以数字档案馆（室）为中心的可信电子档案体系建设研究

研 究 报 告

江苏省档案局（馆）

2014 年 12 月 8 日

目　录

一、概述.....1
1.1 国内外可信电子档案（文件）研究现状.....1
1.2 课题研究的内容及其意义.....2
二、数字签名与CA中心：可信电子档案的技术实现.....4
2.1 数字签名及算法.....4
2.2 CA中心.....9
三、数字档案可信认证体系建设.....10
3.1 数字档案CA中心.....10
3.2 CA中心安全模块说明.....12
3.2.1 SSL模块.....12
3.2.2 SVS模块.....13
四、OFD格式及基于PDF/A的可信电子档案格式.....17
4.1OFD格式[13] [14].....17
4.2 电子档案格式标准：PDF/A[15][16].....18
4.3 PDF/A与电子签名.....19
五、以数字档案馆（室）为中心的可信电子档案体系建设.....20
5.1 电子文件产生环节的可信性保证.....22
5.2 数字档案馆OAIS功能模型概述.....23
5.3 电子档案归档环节的可信性保证.....25
5.4 电子档案移交环节的可信性保证.....26
5.5 电子档案迁移环节的可信性保证.....28
5.6电子档案利用环节的可信性保证.....31
六、江苏省电子档案中心可信体系建设实践.....32
6.1总体架构.....32
6.2移交接收环节.....36
6.3迁移环节.....46
6.4利用环节.....48

七、下一步的研究方向.....49
八、参考文献.....51

一、概述

1.1 国内外可信电子档案（文件）研究现状

近年来，英国陆续实施了“无缝工作流项目”、“数字连续性工程”、“网页连续性工程”、信息管理评估项目，整合不同机构的技能和专长，促进政府电子文件管理职权合法化；美国实施了“电子文件管理动议”、“电子文件档案馆”、“联邦机构文件管理框架”三大项目，涉及电子文件运动生命周期、档案馆电子文件长久保存和利用、前端业务活动文件管理等电子文件关键管理环节、领域与难题的研究；澳大利亚启动了“数字文件保存战略”、澳大拉西亚数字文件管理动议等项目研究，探索政府机构在数字文件生成、保存和利用方面采用统一的途径，以确保重要文化遗产的问责和长期保护。[1-7]

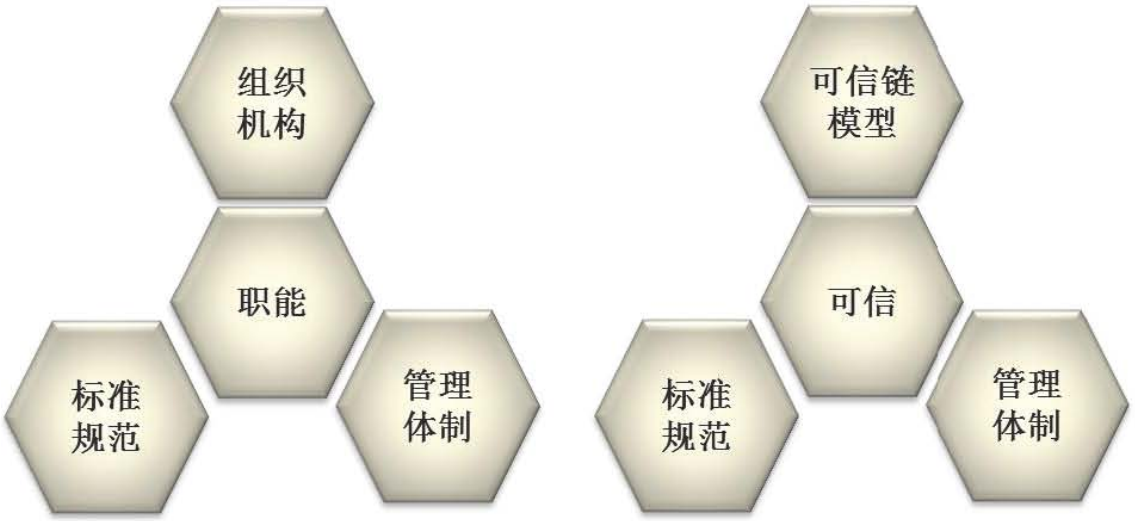
在我国，2008 年中国人民大学信息资源管理学院完成的《关于加强我国电子文件科学管理的报告》，对电子文件管理中客观存在的“失存、失用、失控、失信”现象，从民族文化传承和国家信息资源国家控制力的战略高度深入进行了研究，全面梳理了国内外近年来电子文件管理现状及前沿研究成果，对电子文件管理严峻现实及其原因进行了客观深入的分析，提出了符合中国国情特点的、贯穿成体系的公共治理思想、软治理与硬治理相结合的思想、强化集中统一管理的思想、党政文件档案统一管理的思想、文件档案一体化管理的思想、全程管理和前端控制的思想、电子文件资产化管理的思想、以及电子文件管理“四个纳入”的思想等电子文件管理理论，得到温家宝总理的充分肯定并作了专门批示；2009 年底，中办和国办联合下发了《电子文件管理暂行办法》，标志着电子文件管理纳入国家战略。《国家电子文件管理工作规划（2011-2015）》中明确将建立健全电子文件管理体制机制作为一项主要任务。近年来包括电子文件在内的文件管理国际标准的陆续出台，进一步推动了电子文件管理的科学化、规范化进程。在国家电子文件管理联席会议的统筹规划下，国家有关部委、各地方也正在开展电子文件管理的多项研究。[8][9]

1.2 课题研究的内容及其意义

研究电子档案的信任管理机制，首先应判明传统的档案管理与网络环境下电子档案管理的区别。无论何种管理环境、何种管理模式下的档案价值均表现在对其使用的依赖性上，即档案凭证价值作用。传统档案的依赖性取决于档案管理的组织机构及其机构的职能，通过组织机构的管理体制和制度，保障档案的凭证价值。网络环境下的依赖性取决于电子档案的可信性以及形成可信电子档案的可信任链的体系模型。如图 1 所示。

网络环境下电子档案的可信性保障能力是通过一系列与电子档案管理相关的业务和技术等标准规范，与管理体制相关的组织机构、职权划分等运行机制，以及支撑电子档案可信性的可信任链的传递模式和模型而决定的。

可信性是在真实性、完整性、有效性、安全性、可靠性、可预测性、可控性等众多概念上发展起来的新概念，是客观对象诸多属性在人们心目中的一个综合反映。学者们试图从不同角度、不同层次去诠释“可信性”，但尚未形成共识。



a 传统档案管理模式　　b 网络环境下电子档案管理模式

图 1　档案管理模式

对于电子档案的可信性认识需要在实践中不断认识和把握，一般认为“可信”是指一个实体在实现既定目标的过程中，行为及结果可以预期，强调目标与现实相符，强调行为和结果的可预测性和可控制性。

由此可以认为，电子档案的可信性研究首先要明确网络环境下电子档案的可信性内涵及其综合化的属性概念；其次是研究电子档案实现价值目标的过程，即电子档案可信任链的传递模式，并将其管理的全过程置于全方位的信任链传递模型中，研究并提出信任传递的形式化定义和管理规则，是本课题的研究方法。

本课题研究的技术路线可依图 2“可信性”概念的研究模型所示。

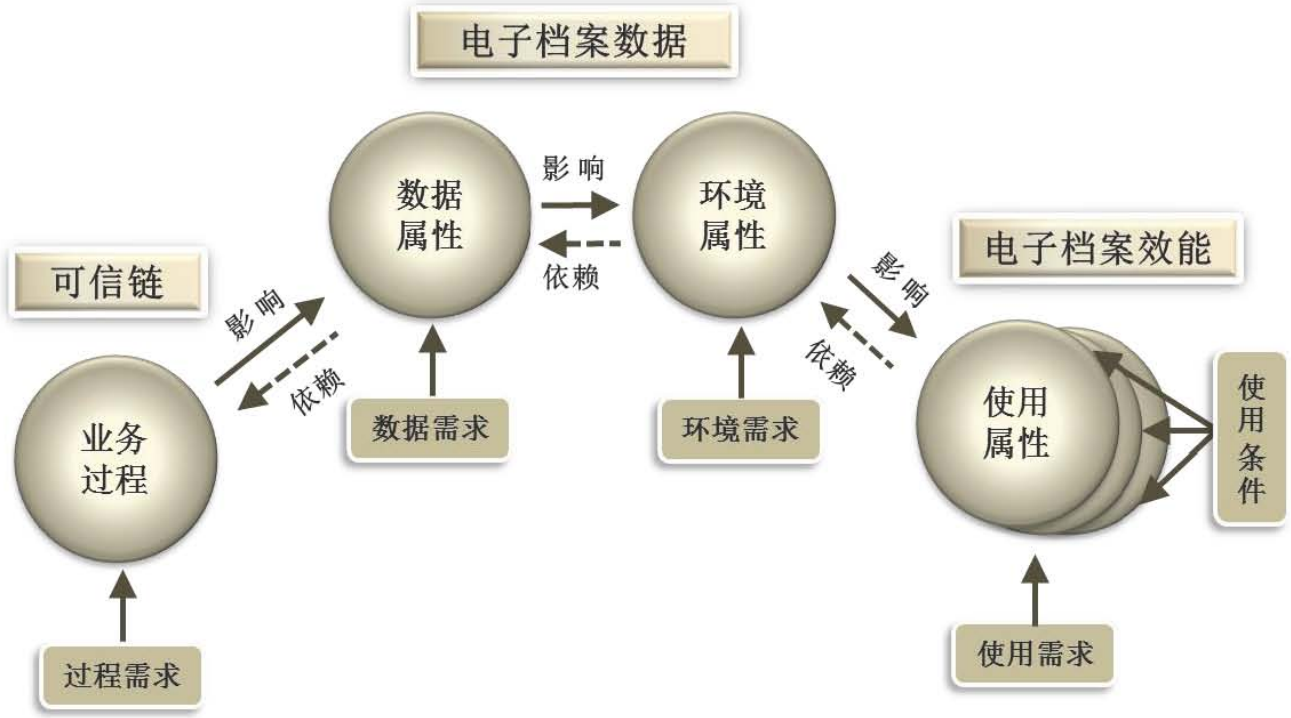


图 2 “可信性”概念的研究模型

可信性的概念模型图描述了电子档案在数字档案馆业务环境中，电子档案数据与可信的电子档案传递机制、电子档案的效能目标之间的依存关系，通过对该依存关系的不断剖析，形成本课题研究的技术路线。

可信链是电子档案的传递机制，可描述为从电子档案初始生成作为“可信根”出发，在每一次动态变化时，将这种信任状态通过传递的方式进行保持和维护，并在每个动态变化的节点进行可信任行为控制并记录控制状态信息，这种传递过程就是信任链的传递机制。[10]

电子档案数据所依赖的数据属性和环境属性，是保证电子档案数据质量的关键要素。如何在可信链的传递过程中约束和规范电子档案主体的数据结构和性质，约束和规范客体对电子档案的行为处置，支撑其电子档案在可控使用条件下

的效能作用，是本课题研究如何解决电子档案价值作用的根本。

可信的电子档案是数字档案馆建设的基础。随着各类电子政务应用系统的发展，不少机关单位已逐步建立起各自的可信电子文件体系。这些系统所形成的可信电子文件按法定要求都将归档到各自的数字档案室系统并最终移交到数字档案馆进行长期保管和利用。为了保证电子档案的可信，必须建立一个以数字档案馆（室）为中心的档案体系。

这个体系建设有两个方面内容：

① 制度体系建设：为保证电子档案的可信，必须对电子文件归档、移交、保管和利用等管理制定一整套的业务规则体系。

② 技术体系建设：为保证业务规则操作的可行性，必须建立一个开放、可靠的技术保障体系，包括必要的设施技术标准和技术体系保障。

本课题旨在以技术角度研究电子政务网和互联网环境下电子档案的可信体系建设，明确电子档案的可信性应具备的属性和特征，研究满足其属性和特征的管理体制和管理模式，将分别从以下两个方面展开。

① 以数字档案馆（室）为中心，如何进行可信电子档案的移交和接收，并为公众提供可信、可控的电子档案查询和利用服务；

② 研究在网络环境下，如何围绕数字档案馆（室）收、存、管、用业务，建立一套完整的的可信电子档案管理模式以及支撑该模式的技术方案。

二、数字签名与 CA 中心：可信电子档案的技术实现

可信电子档案是指可信电子文件按法定程序在可信环境下归档移交到数字档案馆（室）的电子文件，电子档案的可信可以通过权威 CA 中心签署的数字证书来证明。

2.1 数字签名及算法

数字签名是附加在数据单元上的一些数据，是对数据单元所作的密码变换。这种数据或变换允许数据单元的接收者用以确认数据单元的来源和数据单元的完整性并保护数据，防止被人(例如接收者)进行伪造。基于公钥密码体制和私钥

密码体制都可以获得数字签名，数字签名算法有RSA、ElGamal、Fiat-Shamir、Guillou-Quisquarter、Schnorr、Ong-Schnorr-Shamir数字签名算法、DES/DSA、椭圆曲线数字签名算法和有限自动机数字签名算法等。数字签名是个加密的过程，数字签名验证是个解密的过程[11]。以下列举几种常用的数字签名算法。

（1）Hash函数：

把哈希函数值h(x)看成x的消息摘要(message digest)，当x中任一bit变化时都将引起哈希函数值的变化。这样就可以用对h(x)的签名代替对x签名。

散列函数h是一公开函数，用于将任意长的消息m映射为较短的、固定长度的一个值h(m)，作为认证符，称函数值h(m)为杂凑值、杂凑码或消息摘要。杂凑码是消息中所有比特的函数，改变消息中任何一个比特或几个比特都会使杂凑码发生改变。

（2）RSA 算法：

RSA 加密解密是可交换的 。

给定RSA方案{(e,R),(d,p,q)}，要签名消息M，则计算S = Md(mod R)；要验证签名，计算M=Se(mod R)= Me.d(mod R) = M(mod R)。使用RSA加密、认证时，使用发送者的私钥签名一个消息，使用接收者的公钥加密消息，看起来，一个消息可用RSA加密、签名而不改变大小，实际上加密使用的是消息接收者的模，签名是消息发送者的模，后着可能比前者小，签名常使用HASH函数值。

（3）El Gamal Signature Scheme算法

ElGamal 加密算法是不可交换的 ，存在一个相关的签名算法，安全性是基于计算离散对数的困难性，方案的密钥生成是相同的。

有个共享的素数 p，公开的本原根 a ，每个用户选择一个随机数作为私钥x，计算各自的公开密钥：y =ax mod p；公钥是 (y,a,p) ；私钥是(x)；签名消息 M：

选择随机数 k，GCD(k,p-1)=1

计算 K = ak(mod p)

用 Euclidean (inverse) 扩展算法求 S:

M = x.K + k.S mod (p-1)；即求

S = k-1(M - x.K) mod (p-1)

签名是 (K,S)

k 应该被销毁

同 ElGamal 加密方案， 签名信息也是消息的 2 倍

验证 (K,S) 是 对 M 的签名:

yK.KSmod p = aMmod p

（4）DES/DSA (Digital Signature Algorithm)

US Federal Govt approved signature scheme (FIPS PUB 186)

使用 SHA hash alg

首先选取公开参数 (p , q , g) :

选取大素数 p = 2L

L= 512 to 1024 bits（64 倍数）

选取 q , 160 bit 素因子 (of p-1)

选择 g = h(p-1)/q

对任何 h<p-1, h(p-1)/q(mod p)>1

每个用户选取私钥并计算他们的公钥:

选取 x<q

计算 y = gx(mod p)

签名消息 M

生成随机签名蜜钥 k, k<q

计算, r = (gk(mod p))(mod q)

s = k-1.SHA(M)+ x.r (mod q)

发送签名 (r,s)及消息 M

验证签名, 计算:

w = s-1(mod q)

u1= (SHA(M).w)(mod q)

u2= r.w(mod q)

v = (gu1.yu2(mod p))(mod q)

v=r 签名有效

ElGamal 算法是基于离散对数问题的困难性，这样的函数称为单向函数。单向函数不能直接用作密码体制，因为如果用单向函数对明文进行加密，即使是合法的接收者也不能还原出明文，因为单向函数的逆运算是困难的。

与 DES 算法相比较,在加解密的处理效率上 DES 算法优于 RSA 算法,DES 密钥的长度只有 56 位，可以利用软件和硬件实现高速处理，RSA 算法在处理速度上明显慢于 DES 算法。在密钥管理上，RSA 算法比 DES 算法更加优越，RSA 算法可采用公开形式分配加密密钥，对加密密钥的更新容易，并且对不同的通信对象,只需对自己的解密密铜保密即可,DES 算法要求通信前对密钥进行秘密分配，密钥的更换困难，对不同的通信对象， DES 需产生和保管不同的密钥。另外，DES 算法从原理上不可能实现数字签名和身份认证，RSA 算法能够容易地进行数字签名或身份认证。

课题组根据 GB/T 17902 系列标准《信息技术 安全技术 带附录的数字签名》、GB/T 25061-2010《信息安全技术 公钥基础设施 XML 数字签名语法与处理规范》等相关标准，数字签名算法采用经过经典 RSA 算法改良后的 SHARSA 算法。数字签名方案的组成：

由五元组组成，即{P，S，K，Sig，Ver}，P:明文空间，S:签名空间， K:密钥空间，Sign:签名算法，Ver:验证算法。

数字签名的基本原理：

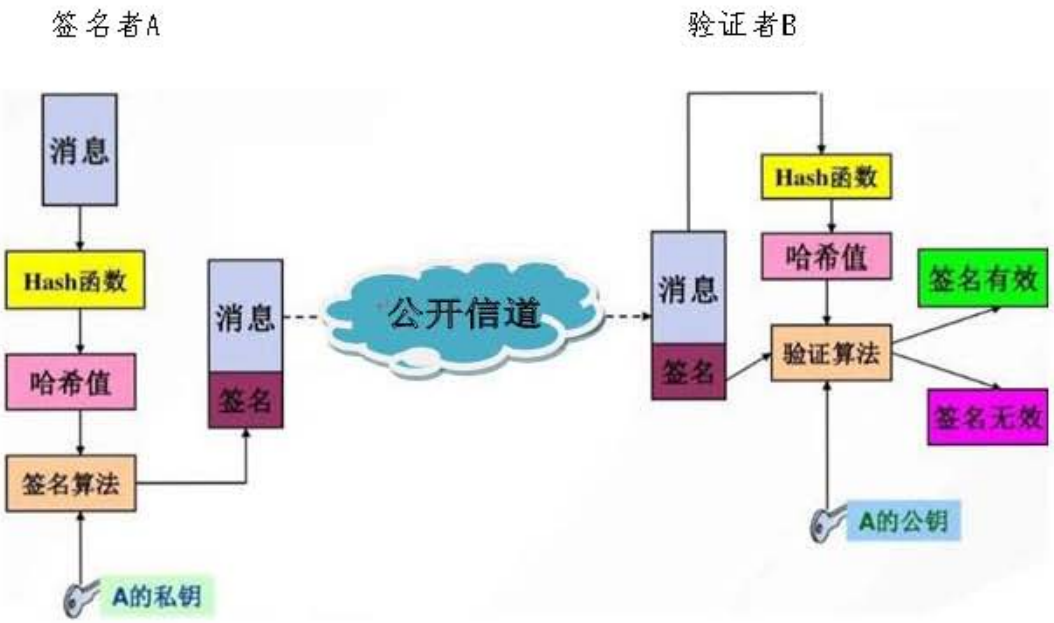


图 3　数字签名的基本原理

SHARSA 数字签名方案（初始化）：

选取两个大素数 p 和 q ，两个数长度接近，1024 位；

计算 n=p*q，ψ(n)=(p-1)(q-1)；

随机选取整数 e(1<e<ψ(n)) ，满足 gcd(e，ψ(n))=1；

计算 d，满足 d*e=1 (mod ψ(n)) 。

注：n 公开， p 和 q 保密；e 为公钥， d 为私钥。

SHARSA 数字签名方案(签名和验证)：

签名算法

利用一个安全的 Hash 函数 h 来产生消息摘要 h(m)。

用签名算法计算签名 s=Sigk(m)=h(m)d mod n。

验证算法

首先利用一个安全的 Hash 函数 h 计算消息摘要 h(m)。

用检验等式 h(m)mod n=se mod n 是否成立，若相等，签名有效，否则，签名无效。

SHARSA 数字签名方案(正确性)

由于 s=h(m)d mod n

d*e=1 (mod ψ (n))

所以 Semod n=h(m)ed mod n

=h(m)kψ(n)+1 mod n

= h(m) *h(m) kψ(n) (n) mod n

=h(m)* (h(m) ψ(n)) k mod n

= h(m)

2.2 CA中心

CA中心，即认证中心，又称CA机构，证书授权中心(Certificate Authority)，或称证书授权机构，作为权威的、可信赖的、公正的机构，承担公钥体系中公钥合法性检验的责任。

CA中心主要职责是颁发和管理数字证书。其中心任务是颁发数字证书，并履行用户身份认证的责任。CA中心的数字签名使得攻击者不能伪造和篡改证书。CA中心在安全责任分散、运行安全管理、系统安全、物理安全、数据库安全、人员安全、密钥管理等方面，需要十分严格的政策和规程，要有完善的安全机制。另外要有完善的安全审计、运行监控、容灾备份、事故快速反应等实施措施，对身份认证、访问控制、防病毒、防攻击等方面也要有强大的工具支撑。

基于公开密钥算法的数字签名技术和加密技术，为CA建设提供了理论依据和技术可行性;《中华人民共和国电子签名法》的颁布和实施为数字签名的使用提供了法律依据，使得数字签名与传统的手工签字和盖章具有了同等的法律效力。

目前，基于CA的数字档案馆认证系统实际应用中主要存在两种类型的CA:① 独立的第三方CA ② 自建CA。

这两种类型的CA在实际使用过程中各有优劣，以下通过表格将进行分析和比较，如表 1 所示。

表 1　两种 CA 应用方式比较

项目比较	独立权威的第三方 CA	自建 CA
建设/使用成本	一般由第三方 CA 按用户收取年费，建设投入和证书使用成本较高	一次投入建设成本，建成后可为用户免费发放证书，成本较低
证书的有效性检查	由第三方 CA 提供的 CRL（证书撤销列表，实效性较差）或者 OCSP（在线证书状态查询，依赖于外部系统，易形成性能瓶颈）	与业务系统紧密结合，对证书有效性的控制和检查实时准确
定制化/灵活性	按第三方归档的流程申请证书，受制于第三方的系统，还需要将用户证书导入业务系统中，灵活性较差	可以与业务系统紧密结合，满足业务系统的定制化需求，灵活性高
故障响应时间	故障解决依赖于第三方	故障响应及时
认证资质	已通过相关单位审批，有电子认证服务资质，权威性强	建成后，对外提供电子认证服务时需要通过相关部门的审批

三、数字档案可信认证体系建设

3.1 数字档案 CA 中心

本课题考虑到自主可控，安全管理，自行维护等方面情况，采用的是自建 CA 中心。基于国际标准，参照国际领先的 CA 系统的设计思想，系统采用模块化结构设计，由最终用户、RA 管理员、CA 管理员、注册中心（RA）、认证中心（CA）等构成，其中注册中心（RA）和认证中心（CA）又包含相应的模块，系统架构如图 4 所示。

CA 架构包括 PKI 结构、高强度抗攻击的公开加解密算法、数字签名技术、身份认证技术、运行安全管理技术、可靠的信任责任体系等等。从业务流程涉及的角色看，包括认证机构、数字证书库和黑名单库、密钥托管处理系统、证书目录服务、证书审批和作废处理系统。[12]

利用数字证书、PKI、对称加密算法、数字签名、数字信封等加密技术，建

立起安全程度极高的加解密和身份认证系统，确保信息安全有效，从而使信息除发送方和接收方外，不被其他方知悉（保密性）；保证传输过程中不被篡改（完整性和一致性）；发送方确信接收方不是假冒的（身份的真实性和不可伪装性）；发送方不能否认自己的发送行为（不可抵赖性）。

数字档案 CA 中心在密码管理方面的作用：

① 自身密钥的产生、存储、备份/恢复、归档和销毁。

② 提供密钥生成和分发服务。

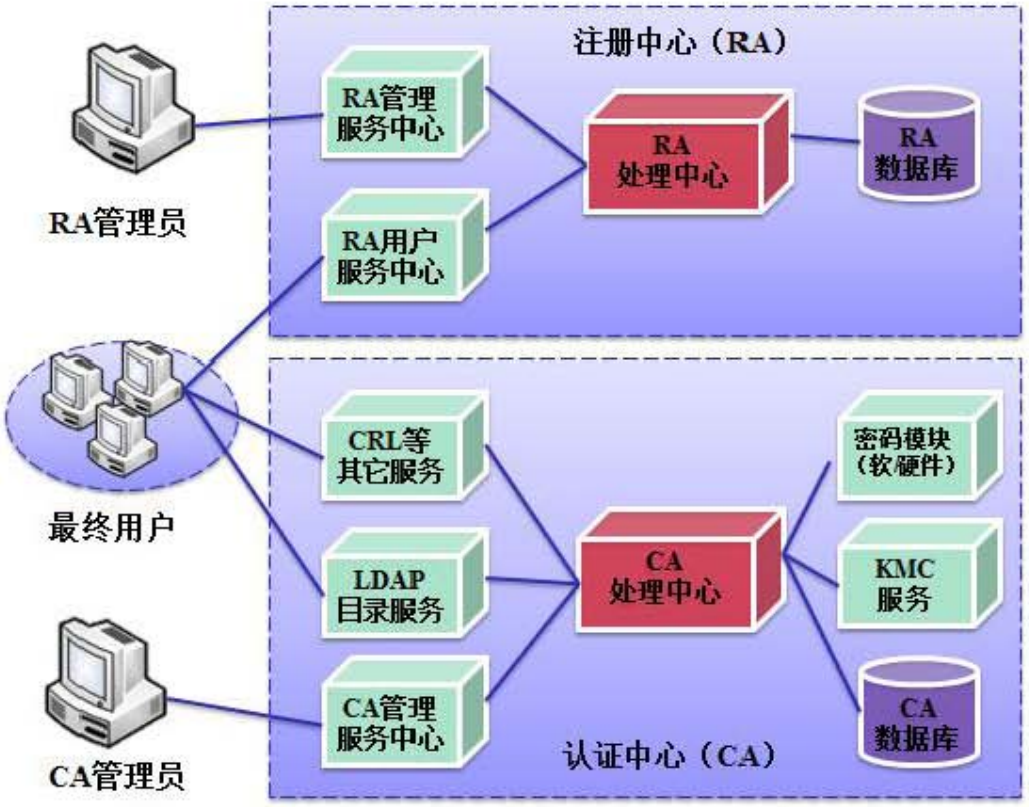


图 2 数字档案馆 CA 系统模块架构

③ 确定密钥生存周期，实施密钥撤销和更新管理。

④ 为安全加密通信提供安全密钥管理服务。

⑤ 提供密钥托管和密钥恢复服务。

⑥ 其他密钥生成和管理，密码运算功能。

在证书管理方面的作用：

用户能够方便地查找各种证书，包括已经撤销的证书；

① 能够根据用户请求或其他信息撤销用户的证书；

② 能够根据证书的有效期自动地撤销证书；

③ 能够完成证书数据库的备份工作；

④有效地保护证书和密钥服务器的安全。

3.2 CA 中心安全模块说明

3.2.1 SSL 模块

3.2.1.1 模块功能

数字档案安全网关中的SSL可信安全身份认证服务模块通过SSL技术、Proxy技术、应用转换技术和 IPTunnel 技术，在不改变网络结构和应用模式的情况下，实现对目前所有网络层以上各种静态或动态端口应用的完全支持。

基于数字证书的高强度身份认证；

高强度的传输数据加密。

3.2.1.2 模块组成

安全网关中 SSL VPN 模块，在用户和信息中心系统之间建立起高强度的安全加密连接，保证信息中心交易信息的可信和保密，并且提供用户身份机制和级别访问控制机制，整个模块分为客户端和服务器端两个部分，流程如图 5 所示。

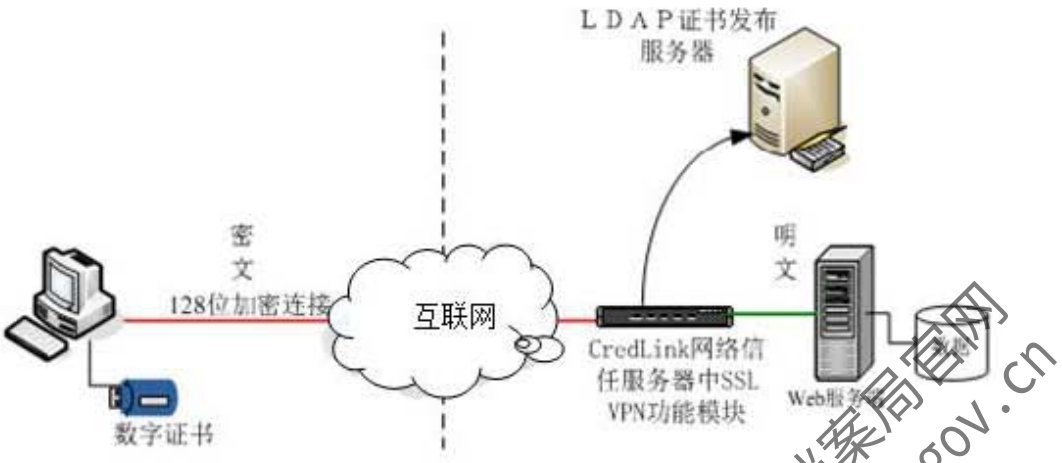


图 5 SSL 可信安全身份认证流程图

客户端部分：

数字证书：存储用户证书及私钥，用以证明客户端身份；

浏览器 SSL 代理模块：代理原客户端明文数据，与远端 SSL 安全网关建立高强度加密连接；

CSP 加密模块：提供 PKI 应用必须的加解密运算。

服务器端部分：

安全网关 SSL VPN 模块：要求客户端与之建立高强度加密连接，将数据解密后发给服务端；

LDAP 发布服务器：发布 CA 颁发的数字证书及相应黑名单。为认证服务器提供查询平台。

3.2.1.3 应用流程

应用流程图如图 6 所示：

3.2.2 SVS 模块

为了保障数据的不可抵赖性和完整性，防止用户对敏感数据的抵赖，使管理者在事后能够拿出有力的证据证明操作用户的真实身份以及操作痕迹，设计了可信签名验证系统。

3.2.2.1 模块功能

数字档案安全认证网关为用户提供完善的 PCS 数字签名运算服务，为服务器端提供 PCKS#1 格式数据签名运算以及数字信封的私钥加、解密运算；

而 SVS 签名验证服务模块主要用于服务器端接收数据后，对数据签名、签名证书的有效性进行合法性验证；安全网关中为应用系统实现数字签名、数字信封技术、验证用户身份、检验交易凭证等方面提供便捷的运算接口。主要功能如下：

签名数据验证；

验证证书有效性；

提供数据签名接口。

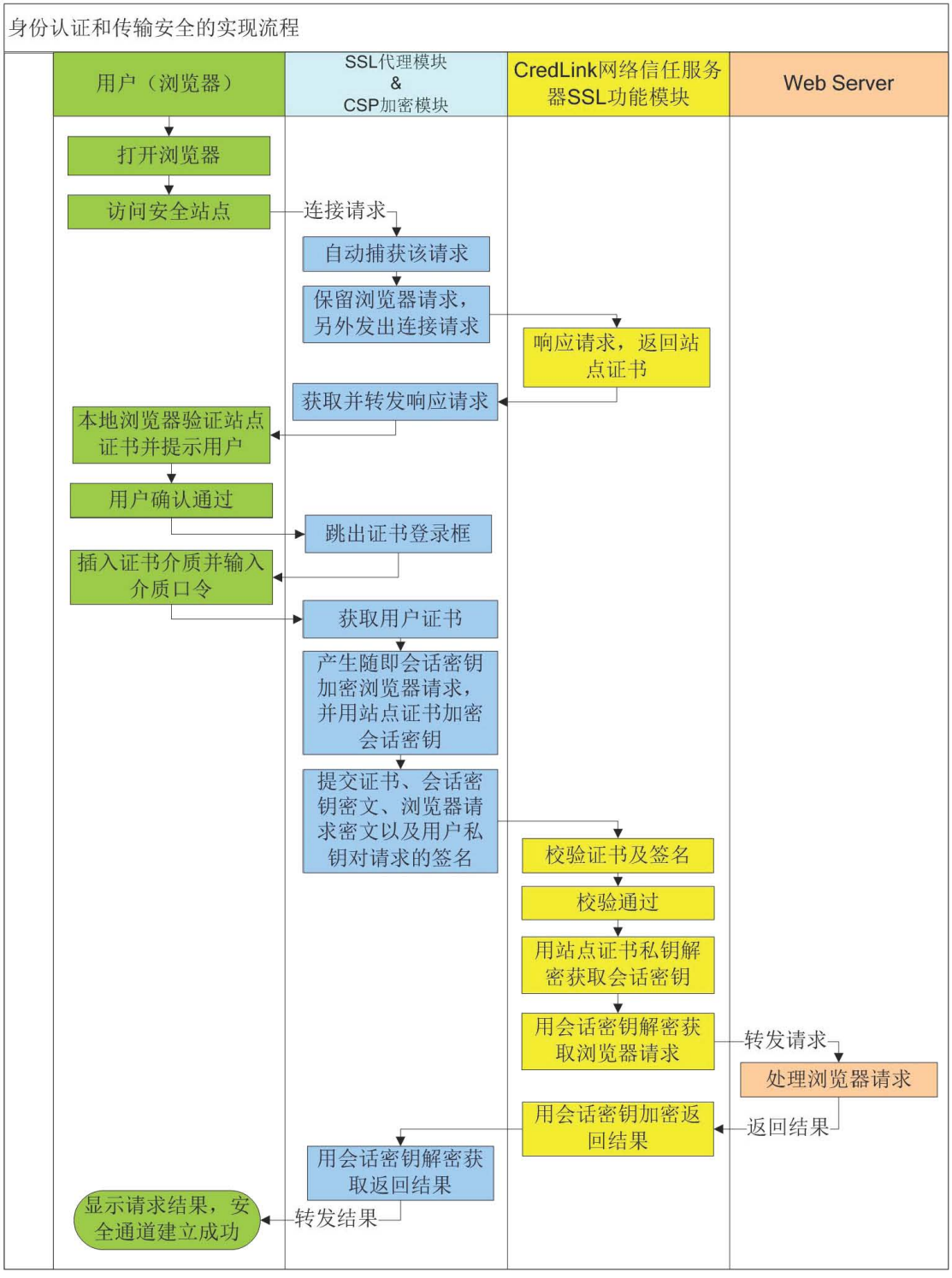


图 6 身份认证及安全传输实现流程

3.2.2.2 模块组成

可信签名验证功能模块分为两个部分：签名客户端，签名验证服务器端，功能逻辑图如图 7 所示。

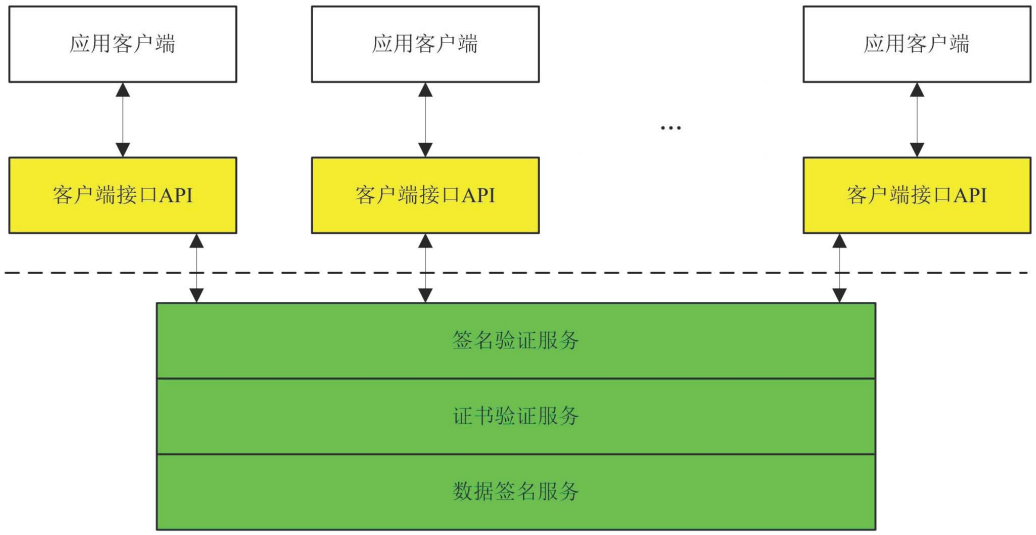


图 7　签名验证模块功能逻辑图

客户端部分：

电子签名客户端主要是 ACTIVEX 控件，可内置在用户浏览器中。通过设置对控件的调用参数，客户端可以对 WEB 页面中的任意指定的信息进行签名，并可单独对上传文件签名。

服务器端部分：

签名验证系统服务模块运行在安全认证网关之中，用于实现对客户端签名的验证和审计；验证设备为多种应用开发语言提供了统一的接口，方便用户调用。

3.3.2.3 应用流程

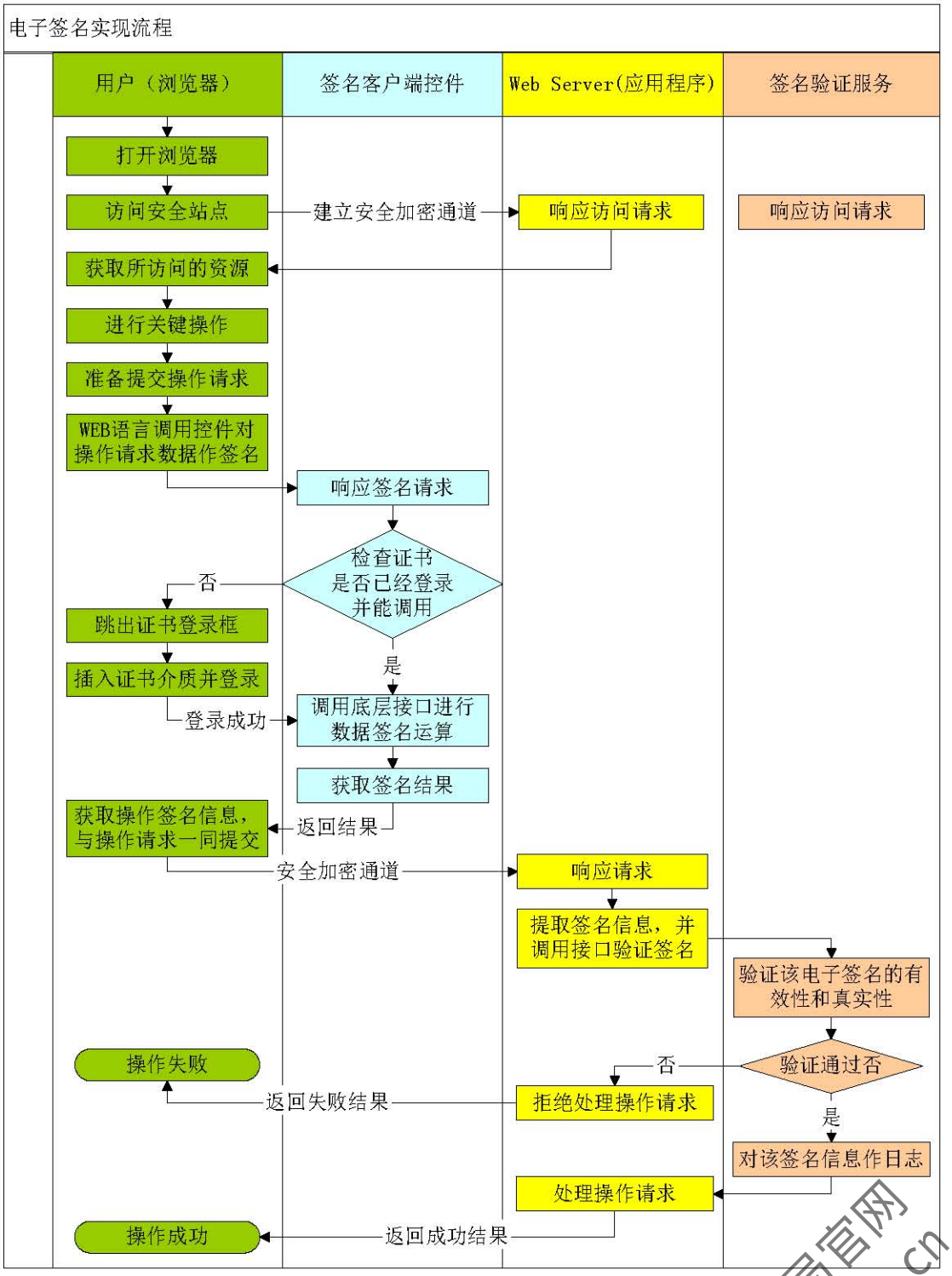


图 8 签名验证系统流程图

四、OFD 格式及基于 PDF/A 的可信电子档案格式

文件格式的标准化，主要通过两种方式来实现，一是在文件生成时采用开放的、标准的格式，而非专有格式；二是在电子文件归档时，将文件格式转化为标准格式。由于第一种方式需要改变人们在日常工作中的操作习惯，让其放弃一种自己已经熟悉的软件和文件格式，转而使用另外一种陌生的文件格式，阻力较大。而第二种方式则不同，因为不论在工作时使用何种文件格式，在电子文档需要长期保存时，都可转换为一种或几种统一的适用的文件格式，而电子文件的管理部门只需要对这些用于长期保存的格式进行注册、维护就可以了。相对于第一种方法，第二种做法更加符合人们的工作需要以及档案管理的要求。

电子文件的长期保存格式一直是数字档案馆在进行数字资源保存和利用时需要面对的极其重要的问题。由于生成电子文件的程序的多样性，造成了文件格式的多样性。程序不断更新，新的文件格式不断产生，旧的文件格式逐渐被淘汰，但阅读旧格式文件仍然需要旧程序的支持，因此，在对电子文件进行长期保存时，会面临众多问题。对于这些问题，档案界采取了仿真、数据迁移等方法，但这些方法都存在一定的缺陷，且成本高、技术和管理复杂。因此，文件格式的标准化、统一化就成为电子文件长期保存的一个重要途径。

对于转换之后形成的文件格式，必须是标准化的、符合长期保存要求的格式。国家档案局于 2009 年 12 月 16 日发布的档案行业标准《版式电子文件长期保存格式需求》(DA/T 47-2009)，即规定版式电子文件长期保存格式应满足的需求为：(1) 格式开放；(2) 不绑定软硬件；(3) 文件自包含；(4) 格式自描述；(5) 显示一致性；(6) 持续可解释；(7) 稳健；(8) 可转换；(9) 利于存储；(10) 支持技术认证机制；(11) 利于利用。

4.1OFD 格式[13] [14]

由工业和信息化部软件司牵头中国电子技术标准化研究院组织的版式编写组制定的版式文档国家标准 OFD(Open Fixed-layout Document)，OFD 是一种适

合交换和保存的文档格式。它定义了一套版式技术的标准集，包括图形、图像，文字、复合对象、视频、音频、标注、签名、水印、交互、大纲、元数据等描述。未来会扩充安全、流式信息、表单等应用。

其主要特性如下：

(1)低技术门槛。支持版式技术的基本集，体系简单．易于实现。

(2)用 XML 描述。容易理解和扩展。

(3)接触式注释。标注位于文档节点下。与页面正文分开．易提取和访问。

OFD 的性能分析

优点：

- 国家自有知识产权；
- 基于 XML 标准，易于理解和扩展；
- 支持标准的 XML 签名标准；
- 支持版式技术的基本集，体系简单，易于实现；
- 支持与流式兼容的底纹模型；
- 标注采用非接触式，易提取和访问；
- 资源集中存放。

缺点：

- 版式技术的应用描述较少，有待扩充；
- 全新标准，有待推广。

4.2 电子档案格式标准：PDF/A[15][16]

Adobe 公司在我国这个标准公布之前，在 2005 年已推出了 PDF/A 格式，并变成了国际标准（ISO 19005-1:2005），使其成为第一个适合长期保存的电子文件格式，这个标准是基于 PDF1.4 版本的，也被称为 PDF/A-1。2008 年 1 月，Adobe 进一步将 PDF 格式的所有权交付给国际标准化组织 ISO，使 PDF 完全成为一个公开的标准（ISO/DIS32000 — Document management — Portable documentformat—PDF 1.7）。2011 年 7 月，国际标准化组织公布了最新的

PDF/A-2（ISO 19005-2:2011）标准。

对于档案馆来说，目前数字档案增加速度迅猛，增量和存量并举。对于存量档案，通过扫描技术对纸质、图片、影像等传统档案实施数字化，保存格式通常为 TIFF 或 JPEG，相比较而言，PDF/A 功能更强、兼容性更高、体积更小，能够保存元数据并通过元数据对文档进行分类，并且能够对整个档案而不是单一文档进行全文检索，这些都有利于档案资源的长期有效利用。对于增量档案，各机关单位归档的电子档案格式各异，大多数公文采用 office 软件生成。Office 文档属于流式文档，PDF 属于版式文档，版式文档特点是版式固定(包括版式、版面、字体、字号等方面)、高保真而体积小、与平台无关(在任何操作系统、任何设备下，文档呈现一致)，这些对于海量电子档案的归档减少了很多障碍。

PDF/A 格式的主要特点有：它是中国国家和 ISO 国际电子文件长期保存的格式标准；版式不变性，可以保证在任何环境下版式不变；防病毒，它通常只有描述数据，没有可执行代码，可以防止病毒的侵袭；另外，有高保真、内容渲染、多平台支持、多媒体、交互性、安全性高、可使用数字签名等特点，同时，它还可以打包非结构化和结构化的数据。因此对于电子文件格式来说，目前 PDF/A 格式满足在档案馆中进行信息发布、分发共享和长期存储的可信格式要求。

4.3 PDF/A 与电子签名

随着全球电子签名立法的不断完善，电子签章与传统的手写签名和盖章具有同等的法律效力。2004 年中国颁布了《电子签名法》，电子签章有了较好的施行基础。电子签章以数字证书为基础，以数字签名为核心技术，将数字签名与印章图片以及被签章对象绑定在一起，为签章对象提供完整性验证和真实性验证。它是数字签名的可视化表现形式之一。是利用电子签章对被签章对象进行数字签名后，所形成的包括印章图片、数字签名值、数字证书等信息在内的不可分割的数据集。

PDF/A 因其版式固定、高保真、与平台无关等特性，成为在电子文档上进行数字签章的最优格式。目前基于 PDF/A 文档的电子签章应用，可实现客户端签章、

服务器端签章、USB-Key 签章等多种安全签章。人们可以非常方便地在电脑和移动设备上，在局域网内、在互联网上对 PDF/A 文档上进行数字签名与验证，并确保签章具唯一性、可信可靠。专业的 PDF 技术厂商提供的电子签章支持权威的第三方 CA 机构对其扩展，且签名值能被任意一个遵循 PDF 标准的阅读器打开验证。

五、以数字档案馆（室）为中心的可信电子档案体系建设

数字档案馆是运用现代信息技术对电子档案信息进行采集、加工、存储、管理，并通过各种网络平台提供公共档案信息服务和共享利用的档案信息集成管理系统，它既是实体概念，也是系统概念[17]。

随着国家电子文件管理体系的不断完善以及各层级档案馆际之间数据交互和资源共享的日益频繁，为了保证电子档案的可信，课题组研究建立一个以数字档案馆（室）为中心的可信档案体系。体系如图 9 所示。

档案系统按照行政隶属关系和业务指导关系，由中央到省到地方有非常完备的组织体系；从技术架构来说，各档案馆有各自的 CA 中心，分别由国家认可的数字认证机构支持，各认证中心必须互信。

依托政府基础信息网络，将各层级档案馆、各立档单位档案室有机结合起来，以国家标准来规范数据库、软硬件平台、安全控制等标准，建立大规模、集中式、虚拟的档案馆（室）集群。

这样，档案室与档案馆之间、档案馆与档案馆之间就可以形成一个立体的、网状的互信体系。为国家档案局提出的以数字资源为基础、安全管理为保障、远程利用为目标的数字档案馆（室）体系的发展目标打下良好的基础。

数字档案馆（室）是现代化的、升级版的档案馆（室）。建设可信的数字档案馆（室）体系，是电子档案在各个环节的可信性保证。

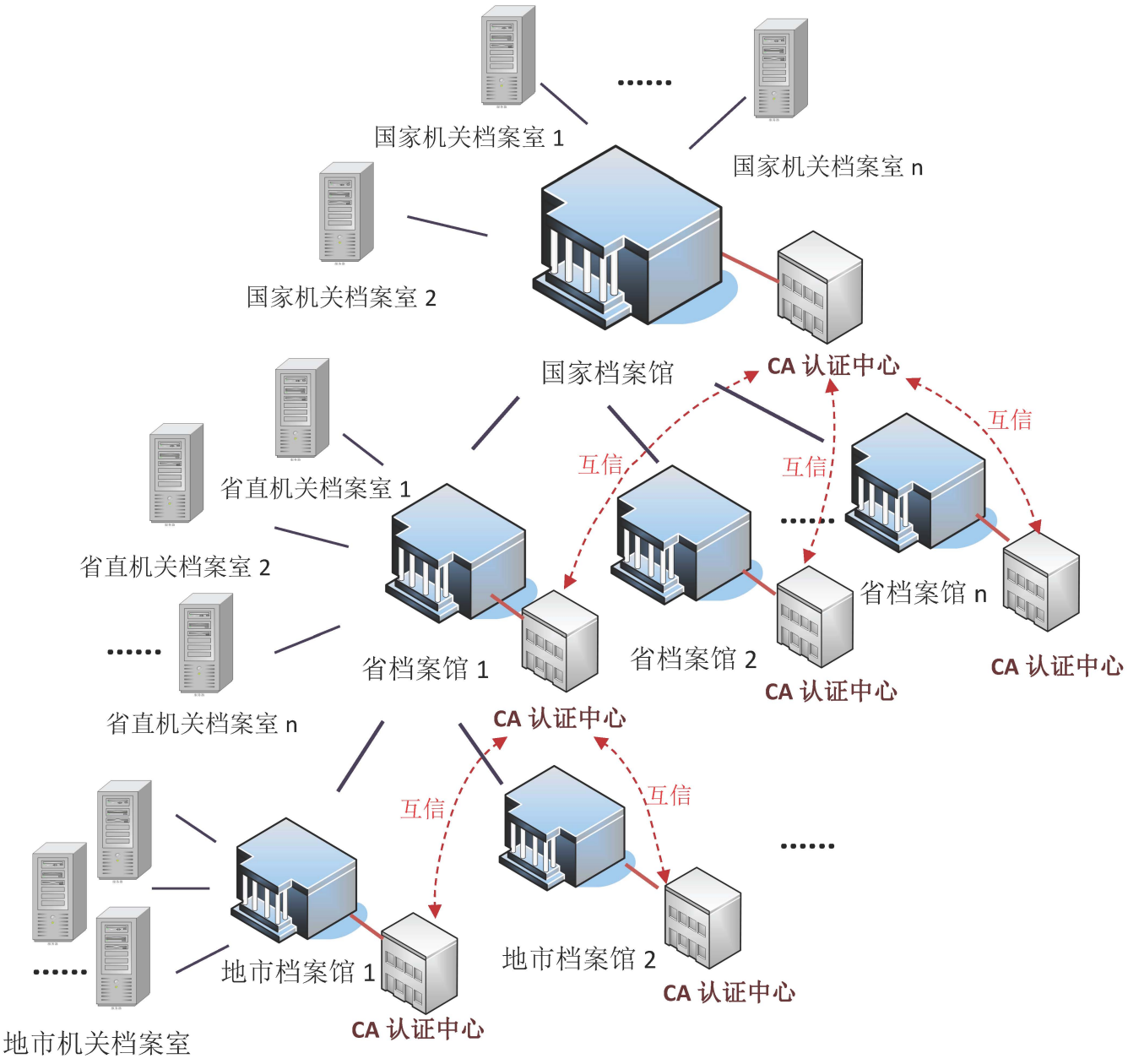


图 9 以数字档案馆（室）为中心的档案体系