基于 XML 多重签名的电子病历安全机制

陈乐君[1] 石 锐[1] 李初民[2]

(重庆大学计算机学院 重庆 400044)[1] (第三军医大学新桥医院 重庆 400038)[2]

摘 要 可扩展标志语言(简称 XML)在实现信息标准化、信息统一、信息的交换和共享上有其独特的技术优势,利用这些优势可以将 XML 技术运用到电子病历系统中。由于电子病历具有法律效力,因此需要一种机制来确保其安全。点对点的安全机制,例如 SSL 和 TLS,仅仅能确保数据的完整性与机密性,电子病历系统需要解决的安全问题是对多重签名的不可抵赖性。本文通过介绍签名文档及其元素的 XML 模式来显示如何确保不可抵赖性。

关键词 电子病历(EMR),XML Schema,XML 数字签名,多签名文档

Security Mechanisms for EMR Based on Multi-signed XML

CHEN Le-Jun[1] SHI Rui[1] LI Chu-Min[2]

(College of Computer, Chongqing Universtiy, Chongqing 400044)[1] (Xinqiao Hospital of Third Military Medical University, Chongqing 400038)[2]

Abstract Extensible Markup Language(XML)has its unique technical abvantages in standardization of information, uniformity of information, information exchange and sharign. XML technology can be imported in EMR system by using these dvantages. As EMR have the rofrce effect, Hence need for a mechanism to ensure its safety. Point-to-point security mechanisms such as SSLand TLS are just good for ensuring data integrity, confidentiality. Ensuring non-repudiation of multi-singed documents is a primary security problem for EMR system. This paper presents a XMLschema for the Singed document and its elements to ensuring non-repudiation. Finally the paper gives the programming cide of XML digital signature and verification implemented in visual sudio. Net 2003.

Keywords EMR, XML schema, XML digital signature, Multi-signed documen

1 引言

现代信息技术的迅猛发展给社会各个领域带来了巨大的革新,在医学领域也不例外,以计算机技术和网络技术为主的医院信息系统(HIS)也应运而生。实现电子病历(EMR)是医院信息系统的一个核心,也是它的一个重要研究课题。电子病历将传统的纸制病历完全电子化。为了确保电子病历内容的真实性,医生应该对其制作的电子病历文档进行数字签名,并确保数字签名与手写签名具有同样的法律效力。考虑到 XML 技术具有结构化的表示方法和广泛的适用性,我们将 XML 运用到电子病历系统中。

SSL 协议是用来确保网络中数据交换安全的一种安全机制,它是主要在点对点的通讯中保证数据的机密性与完整性。SSL 存在的问题是无法对信息发送的身份进行验证,如果发送者与接收者之间有多次的信息交换,接收者不能确定信息来源的真实性。在这种情况下,如果信息遭到篡改,将无法确定传输中哪一方修改了数据。

数字签名适合运用于保证信息的不可否认性,XML 数字签名的优势是支持多重签名和在更细粒度上的签名。XML 数字签名是一个 W3C (World Wide Web Consortium)规范[1]。该方案提供了数字签名的完整性、签名确认性和不可抵赖性,无论待签名的数据是存储在 XML 文档内部还是外部。

本文探讨了使用 XML 数字签名实现电子病历的多重签名。首先在第 2 节介绍了 XML 的安全机制,第 3 节介绍现有的电子病历系统,在第 4 节主要说明使用基于 XML Schema 的签名文档,第 5 节说明签名和验证的程序接口 API 接口。

2 XML 的安全

XML(eXtensible Markup Language 可扩展的标记语言)是由 W3C 组织在 1998 发布的一种标准,它有三个主要的优势:结构化的语言、数据存储与数据显示的分离。

一份 XML 文档是由元素、属性等标签组成。元素及其子元素是以一种树状结构的方式存储的,元素与属性的内容称为值,保证值的安全是 XML 在信息交换中要解决的问题,XML 安全技术包括 XML 加密、XML 数字签名、SAML 等,这些技术都提供了数据在网络中传输的安全机制。

XML 数字签名 XML 数字签名实现的基础是非对称加密技术,使用的是公钥加密算法与散列函数。对 XML 文档签名时,采用签名者的私钥进行加密,验证签名时采用该签名者的公钥进行解密,如果解密出的结果与原文档一致,说明数字签名正确。XML 签名的对象不仅可以是 XML 内部的元素,还可以是本地甚至网络上的文本和数据。XML 签名不仅可以像传统的数字签名技术一样对整个文档签名,还可以实现 XML 数字签名的特有功能,即在较细的粒度上如对文档的特定部分进行签名。另外,XML 支持多重签名,例如不同签名者可以对同一数据进行签名。根据签名元素和被签名对

陈乐君 硕士研究生,主要研究领域为信息安全;石 锐 研究生导师,主要研究领域为计算机软件与理论、数字图像处理技术;李初民 第三军医大学新桥医院信息科主任。

象之间的关系，XML 有三种签名方式：封装的签名（Enveloping Signature），被封装的签名（Enveloped Signature）和分离的签名（Detached Signature）。XML 数字签名的统一格式：

```
〈Signature ID?〉
  〈SignedInfo〉
    〈CanonicalizationMethod /〉
    〈SignatureMethod /〉
    (〈Reference URI?〉
      (〈Transforms〉)?
      〈DigestMethod〉
      〈DigestValue〉
    〈/Reference〉)+
  〈/SignedInfo〉
  〈SignatureValue〉
  (〈KeyInfo〉)?
  (〈Object ID?〉)*
〈/Signature〉
```

〈Signature〉是 XML 签名文档的根节点，它封装了主要元素〈 SignedInfo 〉、〈 SignatureValue 〉以及可选元素比如〈KeyInfo〉。〈 SignedInfo 〉包含标准化方法和签名算法，〈Reference〉指向需要签名的任何对象，〈Transforms〉表示了在签名前可能对被签名对象所要做的转换，〈 DigestValue 〉存放对〈 SignedInfo 〉元素计算的摘要信息，〈KeyInfo〉（可选）用来标识或存放验证签名所需的密钥。

XML 数字签名的步骤分为引用〈Reference〉的创建和签名〈Signature〉的创建。

〈Reference〉的创建　利用 URI 定位到引用的资源，并获得其数据，然后利用〈Transforms〉元素中指定的方法对引用的资源做转换，再利用〈DigestMethod〉元素中指定的方法对转换后的资源做摘要，获得的结果用于创建〈DigestValue〉元素。将以上元素组合成〈Reference〉元素。

〈Signature〉的创建　创建〈SignedInfo〉元素，并将之前创建的〈Reference〉元素包含进来。根据指定的〈CanonicalizationMethod〉将刚〈SignedInfo〉元素标准化。然后按〈SignatureMethod〉的方法对标准化后的〈SignedInfo〉元素做签名，所得到的结果用于创建〈SignatureValue〉元素。将〈SignedInfo〉，〈SignatureValue〉以及可选的〈KeyInfo〉和〈Object〉元素组合成〈Signature〉元素。

验证 XML 数字签名时也需要先验证引用〈Reference〉元素，然后验证签名〈Signature〉。只有两次验证都会正确才说明签名是合法的。

3　电子病历数字签名系统概述

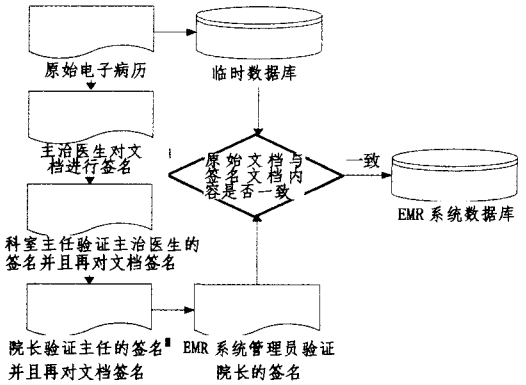

图 1　电子病历数字签名系统

电子病历（Electronic Medical Record，简称 EMR）。电子病历是指将传统的纸病历完全电子化，并超越纸病历的管理模式，提供电子贮存、查询、统计、数据交换等。

在电子病历中，涉及到的签名情况比较多，常见的有以下几种：一是临床病历形成过程中各种电子医疗文书的签名，如手术病人的手术知情同意书、术前术后小结、术中记录等；二是各种检查申请单、检验单医生和承检人签名；三是医疗护理执行过程中直接责任人签名。

电子病历本身具有法律效力，为了避免医患纠纷要求医生必须对其书写的病历内容负责。对于同一份病历可能需要多位医生的签名，这样的签名就是多重签名。本章以一份“入院记录”的病历文档为例来介绍本系统如何对病历文档进行多重签名。

一份有效的“入院记录”文档一般需要三重签名。签名者包括主治医生、科室主任、院长，一份病历文档必须经过三个人的签名并最终拥有三份数字签名。

在主治医生完成对病人的诊断后需要填写病人的病历，并对其进行数字签名。主治医生书写的病历会生成两封拷贝，未签名的原始病历将会存储在临时数据库中，签了名的病历将通过本系统发送给科室主任。科室主任的职责主要是检查主治医生的签名是否合法，然后再病历签名并发送给上级部门主管，比如院长。院长的职责与科室主任基本一样，检查两份签名的正确性，如果合法，再次签名后发送给 EMR 系统服务器。EMR 系统管理员检查原始病历与院长上传的病历内容是否一致，如果相符则将带有数字签名的病历文档存储到永久数据库中。

4　XML 电子病历的多重签名

XML 技术运用于电子病历领域，具有以下优点：

• XML 是一种元语言，它采用层次化结构的、面向对象的方法描述数据对象的结构，通过 XML 的标记来定义电子病历的数据对象，非常适合描述含有复杂内容的病历文档；

• XML 具有平台无关性，适合在不同平台下进行病历数据的交换，在不同环境下进行系统开发；

• 在临床医疗中，需要方便地实现数据交换，数据处理以及多种数据视图，XML 具有数据和表现样式相分离的特点，XML 文件本身可以只包含数据信息，而由与数据相对应的样式文件表现其显示格式，通过两者的结合，生成包含数据和样式的文件，不同的样式文件可以得到不同的数据视图；

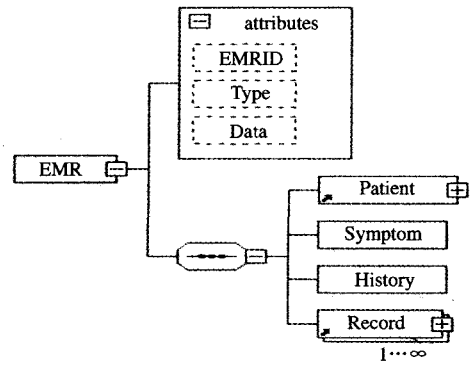

图 2　原始电子病历文档

• XML 对于机器和人来说，都具有很强的可读性，即使没有表现样式，用户也能够大致明白病历文档的内容；

• XML 有一整套体系支撑，提供丰富的技术支持，有利于电子病历的实现开发。

利用这些优势本系统使用 XML 作为电子病历存储和交

换的媒介，从图 1 中可以看到，一份电子病历的数字签名是从主治医生到永久数据库之间的有序过程。我们下面使用 XML Schema2 来描述 XML 电子病历文档签名的细节。

图 2 描述的是“入院病历”的原始 XML 文档，文档的根元素是〈EMR〉，〈EMR〉拥有三个属性和四个子元素。其中〈EMRID〉是该文档的惟一标识，〈Type〉是病历的类型，〈Date〉是电子病历创建的日期，〈Patient〉元素存储的是病人的信息，〈Symptom〉是病人的症状，〈History〉是病史，〈Record〉元素是医生书写的诊断信息，其属性与子元素如图 3 所示。

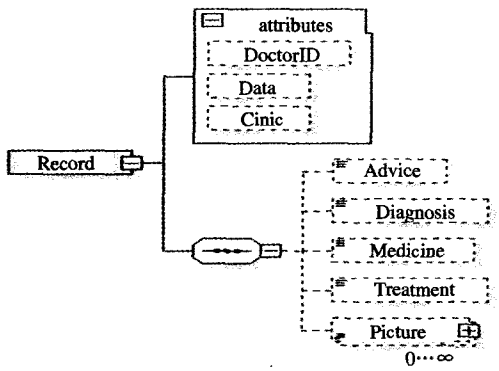

图 3 〈Record〉的属性及其子元素

〈Record〉元素有三个属性和四个子元素，其中〈DoctorID〉是本系统中主治医生的身份惟一标识，〈Cinic〉是医生所属科室，〈Advice〉是医嘱信息，〈Diagnosis〉是诊断信息，〈Medicine〉是用药信息，〈Treatment〉是治疗信息，〈Picture〉元素存储的是病人的影像资料。从图中我们可以看出〈Record〉元素存储的是电子病历文档中最重要的信息，也是需要医生、主任和院长等领导进行签名的部分。使用本系统对〈Record〉元素进行数字签名，签名后文档的 XML Schema 如图 4 所示。

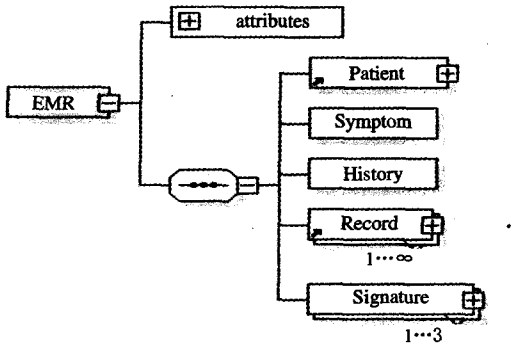

图 4 签名后的电子病历文档

在电子病历的传输中，文档不断地接受新的签名和验证，最终合法的数字签名文档至少有三个〈Signature〉元素。为了不破坏原始文档的结构，本系统使用的是封装的签名，使〈Signature〉元素成为〈EMR〉根元素的子元素节点。

5 数字签名系统的 API

本系统主要有两个程序接口，分别是注册登记接口和数字签名接口。注册登记类 Register 主要用来识别用户信息、产生密钥对，并与服务器进行通讯。数字签名接口的内容如图 5 的 UML 类图所示。

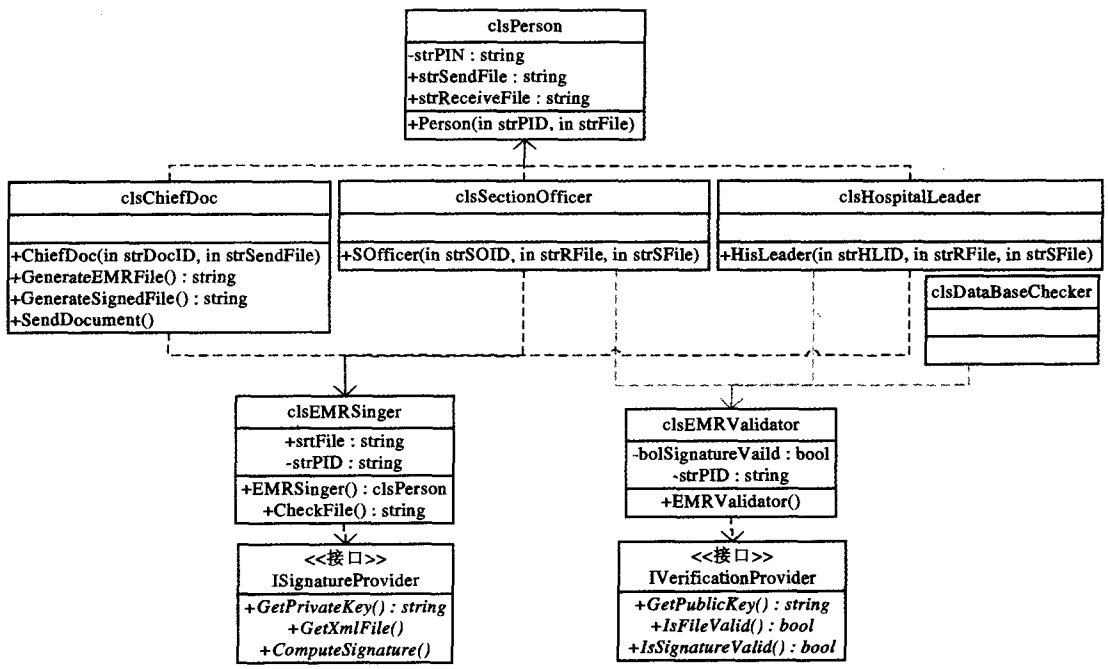

图 5 数字签名与验证使用的类库

clsChiefDoc 类 主治医生使用该类来创建电子病历文档，并将其发送到系统的临时数据库中。该类还完成主治医生对文档的签名工作，并将签名后的文档通过 TCP 协议发送到科室主任端。

clsSectionOfficer 和 clsHospitalLeader 类 这两个类主要用来实现上级主管的签名和验证机制。文档最终通过 TCP 被发送到系统数据库。

clsDataBaseChecker 类 EMR 系统管理员在病历文档存库之前对文档做最后一次检查。他需要验证两个签名，一个是主治医生对文档的初始签名，一个是医院院长对文档的最终签名。最后，系统将检查零时数据库中存储的初始文档与

（下转第 170 页）

+1=2；(3,5)∉A={(1,2)，(1,5)，(2,5)，(2,4)，(2,3)，(1,4)，(1,3)，(3,5)}，此时让n_{35}=1。

对事件9，得到2-项目集{ $I1$，$I2$}，{ $I1$，$I3$}，{ $I2$，$I3$}。(1，2)∈A={(1,2)，(1,5)，(2,5)，(2,4)，(2,3)，(1,4)，(1,3)，(3,5)}，此时让n_{12}=3+1=4；(1，2)A={(1，2)，(1,5)，(2,5)，(2,4)，(2,3)，(1,4)，(1,3)，(3,5)}，此时让n_{12}=3+1=4；(2，3)∈A={(1,2)，(1,5)，(2,5)，(2,4)，(2,3)，(1,4)，(1,3)，(3,5)}，此时让n_{23}=3+1=4。

至此，通过敏捷分桶算法处理已可得到如表2所示的桶分配及其内容示例表。

表2 2-项目集敏捷分桶的桶分配及其内容示例表

桶地址	(1,2)	(1,5)	(2,5)	(2,4)	(2,3)	(1,4)	(1,3)	(3,5)
桶计数	4	2	2	2	4	1	4	1
桶内容	{ I1，I2}	{ I1，I5}	{ I2，I5}	{ I2，I4}	{ I2，I3}	{ I1，I4}	{ I1，I3}	{ I3，I5}
	{ I1，I2}	{ I1，I5}	{ I2，I5}	{ I2，I4}	{ I2，I3}		{ I1，I3}	
	{ I1，I2}				{ I2，I3}		{ I1，I3}	
	{ I1，I2}				{ I2，I3}		{ I1，I3}	

第4步，2-频繁项目集L_2生成处理。显然，只要用户提供2-频繁项目集判据的最小支持度Minsupport值，就可据此求得相应的2-频繁项目集L_2。例如，若取最小支持度Minsupport为0.2(即4/(4+2+2+2+4+1+ 4+1)之值)，就可得所求2-频繁项目集L_2为{{ $I1$,$I2$}，{ $I2$,$I3$}，{ $I1$，$I3$}}。

应当指出：表2说明，敏捷分桶算法所得桶分配及其内容，地址含义明确，数据简明无歧(义)，处理敏捷高效。而与此形成鲜明对比是：尽管同样针对表1所示事务数据库的同一数据源，但文[2]通过Hash函数$(10x+y)$ mod 7所得如表3所示的桶分配及其内容，却“地址含义不明，数据可能有歧，处理繁琐低效”(不难看出，表3中，桶地址为0的桶发生了Hash冲突，因为此桶内容有两个不同2-项目集{$I1$,$I4$}和{$I3$,$I5$}，而采用敏捷分桶的算法，可有效地将这两个不同的项目集各放在两个不同桶中)。

表3 2-项目集Hash函数的桶分配示例

桶地址	0	1	2	3	4	5	6
桶计数	2	2	4	2	2	4	4
桶内容	{ I1，I4}	{ I1，I5}	{ I2，I3}	{ I2，I4}	{ I2，I5}	{ I1，I2}	{ I1，I3}
	{ I3，I5}	{ I1，I5}	{ I2，I3}	{ I2，I4}	{ I2，I5}	{ I1，I2}	{ I1，I3}
			{ I2，I3}			{ I1，I2}	{ I1，I3}
			{ I2，I3}			{ I1，I2}	{ I1，I3}

结论 本文提出的生成频繁项目集的敏捷分桶算法，可有效克服基于Hash函数的频繁项目集生成传统算法的上述两个主要缺点。同时，本敏捷分桶算法还可进一步予以改进。例如(我们拟另文阐述)，1)可利用其桶地址与桶内容的一一映射关系，基于所设计的用其桶地址间接而唯一地表征桶内容的有效描述方法，就可通过只对桶地址进行相应等价处理来直接找出所求频繁项目集，从而可再次提高本算法效率；2)对n-频繁项目集，还可利用适当的树结构(譬如n叉树)来描述和存放其各桶地址，就可采用基于树结构的桶地址快速查找技术，来更加提高本算法效率。

参 考 文 献

1 Park J. Efficient parallel data mining for association rules. In: Proc. of the 4th Int Conf on Information and Knowledge Management. Baltimore, USA, 1995. 31～36
2 毛国君，段立娟，王实，等. 数据挖掘原理与算法[M]. 北京：清华大学出版社，2005
3 Agrawal R, Srikant R. Fast algorithms for mining association rules [C]. In: Proceeding of the 20th International Conference on Very Large Databases, 1994. 487～499
4 何小卫. Apriori算法强项集产生的二维哈希算法[J]. 计算机与现代化，2003,4:10～12
5 张江，傅鹤岗. 基于关联规则的二维哈希算法的改进[J]. 计算机工程与设计，2225，26(8)：2178～2179

(上接第138页)

签名文档内容的一致性，最后将病历存储到系统数据库中。

ISignatureProvider接口 该接口用来对任意的XML文档的内容进行数字签名。clsEMRSinger类实现了这个接口来对电子病历文档数字签名。

IVerificationProvider接口 该接口用来验证任意XML文档的数字签名。clsEMRValidator类实现了这个接口来验证电子病历文档签名的正确性。

总结 为了保证对信息的不可抵赖性，本文一开始就引入了XML数字签名的概念。我们用一系列XML Schema图来说明XML签名如何保证对病历文档的不可抵赖性。然后给出了目前系统实现的步骤与方法。文章最后给出了系统的API接口，实现了电子病历文档的签名与验证。

参 考 文 献

1 W3C. XML Signature Syntax and Processing [R]. http: //www.w3.org/TR/XMLDSIG,2003
2 W3C. XML Schema [EB/OL] http://www.w3.org/XML/Schema
3 Microsoft. netframework [R]. http: //msdn.microsoft/netframework/security,2002
4 BoyerJ. Canonical XML Version 1.0 [S]. RFC 3076,March 2001
5 李凤银. 基于XML的多人数字签名技术研究与实现[J]. 计算机科学，2004,1(7):109～114
6 李双庆，游莲，古平，等. 一种基于XML的数据交换中间件技术[J]. 计算机科学，2003
7 马永恒，熊前兴，杨金娥. W3C XML Schema模式的设计方法研究[J]. 计算机应用研究，2006(5)
8 肖林. 浅谈电子病历[J]. 档案管理，2000,4(1)
9 严维良，于津. XML在电子病历中的应用[J]. 汕头大学学报，2003,15(3)
10 张勇，冯玉才. XML数字签名技术及其在java中的具体实现[J]. 计算机应用，2003,23(9)
11 李浩，孙统风，孟现飞，等. 基于面向对象思想构建XML Schema[J]. 微机发展，2003,13(6):59～64