通过这个控件,客户端页面上的 JavaScript 就可以激活本地的 Office 软件,来实现打开、编辑 Office 文档。同时,笔者经过试验,发现在本系统的 web.xml 文件上加入以下代码就行:

```
<servlet>
```

```

    <servlet-name>webdav</servlet-name>
    <servlet-class>org.apache.catalina.servlets.WebdavServlet
</servlet-class>
    <init-param>
        <param-name>debug</param-name>
        <param-value>0</param-value>
    </init-param>
    <init-param>
        <param-name>listings</param-name>
        <param-value>true</param-value>
    </init-param>
    <init-param>
        <param-name>readonly</param-name>
        <param-value>>false</param-value>
    </init-param>
</servlet>
<servlet-mapping>
    <servlet-name>webdav</servlet-name>
    <url-pattern>/webdav/*</url-pattern>
</servlet-mapping>

```

同时,在 jsp 文件中采用如下 javascript 语句:

```

<script language="javascript">
Var openDocObj = new
ActiveXObject("SharePoint.OpenDocuments.2");
openDocObj.EditDocument("<%=path%>");
</script>

```

其中, path = "http://IP 地址:端口号/系统名称/webdav/+文档名称.doc/ppt"。但是这里知识修改了上传的文档原件的内容,对于学生端而言,他们能够浏览的是一些经过转换后的 html 文档以及 PDF 文档,如果单纯地按下 Office Word 的保存按钮是不会帮你促发转换事件的,这里的解决办法就是当按下在线修改某文档的连接后会去到一个叫 amendfile.jsp 的页面,在这个页面可以写入上面的 script 语句用来在线修改与保存,而当你关闭这个页面时会促发 onunload = "RunOnUnload()" 的函数,在这个函数中使用了 Amendfile.java 中的 convert 函数先去转换 doc 文档,再删除数据库中的相关信息,最后就是插入最新的文档的信息。

到目前为止,已经实现了教师权限的上传下载与在线修改功能。而对于学生端,则要是实现文档的安全性管理,由于之前已经把 Word 和 PPT 文档都转为 html 格式了,经过查阅相关资料可以发现通过设置 html 页面的 body 参数就可以实现该文档的防复制功能,同时也可以实现禁止另存、右键等功能。看到这里,发现可以巧妙地用 JSP 的 include 语句把一个 html 文件包含在一个 JSP 页面上,通过对该 JSP 页面的 body 参数的设置,就可以实现对该文档的防复制功能

了, body 参数的设置如下:

```
<body onmouseup="document.selection.empty()" oncopy="document.selection.empty()"
oncontextmenu="return false" onselectstart="return false" ondragstart="return false" onbeforecopy="return false">
```

以上的方法只能对 Word 文档适用, 对于由于 PPT 文档而言, 它转为 html 时会生成一个是 *.html 和一个 *.files 文件夹, 而 *.html 页面要调用 *.files 文件夹里面的 frame.html 页才能实现浏览这份文档, 但是如果这样实现的话浏览 PPT 文档时是可以实现对里面内容的复制的。通过对 files 文件夹里面的文件的研究, 可以发现里面有一个 fullscreen.html 的文件, 当直接访问这个文件时, 会用全屏的方式来浏览 PPT 文档, 并且不能复制里面的内容, 只有这样访问, 才能实现 PPT 内容的防复制功能。

完成了以上的步骤后, 一个文档处理系统的重要功能都基本上实现了, 但是还要注意使用 Windows 操作系统来浏览网页时, 所有的网页都会先下载到一个叫 Temporary Internet Files 的文件夹上, 然后再通过浏览器来读出网页的内容。因此, 为了防止一些对计算机比较熟悉的用户利用这一点来实现对文档的拷贝, 必须有所应对, 应对的方法是: 在 JSP 文件的头部设置一些特殊值使得浏览器不会缓存该 html 文件^[12], 设置如下:

```
<%
response.setHeader("Cache-Control","no-store");
response.setHeader("Pragma","no-cache");
response.setDateHeader("Expires",0);
%>
```

这样就不可能在 Temporary Internet Files 的文件夹中找到 html 文件了(单文件)。但是对于 PPT 转为 html 的方式呢? 答案是否定的, 所有的文件都缓存到了 Temporary Internet Files 的文件夹中, 但是幸运的是, 经过多次的实验, 笔者发现把这些文件一个不漏地拷贝回自己的硬盘上在打开的话是打开不了的, 从而也实现了对于 PPT 文档的安全性保护功能。

最后再谈谈搜索模块, 由于要涉及到全文检索, 因此笔者采用了 Lucene 搜索引擎, 实现的方式为先对上传的 Word, PPT 以及 PDF 文档进行格式转换, 都转换为 txt 文件格式, 然后对这些文档进行预处理, 接着就为这些文档中的内容建立索引并构建查询对象, 最后就可以在所建立的索引中进行查找了。这里需要补充的是:

对 Word 和 PPT 文档进行处理的是 POI, POI 包括一系列的 API, 它们可以操作基于 Microsoft OLE 2 Compound Document Format 的各种格式的文件, 可以通过 API 在 Java 中读写 Excel、Word 和 PPT 等文件。而

对与 PDF 文件, 则要采用 xpdf 来提取, xpdf 用来处理中文的 PDF 文档比较强大。

5 结束语

在 Java 平台上使用 Jacob 技术, 把 Office 文档转换为 html 文档, 不但可以结合 JSP 使得 Office 文档不能复制与下载, 而且在学生用户方可以完全不用安装 Office 工具都可以浏览, 并且他们查看 html 文件的速度肯定会比调用本地的 Office 来查看文档快许多, 占用资源量也是比较少的。这也是本系统的一个创新点, 纵观现在网上的办公系统的文档处理, 一般都是用 .NET 来实现的, 而本系统是用 J2EE 的 SSH+DWR 轻量级框架来实现, 这也体现了本系统的稳定性和安全性以及很好的可维护性这些优点, 并且, 如今的办公系统大多数除了在线修改一般都没有实现文档的安全性保护, 而本系统则可以包含这些功能, 同时本系统也增加了一些人性化的方面, 比如文档的全文搜索以及文档的备份管理等, 这样可以更加方便管理员用户使用这个系统, 用起来也更让管理员放心。

同时也发现了不足, 就是对于 flash 文件, 还没能够实现防下载功能, 而对于 Lucene 检索全文, 还需要更进一步地进行性能优化处理, 还有其他一些细节之处都是期待完善的。

参考文献:

- [1] Ruhl K. Tutorial: Jacob & Microsoft Word[EB/OL]. 2003-06. <http://www.land-of-kain.de/docs/jacob/>.
- [2] Jacob document[EB/OL]. 2007. <http://jacob-project.sourceforge.net>.
- [3] 李瑞, 李永刚. JAVA 中基于 JACOB 的 COM 组件调用研究[J]. 微计算机信息, 2007, 24(5): 168-170.
- [4] CN-JAVA. WebDAV 客户端开发实例[EB/OL]. 2006-02-14. <http://www.cn-java.com/www1/?action=viewnews-itemid-3439>.
- [5] 崔清华, 朱连军. 基于 B/S 架构的新型办公自动化系统设计[J]. 微计算机信息, 2007, 23(24): 280-282.
- [6] Daemen J, Rijmen V. The design of rijndael: AES-the advanced encryption standard[M]. [s. l.]. Springer-Verlag, 2002: 125-127.
- [7] 胡晓丹, 魏长军. 基于 Struts+Spring+Hibernate 架构的能源管理系统的研究[J]. 计算机与现代化, 2006(12): 25-30.
- [8] 冯正全. COM 原理概述[J]. 成都信息工程学院学报, 2004, 19(3): 398-402.
- [9] Qu Changtao, Engel T, Meinel C. Implementation of a document management system based on WebDAV protocol[C]// Proceedings of the 2000 IEEE International Conference on Management of Innovation and Technology ICMIT2000. [s.

从以上检测数据和检测图像中可以看出,该算法在进行各种常见攻击(如压缩、滤波等)和几何攻击(如旋转、缩放、平移等)后,都可以正确检测出水印,有些区域相关性为1,可以完全检测出水印,充分验证了该算法的可靠性。

同时,有部分区域不能正确检测出水印,这是由于待检测图像与其原图像的特征点相比较发生了漂移,从而导致提取区域与原图像发生了改变,影响了此区域水印的正确检测。但是,这并不影响本算法检测水印的相关性,因为只要有一个检测正确就可以判定水印存在。

5 结束语

以数字图像为研究对象,利用第二代水印的框架来解决鲁棒盲水印中的抗几何攻击的难题,并提出了一套基于Harris算子的鲁棒水印算法,对这些算法进行了实验验证。结果表明该算法不仅能抵抗压缩、滤波等常见的攻击,而且对几何攻击,如旋转、缩放等都有较好的鲁棒性。

此外,在水印不被察觉的基础上如何增大嵌入信息量的问题是有待进一步研究和改进的。

参考文献:

- [1] 孙圣和,陆哲明,牛夏牧. 数字水印技术及应用[M]. 北京:科学出版社,2004:67-98.
 - [2] 钮心沂. 信息隐藏与数字水印[M]. 北京:北京邮电大学出版社,2004:51-63.
 - [3] 俞龙江,牛夏牧,孙圣和. 一种旋转尺度变换和平移鲁棒水印算法[J]. 电子学报,2003,31(12A):2071-2073.
 - [4] Pereir S, Pun T. Robust template matching for affine resistant image watermarks[J]. IEEE Trans. Image Process, 2000, 9: 1123-1129.
 - [5] Bas P, Chassery J M, Macq B. Geometrically invariant watermarking sing feature points[J]. IEEE Trans. Image Processing, 2002, 11(9):1014-1028.
 - [6] 石磊,钟铭,洪帆. 抵抗几何变换的基于量化的水印技术[J]. 计算机辅助设计与图形学学报, 2004, 16(6):850-855.
 - [7] Simitopoulos D, Koutsonanos D E, Strintzis M G. Robust image water-marking based on generalized radon transformations[J]. IEEE Trans. Circuits Systems Video Technol, 2003, 13(8):732-745.
 - [8] Schmid C, Mohr R, Bauckhage C. Evaluation of interest point detectors[J]. International Journal of Computer Vision, 2000, 37(2):151-172.
 - [9] 李建. 抗几何攻击的数字图像水印技术的研究[D]. 南京:南京理工大学,2010.
 - [11] Lin Wei-Hung, Horng Shi-Jinn, Kao Tzong-Wann, et al. An Efficient Watermarking Method Based on Significant Difference of Wavelet Coefficient Quantization[J]. IEEE Trans. on Multimedia, 2008, 10(5):749-757.
 - [10] 李水乡,陈斌,赵亮,等. 快速Delaunay逐点插入网格生成算法[J]. 北京大学学报(自然科学版),2006(3):1-5.
 - [12] 于帅珍,冯丽平. 数字水印的关键技术[J]. 计算机技术与发展,2010,20(2):148-151.
-
- (上接第188页)
- 1.]:[s. n.],2000:752-757.
 - [10] 简清明. WebDAV及其在Web群件系统中的应用[J]. 安庆师范学院学报,2004(4):112-117.
 - [11] 江雨燕. Web环境下的在线协同编辑系统设计与实现[J]. 电子科技大学学报,2002,31(6):630-635.
 - [12] 张选平,谭小鹏,朱永虎. 面向Internet的文档管理系统的设计与实现[J]. 计算机工程与设计, 2004, 25(1):135-138.
-
- (上接第192页)
- dynamics[J]. Nature, 1976, 261:459-481.
 - [7] 王银花,柴晓东,周成鹏,等. 基于混沌序列和分数傅立叶变换的图像加密技术[J]. 计算机技术与发展,2006,16(9):213-215.
 - [8] 王银花,柴晓东,周成鹏,等. 基于分数傅立叶变换的盲数字水印算法[J]. 计算机技术与发展,2008,18(1):168-171.
 - [9] 杨倬,冯久超,方勇. 一种基于混沌和分数阶傅立叶变换的图像加密算法[J]. 计算机科学,2008(9):239-274.
 - [10] 曹荣,王颖,李象霖. 一种基于离散傅立叶变换域相位和幅度的数字水印算法[J]. 计算机应用,2005(11):2536-2543.
 - [11] 王炳锡,陈琦,邓峰森. 数字水印技术[M]. 西安:西安电子科技大学出版社,2003.
 - [12] 王远干,喻洪麟,黄良明. 基于M周期离散分数傅立叶变换的数字水印算法[J]. 计算机应用研究,2005(2):229-231.