



(12) 发明专利

(10) 授权公告号 CN 1936780 B

(45) 授权公告日 2012.01.11

(21) 申请号 200610129121.1

栏第 20 行, 第 7 栏第 31-58 行, 附图 4.

(22) 申请日 2006.09.08

US 6785815 B1, 2004.08.31, 全文.

US 5671389 A, 1997.09.23, 全文.

(30) 优先权数据

2005-263074 2005.09.09 JP

2006-232812 2006.08.29 JP

审查员 吴敏

(73) 专利权人 佳能株式会社

地址 日本东京

(72) 发明人 须贺祐治

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 李镇江

(51) Int. Cl.

G06F 3/00 (2006.01)

H04L 9/32 (2006.01)

G06F 17/30 (2006.01)

(56) 对比文件

US 5898779 A, 1999.04.27, 说明书第 3 栏第 1-18 行, 第 6 栏第 42-64 行, 第 6 栏第 65 行 - 第 7

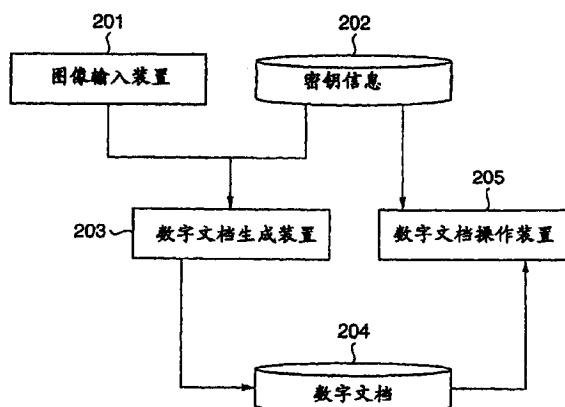
权利要求书 2 页 说明书 12 页 附图 14 页

(54) 发明名称

信息处理装置、验证处理装置及其控制方法

(57) 摘要

一种信息处理装置, 包括适于通过将数字文档划分成区域来生成要签名的数据的第一生成单元, 适于生成要签名的数据的第一提要值和用于标识要签名的数据的标识符的第二生成单元, 适于根据从数字文档获得的多个第一提要值和标识符生成签名信息的第三生成单元, 及适于根据签名信息和要签名的数据生成第一签名数字文档的第四生成单元。



1. 一种信息处理装置,包括:

第一生成单元,适于通过将数字文档划分成区域来生成要签名的数据;

第二生成单元,适于生成要签名的数据的第一提要值和用于标识要签名的数据的标识符;

第三生成单元,适于根据从所述数字文档获得的多个第一提要值和标识符生成签名信息;以及

第四生成单元,适于根据签名信息和要签名的数据生成第一签名数字文档,

其中所述第三生成单元利用从所述数字文档获得的所述多个第一提要值和标识符及加密密钥生成签名值,并利用所生成的签名值以及所述多个第一提要值和标识符生成签名信息。

2. 如权利要求1所述的信息处理装置,还包括:

选择接受单元,适于接受要签名的数据的选择,

其中,所述第三生成单元根据要签名的数据的多个第一提要值和标识符生成签名信息,要签名数据的选择是由所述选择接受单元接受的。

3. 如权利要求1所述的信息处理装置,还包括:

区域指定接受单元,适于接受对数字文档预定区域的指定,

其中所述第一生成单元为由所述区域指定接受单元指定的预定区域生成要签名的数据。

4. 一种验证处理装置,根据由如权利要求1至3任意一项所述的信息处理装置生成的第一签名数字文档验证数字文档,该验证处理装置包括:

提取单元,适于从第一签名数字文档提取签名信息;

确定单元,适于确定签名信息中的所述多个第一提要值和标识符是否已经改变;

获得单元,适于当所述确定单元确定所述多个第一提要值和标识符未改变时根据标识符从第一签名数字文档获得要签名的数据;

计算单元,适于计算要签名的数据的多个第二提要值;

比较单元,适于比较第一提要值和第二提要值;以及

验证结果生成单元,适于根据比较结果生成验证结果。

5. 如权利要求4所述的验证处理装置,其中所述确定单元根据通过利用解密密钥解密包括在签名信息中的签名值所获得的结果是否匹配于第一提要值和标识符,来确定签名信息中的第一提要值和标识符是否已经改变。

6. 如权利要求4所述的验证处理装置,其中即使当所述获得单元不能获得对应于多个标识符中一个的要签名的数据但能够获得对应于其它标识符的要签名的数据时,所述获得单元也获得要签名的数据。

7. 如权利要求4所述的验证处理装置,还包括:

操作单元,适于对第一签名数字文档施加操作;以及

第五生成单元,适于根据进行了该操作的第一签名数字文档生成第二签名数字文档,

其中,当所述操作单元通过选择包括在第一签名数字文档中的要签名的数据的任一个来重建数字文档时,所述第五生成单元对于该重建数字文档根据签名信息和由所述操作所选择的要签名的数据生成第二签名数字文档。

8. 一种用于控制信息处理装置的方法,包括:

第一生成步骤,通过将数字文档划分成区域来生成要签名的数据;

第二生成步骤,生成要签名的数据的第一提要值和用于标识要签名的数据的标识符;

第三生成步骤,根据从所述数字文档获得的多个第一提要值和标识符生成签名信息;

以及

第四生成步骤,根据签名信息和要签名的数据生成第一签名数字文档,

其中,在第三生成步骤中,签名值是利用从所述数字文档获得的所述多个第一提要值和标识符及加密密钥生成的,而且签名信息是利用所生成的签名值以及所述多个第一提要值和标识符生成的。

9. 如权利要求 8 所述的用于控制信息处理装置的方法,还包括:

选择接受步骤,接受要签名的数据的选择,

其中,在第三生成步骤中,根据要签名的数据的第一提要值和标识符生成签名信息,要签名的数据的选择是在所述选择接受步骤接受的。

10. 如权利要求 8 所述的用于控制信息处理装置的方法,还包括:

区域指定接受步骤,接受对数字文档预定区域的指定,

其中,在第一生成步骤中,为在区域指定接受步骤中指定的预定区域生成要签名的数据。

11. 一种用于控制验证处理装置的方法,其中验证处理装置根据由如权利要求 8 至 10 任意一项所述的方法生成的第一签名数字文档验证数字文档,该方法包括:

提取步骤,从第一签名数字文档提取签名信息;

确定步骤,确定签名信息中的多个第一提要值和标识符是否已经改变;

获得步骤,当在确定步骤中确定所述多个第一提要值和标识符未改变时根据标识符从第一签名数字文档获得要签名的数据;

计算步骤,计算要签名的数据的多个第二提要值;

比较步骤,比较第一提要值和第二提要值;以及

验证结果生成步骤,根据比较结果生成验证结果。

12. 如权利要求 11 所述的用于控制验证处理装置的方法,其中,在确定步骤中,对签名信息中第一提要值和标识符是否改变的确定基于通过利用解密密钥解密包括在签名信息中的签名值获得的结果是否匹配于第一提要值和标识符。

13. 如权利要求 11 所述的用于控制验证处理装置的方法,其中,在获得步骤中,即使当不能获得对应于多个标识符中一个的要签名的数据但能够获得对应于其它标识符的要签名的数据时,也获得要签名的数据。

14. 如权利要求 11 所述的用于控制验证处理装置的方法,还包括:

操作步骤,对第一签名数字文档施加操作;及

第五生成步骤,根据进行了该操作的第一签名数字文档生成第二签名数字文档,

其中,在第五生成步骤中,当数字文档在操作步骤中通过选择包括在第一签名数字文档中的要签名的数据的任一个重建时,对于重建数字文档根据签名信息和所述操作所选择的要签名的数据生成第二签名数字文档。

信息处理装置、验证处理装置及其控制方法

技术领域

[0001] 本发明涉及信息处理装置、验证处理装置及其控制方法。

背景技术

[0002] 近年来,随着计算机及其网络的快速发展和流行,如文本数据、图像数据、音频数据等的许多种信息都被数字化了。数字数据不会由于老化等而损坏,而且可以永远以理想的状态保存。此外,数字数据可以容易地拷贝、编辑和修改。

[0003] 数字数据的这种拷贝、编辑和修改对用户是非常有用的,但同时数字数据的保护也造成了严重的问题。特别是,当文档和图像数据要通过如因特网等的广域网分发时,由于数字数据很容易被改变,因此第三方也可以改变数据。

[0004] 为了使接收者能够检测进入的数据是否被改变,提出了称为数字签名的处理技术,作为验证防止改变的附加数据的策略。数字签名处理技术不仅可以防止数据的改变,而且还可以防止因特网上的电子欺骗、否认等。

[0005] 以下将具体描述数字签名、散列函数、公钥密码系统和公钥基础结构(PKI)。

[0006] (数字签名)

[0007] 图 14A 和 14B 是用于解释签名生成处理和签名验证处理的视图,而且以下将参考图 14A 和 14B 描述这些处理。当生成数字签名数据时,使用散列函数和公钥密码系统。

[0008] 令 $K_s(2106)$ 为私钥,而 $K_p(2111)$ 为公钥。发送者对数据 $M(2101)$ 施加散列处理 2102,以便计算作为固定长度数据的提要值 $H(M) 2103$ 。接下来,发送者利用私钥 $K_s(2106)$ 对固定长度的数据 $H(M)$ 施加签名处理 2104,以便生成数字签名数据 $S(2105)$ 。发送者将这种数字签名数据 $S(2105)$ 和数据 $M(2101)$ 发送到接收者。

[0009] 接收者利用公钥 $K_p(2111)$ 转换(解密)接收到的数字签名数据 $S(2110)$ 。接收者通过对接收到的数据 $M(2107)$ 施加散列处理 2108 生成固定长度的提要值 $H(M) 2109$ 。验证处理 2112 验证解密的数据是否与提要值 $H(M)$ 匹配。如果作为这种验证的结果,两个数据不匹配,则可以检测出数据已经改变。

[0010] 在数字签名中,可以使用如(后述)RSA、DSA 等的公钥密码系统。这些数字签名的安全性基于除私钥持有者以外的其它实体很难根据计算伪造签名或解码私钥的事实。

[0011] (散列函数)

[0012] 以下将描述散列函数。通过对要签名的数据施加有损压缩,散列函数与数字签名处理一起用于缩短分配签名的处理时间周期。即,散列函数具有处理具有任意长度的数据 M 并生成具有恒定长度的输出数据 $H(M)$ 的功能。应当指出,输出 $H(M)$ 称为明文数据 M 的散列数据。

[0013] 特别地,单向散列函数的特征在于如果给出数据 M ,则计算出满足 $H(M') = H(M)$ 的明文数据 M' 在计算量方面是很困难的。作为单向散列函数,如 MD2、MD5、SHA-1 等的标准算法都可以使用。

[0014] (公钥密码系统)

[0015] 以下将描述公钥密码系统。公钥密码系统使用两种不同的密钥,而且特征在于通过一个密钥加密的数据只可以通过另一个密钥解码。在这两个密钥中,一种密钥称为公钥,是向公众公开的。另一种密钥称为私钥,并且是由所标识的人拥有的。

[0016] 利用公钥密码系统、RSA 签名、DSA 签名、Schnorr 签名等的数字签名是已知的。在这种情况下,将作为例子描述 1978 年 2 月 Communications of the ACM, v. 21, n. 2 上第 120-126 页由 R. L. Rivest、A. Shamir 与 L. Aldeman 所写“A method for Obtaining Digital Signatures and Public-Key Cryptosystems”中描述的 RSA 签名。而且,将附加地解释 2000 年 1 月 Federal Information Processing Standards (FIPS) 186-2, Digital Signature Standard (DSS) 中描述的 DSA 签名。

[0017] (RSA 签名)

[0018] 生成质数 p 和 q ,使得有 $n = pq$ 。 $\lambda(n)$ 设置为 $p-1$ 和 $q-1$ 的最小公倍数。选择对 $\lambda(n)$ 合适的 e 质数,使得私钥 $d = 1/e \pmod{\lambda(n)}$,其中 e 和 n 是公钥。而且,令 $H()$ 是散列函数。

[0019] (RSA 签名生成)用于文档 M 的签名生成序列

[0020] 令 $s := H(M)^d \pmod{n}$ 为签名数据。

[0021] (RSA 签名验证)用于文档 M 的签名 (s, T) 的验证序列

[0022] 验证是否 $H(M) = s^e \pmod{n}$ 。

[0023] (DSA 签名)

[0024] 令 p 和 q 为质数,且 $p-1$ 是整除 q 的值。令 g 为次数为 q 的元素(生成元), q 任意选自 Z_p^* (从次数为 p 的循环群 Z_p 排除零的乘法群)。令任意选自 Z_p^* 的 x 为私钥,以便通过 $y := g^x \pmod{p}$ 给出公钥 y 。令 $H()$ 为散列函数。

[0025] (DSA 签名生成)用于文档 M 的签名生成序列

[0026] 1) a 任意选自 Z_p ,使得 $T := (g^a \pmod{p}) \pmod{q}$ 。

[0027] 2) 使得 $c := H(M)$ 。

[0028] 3) 使得 $s := a^{-1}(c + xT) \pmod{q}$,以便设置 (s, T) 为签名数据。

[0029] (DSA 签名验证)用于文档 M 的签名 (s, T) 的验证序列

[0030] 验证是否 $T = (g^{(h(M)s^{-1})}y^{(Ts^{-1})} \pmod{p}) \pmod{q}$ 。

[0031] (公钥基础结构)

[0032] 为了在客户-服务器通信中访问服务器中的资源,需要用户认证。作为用户认证的一种方式,普遍使用如 ITU-U 建议 X. 509 等的公钥证书。公钥证书是保证公钥及其用户之间绑定的数据,并且是由称为认证机构:CA 的受托第三方数字签名的。用在浏览器中的利用 SSL(安全套接字层)的用户认证方案是通过确认用户是否具有对应于包括在用户所给出的公钥证书中的公钥的私钥来实现的。

[0033] 由于公钥证书是由 CA 签名的,因此包括在其中的用户或服务器的公钥是可以信任的。为此,当用在 CA 签名生成中的私钥泄漏或者变得脆弱时,由该 CA 颁发的所有公钥证书都变得无效。由于某些 CA 管理着大量的公钥证书,因此为降低管理成本作出了多种建议。随后描述的本发明的作用是可以减少要颁发的证书数量和作为公钥仓库的服务器访问。

[0034] 在 ITU-U 建议 X. 509/ISO/IEC 9594-8 中描述的 ITU-U 建议 X. 509v. 3 中:

“Information technology-Open Systems Interconnection-The Directory:Public-key and attribute certificate frameworks”,包括作为要签名的数据的数据的要验证的实体（主体）的 ID 和公钥信息。通过用于通过对这些要签名的数据应用散列函数所获得的提要的如上述 RSA 算法等的签名操作,生成签名数据。要签名的数据具有可选字段“扩展”,它可以包括对应用或协议唯一的扩展数据。

[0035] 图 15 示出了由 X.509 v.3 指定的格式,以下将解释每个单独字段中所示的信息。“版本”字段 1501 存储 X.509 的版本。这个字段是可选的,如果忽略则表示 v1。“序列号”字段 1502 存储由 CA 唯一指定的序列号。“签名”字段 1503 存储公钥证书的签名方案。“颁发者”字段 1504 存储作为公钥证书颁发者的 CA 的 X.500 标识名。“有效性”字段 1505 存储公钥的有效期（起始日期和结束日期）。

[0036] “主体”字段 1506 存储对应于包括在这种证书中的公钥的私钥持有者的 X.500 标识名。“主体公钥信息”字段 1507 存储颁发的公钥。“颁发者唯一标识符”字段 1508 和“主体唯一标识符”字段 1509 是从 v2 开始添加的可选字段,分别存储对 CA 和持有者的唯一标识符。

[0037] “扩展”字段 1510 是在 v3 中添加的可选字段,存储三个值的集合,即,扩展类型 (extnId) 1511、临界位 (critical) 1512 和扩展值 (extnValue) 1513。v3 “扩展”字段不仅可以存储由 X.509 指定的标准扩展类型,而且还可以存储唯一的、新的扩展类型。为此,如何识别 v3 “扩展”字段依赖于应用方。关键位 1512 指示扩展类型是不可缺少的还是可以忽略的。

[0038] 已经对数字签名、散列函数、公钥密码系统和公钥基础结构进行了描述。

[0039] 已经提出了用于将要签名的文本数据划分成多个文本数据并利用上述数字签名处理技术将数字签名附加到各文本数据的方案（见日本专利公开号 10-003257）。根据这种提出的方案,当数字签名的文本数据被部分引用时,可以对该部分引用的文本进行验证处理。

[0040] 所提出的方案只处理作为要签名的数据的文本数据。但是,随着近年来数字数据的多样化,包括多种类型内容的复合内容也可以被数字签名。当这种复合内容被作为二进制数据群处理并要通过例如压缩处理等被数字签名时,如果第三方将内容划分成子内容并尝试重新分发这些子内容,则子内容中的签名数据不再能被验证。

[0041] 就象在所提出的方案中,为了避免这种问题,除了文本数据,所有要签名的子内容都可以被数字签名。但是,在这种情况下,签名生成和签名验证在其加密或解码处理中都需要大量的计算成本。因此,处理的数量随着子内容数量的增加而增加。

发明内容

[0042] 因此,本发明的目的是即使当作为这种复合内容一部分的子内容单独存在时,也允许不仅对文本数据而且对以多种格式存储的数字数据的复合内容进行签名验证。而且,本发明的目的是提供可以将签名生成和签名验证处理的计算量设置为恒定而不是与所划分子内容的量成比例的签名处理技术。

[0043] 根据一起或者独立地至少减轻上述问题的本发明,提供了信息处理装置,包括:适于通过将数字文档划分成区域来生成要签名的数据的第一生成单元,适于生成要签名的数

据的第一提要值和用于标识要签名的数据的标识符的第二生成单元,适于根据从数字文档获得的多个第一提要值和标识符生成签名信息的第三生成单元,及适于根据签名信息和要签名的数据生成第一签名数字文档的第四生成单元。

[0044] 而且,还提供了根据签名数字文档验证数字文档的验证处理装置,该装置包括:适于从签名数字文档提取签名信息的提取单元,适于确定签名信息中的第一提要值和标识符是否改变的确定单元,适于当确定单元确定第一提要值和标识符未改变时根据标识符从签名数字文档获得要签名的数据的获得单元,适于计算要签名的数据的第二提要值的计算单元,适于比较第一提要值和第二提要值的比较单元,及适于根据比较结果生成验证结果的验证结果生成单元。

[0045] 此外,还提供了用于控制信息处理装置的方法,包括:通过将数字文档划分成区域来生成要签名的数据的第一生成步骤,生成要签名的数据的第一提要值和用于标识要签名的数据的标识符的第二生成步骤,根据从数字文档获得的多个第一提要值和标识符生成签名信息的第三生成步骤,及根据签名信息和要签名的数据生成第一签名数字文档的第四生成步骤。

[0046] 此外,还提供了用于控制根据签名数字文档验证数字文档的验证处理装置的方法,该方法包括:从第一签名数字文档提取签名信息的提取步骤,确定签名信息中的第一提要值和标识符是否改变的确定步骤,当在确定步骤中确定第一提要值和标识符未改变时根据标识符从签名数字文档获得要签名的数据的获得步骤,计算要签名的数据的第二提要值的计算步骤,比较第一提要值和第二提要值的比较步骤,及根据比较结果生成验证结果的验证结果生成步骤。

[0047] 本发明的更多特征将参考附图从以下示例实施方式的描述变得显而易见。

附图说明

[0048] 图 1 是显示对应于本发明实施方式的系统布置例子的图;

[0049] 图 2 是显示对应于本发明实施方式的系统的功能性布置例子的方框图;

[0050] 图 3 是显示对应于本发明实施方式的系统的硬件布置例子的方框图;

[0051] 图 4 是对应于本发明实施方式的数字文档生成处理与数字文档操作处理的功能性方框图;

[0052] 图 5 是显示对应于本发明实施方式的中间数字文档生成处理中处理例子的流程图;

[0053] 图 6A 和 6B 是用于解释对应于本发明实施方式的数字数据例子的视图;

[0054] 图 7A 和 7B 是用于解释对应于本发明实施方式的中间数字文档和数字文档的视图;

[0055] 图 8 是显示对应于本发明实施方式的签名生成处理中处理例子的流程图;

[0056] 图 9A 和 9B 是显示对应于本发明实施方式的数字文档结构例子的视图;

[0057] 图 10 是显示对应于本发明实施方式的签名验证处理中处理例子的流程图;

[0058] 图 11A 和 11B 是显示对应于本发明实施方式在重建处理之后签名数据结构例子的视图;

[0059] 图 12 是用于解释对应于本发明第三实施方式的数字数据浏览例子的视图;

- [0060] 图 13 是用于解释对应于本发明第三实施方式的数字数据另一浏览例子的视图；
[0061] 图 14A 是显示签名生成处理通用例子的图；
[0062] 图 14B 是显示签名验证处理通用例子的图；
[0063] 图 15 是用于解释公钥证书 X.509v.3 数据格式的视图。

具体实施方式

[0064] 现在将根据附图具体描述本发明的优选实施方式。

[0065] < 第一实施方式 >

[0066] 对应于这种实施方式的签名生成处理和签名验证处理包括数字文档生成处理和数字文档操作处理。更具体而言,数字文档生成处理将通过扫描纸质文档生成的图像数据划分成子内容并通过由用户数字签名期望的子内容群生成复合内容(下文中称为数字文档)。数字文档操作处理从数字文档提取子内容,验证需要验证的子内容的签名信息,然后执行如浏览、打印等的内容消费处理、内容重建处理等。

[0067] 图 1 是显示对应于这种实施方式的系统布置例子的图。图 1 所示的系统通过连接扫描仪 101、作为用于生成和验证数字文档的处理装置的计算机 102、用于编辑和修改数字文档的计算机 103 及用于通过网络 105 打印数字文档的打印机 104 来配置。

[0068] 图 2 是显示对应于这种实施方式的系统的功能性布置例子的功能性方框图。参考图 2,图像输入装置 201 接收图像数据。密钥信息 202 包括用于生成数字签名的加密密钥和用于验证数字签名的解密密钥。作为信息处理装置的数字文档生成装置 203 通过根据输入的图像数据和密钥信息 202 的加密密钥将签名信息附加到输入的图像数据来生成数字文档 204。数字文档操作装置 205(验证处理装置)利用密钥信息 202 的解码密钥验证所生成的数字文档 204,并对数字文档执行例如数据修改、编辑、打印等操作。数字签名处理将根据公钥密码系统解释。在这个时候,密钥信息 202 的加密密钥对应于私钥 406,而密钥信息 202 的解密密钥对应于公钥 414。

[0069] 图 3 是显示数字文档生成装置 203 和数字文档操作装置 205 的内部硬件布置例子的方框图。CPU 301 通过执行软件整体控制装置。存储器 302 临时存储由 CPU 301 执行的软件和数据。硬盘 303 存储软件和数据。输入/输出(I/O)单元 304 从键盘、鼠标、扫描仪等接收输入信息,并向显示器和打印机输出信息。

[0070] (数字文档生成处理)

[0071] 以下将描述对应于这种实施方式的处理。图 4 是显示对应于这种实施方式的处理例子的功能性方框图。如图 4 所示,对应于这种实施方式的处理大致包括数字文档生成处理 401 和数字文档操作处理 402。

[0072] 在对应于这种实施方式的数字文档生成处理 401 中,纸质文档输入处理 404 输入纸质文档 403。接下来,中间数字文档生成处理 405 通过分析纸质文档 403 生成中间数字文档。签名信息生成处理 407 根据中间数字文档和私钥 406 生成签名信息。签名信息附加处理 408 关联中间数字文档和签名信息。数字文档归档处理 409 通过综合中间数字文档和签名信息生成数字文档 411。数字文档 411 对应于图 2 中的数字文档 205。数字文档发送处理 410 将数字文档 411 发送到数字文档操作处理 402。

[0073] 在数字文档操作处理 402 中,数字文档接收处理 412 接收数字文档 411。数字文

档提取处理 413 从所接收的数字文档 411 提取中间数字文档和签名信息。签名信息验证处理 415 根据中间数字文档、签名信息和公钥 414 执行验证。文档操作处理 416 对提取出的数字文档执行例如修改、编辑、打印等操作。

[0074] 图 4 中功能块的细节将进一步描述。以下将参考图 5 及图 6A 和 6B 描述中间数字文档生成处理 405 的细节。图 5 是显示对应于这种实施方式的中间数字文档生成处理 405 中处理例子的流程图。图 6A 和 6B 示出了数字数据和区域划分处理结果的例子。

[0075] 参考图 5, 在步骤 S501, 由纸质文档输入处理 404 获得的数据被数字化, 以便生成数字数据。图 6A 示出了所生成的数字数据的例子。

[0076] 在步骤 S502, 数字数据被划分成关于各个属性的区域。本例中的属性包括文本、照片、表格和图片。

[0077] 区域划分处理提取例如数字数据中 8 个相连的黑色轮廓像素组、4 个相连的白色轮廓像素组等的集合, 而且可以提取具有例如文本、图片或图形、表格、框架和线的特征名的区域。这种方案在美国专利号 5, 680, 478 中描述。应当指出, 区域划分处理的实现方法不限于这种特定的处理, 而是其它方法也可以应用。

[0078] 图 6B 示出了通过根据所提取的特征量确定属性的区域划分结果的例子。应当指出, 作为各个区域的属性, 602、604、605 和 606 指示文本区域, 而 603 指示彩色照片区域。

[0079] 在步骤 S503, 为在步骤 S502 中获得的每个区域生成文档信息。每个文档信息都包括属性, 如页面上的位置坐标等的布局信息, 如果感兴趣的划分区域属性为文本时的字符代码串, 如段落、标题等的文档逻辑结构, 等等。

[0080] 在步骤 S504, 在步骤 S502 中获得的每个区域都被转换成传输信息。传输信息是呈现所需的。更具体而言, 传输信息包括分辨率可变的光栅图像、向量图像、单色图像或彩色图像、每个传输信息的文件大小、如果感兴趣划分区域的属性是文本时作为字符识别结果的文本、单个字符的位置与字体、通过字符识别获得的字符的可靠性, 等等。以图 6B 作为例子, 文本区域 602、604、605 和 606 被转换成向量图像, 而彩色照片区域 603 被转换成彩色光栅图像。

[0081] 在步骤 S505, 在步骤 S502 中划分的区域、在步骤 S503 中生成的文档信息及在步骤 S504 中获得的传输信息彼此关联。各条相关信息在树结构中描述。在以上步骤中生成的传输信息和文档信息在下文中称为构成要素。

[0082] 在步骤 S506, 在以上步骤中生成的构成要素作为中间数字文档保存。保存格式不特别限制, 只要它能表达树结构就可以。在这种实施方式中, 中间数字文档可以利用作为结构化文档例子的 XML 来保存。

[0083] 以下将描述图 4 中的签名信息生成处理 407。在这种处理中, 为前面生成的中间数字文档的构成要素生成数字签名。图 8 是这种实施方式中签名信息生成处理的流程图。以下将参考图 8 描述签名信息生成处理 407。

[0084] 在步骤 S801, 为每个要签名的数据生成要签名的数据的提要值。应当指出, 要签名的数据是包括在中间数字文档中的数据, 而且可以看作是 (后述) 图 7A 中的传输信息 a(701)、传输信息 b(702) 或文档信息 (703)。为了生成提要值, 这种实施方式应用散列函数。由于散列函数已经在“背景技术”的段落中进行了描述, 因此其具体描述将省略。

[0085] 在步骤 S802, 为每个要签名的数据生成要签名的数据的标识符。应当指出, 标识符

只需要唯一标识要签名的数据。例如,在这种实施方式中,由 RFC2396 指定的 URI 作为要签名的数据的标识符应用。但是,本发明不限于这种指定,而是各种其它值也可以作为标识符应用。

[0086] 在步骤 S803 中检查是否已经对所有要签名的数据都施加了步骤 S801 和 S802 的处理。如果这种处理已经施加到所有要签名的数据(步骤 S803 中的“是”),则流程前进到步骤 S804;否则,流程返回步骤 S801。

[0087] 在步骤 S804,对在步骤 S801 中为相同数字文档生成的所有提要值和在步骤 S802 中生成的所有标识符利用私钥 406 执行签名值生成处理,以便计算签名值。为了生成签名值,这种实施方式应用“背景技术”段落中所描述的数字签名。数字签名的实际运算处理的具体描述将省略。图 14A 中所示签名生成处理流程中的数据 M(2101) 对应于在步骤 S801 中生成的所有提要值和在步骤 S802 中生成的所有标识符(这个数据群将称为聚合数据)。同样,私钥 Ks 2106 对应于图 4 中的私钥 406。

[0088] 随后,在步骤 S805,签名信息利用聚合数据(在步骤 S801 中生成的所有提要值和和在步骤 S802 中生成的所有标识符)和在步骤 S804 中生成的签名值配置,从而结束签名信息生成处理。

[0089] 应当指出,步骤 S804 中的签名值生成处理可以对一些生成的提要值和标识符(即,多个生成的提要值和标识符)而不是生成的所有提要值和所有标识符执行。在这种情况下,原始文档中更有可能被重新使用的子内容可以被自动或由用户手动选择,而且签名值可以基于与所选子内容关联的提要值和标识符计算。在这种情况下,在步骤 S805,签名信息利用用于计算签名值的一些提要值和标识符及计算出的签名值配置。即使当签名值利用多个(但不是所有)提要值和标识符计算时,签名值生成处理也可以对整个内容只进行一次。

[0090] 以下将参考图 9A 和 9B 描述对应于这种实施方式的数字文档 411 的结构。图 9A 和 9B 示出了对应于这种实施方式的数字文档 411 结构例子。图 9A 示出了整个数字文档 411 的结构。如图 9A 所示,数字文档 411 优选地包括签名信息 901、要签名的数据 1(902) 和要签名的数据 2(903)。图 9B 示出了图 9A 中签名信息 901 的具体结构的例子。如图 9B 所示,签名信息 901 优选地包括签名值 904、要签名的数据 1 的标识符(905)、要签名的数据 1 的提要值(906)、要签名的数据 2 的标识符(907) 及要签名的数据 2 的提要值(908)。数据 905 至 908 形成聚合数据 909。

[0091] 图 9A 示出了为两个要签名的数据 1(901) 和要签名的数据 2(903) 生成一个签名信息 901 时数字文档 411 的结构例子。图 9B 示出了签名信息 901 的具体结构的例子。在图 9B 中,要签名的数据 1 的标识符(905) 和要签名的数据 2 的标识符(907) 在上述步骤 S802 中生成。而且,要签名的数据 1 的提要值(906) 和要签名的数据 2 的提要值(908) 在上述步骤 S801 中生成。签名值 904 在步骤 S804 中利用数据 905 至 908,即聚合数据 909,生成。

[0092] 随后,以下将参考图 7A 描述签名数据附加处理 408。标号 701 和 702 指示在中间数字文档生成处理 405 中生成的中间数字文档的两条传输信息;703 指示文档信息。标号 704 和 705 指示在签名信息生成处理 407 中生成的两条签名信息。

[0093] 每个签名信息都嵌入了标识符,如上所述,标识符指示对应于要签名的数据的传

输信息或文档信息。在图 7A 中,指示要签名的数据的标识符 706(即,传输信息 701)嵌入到签名信息 704 中。签名信息和要签名的数据不需要总是具有一对一的对应性。例如,分别指示作为要签名的数据的传输信息 702 和文档信息 703 的标识符 707 和 708 可以嵌入到签名信息 705 中。

[0094] 应当指出,传输信息 a(701)被看作要签名的数据 1(902),而传输信息 b(702)和文档信息 703 被看作要签名的数据 2(903)。而且,签名信息 1(704)和签名信息 2(705)可以看作签名信息 901。

[0095] 以下将参考图 7A 和 7B 描述数字文档归档处理 409。如图 7A 所示,在上述处理中生成的中间数字文档和签名信息作为独立的数据存在。因此,数字文档归档处理将这些数据归档,以便生成一个数字文档。图 7B 示出了中间数字文档和签名信息的归档数据的例子。归档数据 709 对应于图 4 所示的数字文档 411。对于图 7A 所示的 701 至 705,701 对应于 713;702 对应于 714;703 对应于 712;704 对应于 710;而 705 对应于 711。

[0096] 已经解释了这种实施方式中的数字文档生成处理。如上所述,在根据这种实施方式的数字文档生成处理中,在假定原始内容是独立的且随后要重新分发或重新使用及标识符被赋予子内容的每一个或子内容群的情况下,原始内容分成了多个子内容。如在步骤 S802 的描述中所解释的,由 RFC2396 指定的 URI 可以应用作为标识符。但是,本发明不限于此,例如,可以使用原始内容中子内容的相对位置信息。而且,利用单向散列函数从例如唯一分配给子内容头字段的编号信息的元数据、从例如包括在头字段中的内容所有者、日期等的信息等计算出的值等等可以用作标识符。

[0097] 此外,提要值通过利用以对应于每个标识符的子内容作为输入的单向散列函数计算生成。标识符和提要值的集合(聚合数据)被给予复合内容。以这种方式,即使当在文档操作处理中一些子内容从原始内容删除,且利用剩余子内容重建的内容被分发时,也可以进行重建内容的签名验证处理。此外,即使当签名不是为每个子内容块生成时(即,即使当签名不是以与子内容的一对一对应关系生成时),也可以验证每个子内容块是否被改变。

[0098] 以下将结合数字文档操作处理描述重建内容验证的可能性。

[0099] (数字文档操作处理)

[0100] 在图 4 数字文档接收处理 412 中接收的数字文档 411 在数字文档提取处理 413 中经受与数字文档归档处理 409 相反的处理。即,中间数字文档和签名信息的单个数据从数字文档 411 中提取出来。

[0101] 在签名信息验证处理 415 中,图 14B 所示签名验证处理流程中的输入数据:M(2107)对应于聚合数据 909。同样,数字签名数据:S(2110)对应于签名信息 901,而公钥 2111 对应于图 4 中的公钥 414。以这种方式,可以检查聚合数据 909 是否被改变。

[0102] 如果可以确认聚合数据未改变,则验证对应于包括在聚合数据中的标识符的提要值是否与从要签名的数据生成的提要值匹配。以下将参考图 9 和图 10 描述上述处理。图 10 示出了根据本实施方式的签名验证处理的流程图的例子。

[0103] 参考图 10,在步骤 S1001,通过数字文档提取处理 413,签名信息 901 从数字文档 411 中提取出来。然后,根据包括在签名信息 901 中的签名值 904,利用图 14B 中所描述的方法验证聚合数据 909 是否改变。即,生成提要值 2109,以便使标识符 905 和提要值 906 及标识符 907 和提要值 908 作为输入数据 M。此外,签名值 904 利用公钥 414 解密,以便生成提

要值。然后,检查所生成的两个提要值是否匹配。如果这两个值匹配,则确定聚合数据 909 未改变。

[0104] 如果在步骤 S1002 中验证失败(步骤 S1002 中的“NG”),则签名验证处理结束,“NG”作为结果返回。另一方面,如果在步骤 S1002 中验证成功(步骤 S1002 中的“OK”),则分别对包括在聚合数据 909 中的标识符 905 和 907 执行步骤 S1003 至 S1008 中的处理。

[0105] 在步骤 S1004,要签名的数据 902 或 903 基于标识符 905 或 907 从数字文档 411 中提取出来。在步骤 S1005 检查是否可以获得要签名的数据 902 或 903。如果可以获得要签名的数据 902 或 903,则流程前进到步骤 S1006。如果不能获得要签名的数据 902 或 903,则流程跳到步骤 S1008。如果下一标识符存在,则对要签名的对应数据执行步骤 S1004 中的处理。如果存在不能从数字文档 411 获得的要签名的数据,则指示对应于感兴趣标识符的子内容不作为要验证的数据而被包括的消息可以显示在数字文档操作装置 205 上。这种显示可以通过使用图 1 所示布置中计算机 103 或打印机 104 的显示设备进行。

[0106] 在步骤 S1006,根据图 14B 所示的方法,计算要签名的数据 902 或 903 的提要值: H(M)。在步骤 S1007,检查提要计算结果是否匹配于包括在聚合数据 909 中的提要值 906 或 908。如果两个值匹配,则流程前进到步骤 S1008。如果在步骤 S1008 中存在下一标识符,则对要签名的对应数据执行步骤 S1004 中的处理。如果两个提要值不匹配,则签名验证处理结束,“NG”作为结果返回。如果在步骤 S1008 确定已经对所有标识符都进行了重复处理,则签名验证处理结束,“OK”作为结果返回。以下将描述图 4 中的数字文档操作处理 416。操作包括例如浏览、打印等的内容消费处理。但是,如何消费内容不影响这种实施方式,只要进行允许用户享受内容的处理就可以。因此,将省略这种处理的具体描述。另一方面,内容重建处理重建新的数字文档 411,且重建的数字文档 411 可以输入到数字文档操作处理 402。

[0107] 由于重建的数字文档 411 不是在数字文档生成处理 401 中生成的数字文档 411,因此其签名信息常常可以包括未归档内容的提要值。

[0108] 因此,以下将参考图 11A 和 11B 描述重建数字文档 411 的验证。在这种情况下,我们假定接收图 9A 和 9B 所示数字文档 411 的用户执行重建处理以删除要签名的数据 1(902),且只将要签名的数据 2(903)作为内容分发。

[0109] 如图 11A 和 11B 所示,在这个时候要分发的数字文档 411 被重写。更具体而言,数字文档 411 包括签名信息 901 和要签名的数据 2(903)。在这个时候,例如,当签名数据 901 通过例如删除在签名验证时不需要的数据 905 和 906 以便消除冗余而被修改时,签名值 904 本身变得无效。因此,包括信息 904 至 908 的签名信息 901 将没有任何修改地被使用。

[0110] 以下将检查其中处理基于图 10 中流程图进行的情况。在步骤 S1005,尝试基于要签名的数据 1(902)的标识符 905 获得要签名的数据 1(902)。但是,因为要签名的数据 1 没有归档到图 11A 和 11B 所示的数字文档 411 中,因此不能获得它(902)。因此,在步骤 S1005 中确定为“NO”,流程跳到步骤 S1008,从步骤 S1004 开始的处理对下一标识符(在这种情况下是标识符 907)继续。以这种方式,对于要签名的数据 1(902),跳过步骤 S1007 中的提要匹配处理。另一方面,在步骤 S1004 获得要签名的数据 2(903),并执行提要匹配处理。因此,可以验证要签名的数据 2(903)是否被改变。

[0111] 以这种方式,可以提供对包括在聚合数据 909 中但未归档到数字文档 411 中的子

内容跳过提要匹配处理并对归档的子内容保证未改变 / 改变的机制。

[0112] 在传统的签名生成处理中, 签名值必须分别提供给要签名的数据 1 和 2 (902 和 903)。因此, 计算处理的负担加重。特别地, 计算量与划分的要签名的数据的划分数量成比例增加。

[0113] 相反, 在这种实施方式中, 签名值的计算处理可以只进行一次, 而不考虑内容的划分数量。这样, 根据这种实施方式, 签名生成和签名验证处理可以比现有技术的执行有效得多。即使当数据只使用一些子内容重建时, 也可以可靠地验证子内容是否被改变。

[0114] 如上所述, 根据本发明, 即使当作为这种复合内容一部分的子内容独立存在时, 也允许不仅对文本数据而且对以多种格式存储的数字数据的复合内容进行签名验证。此外, 签名生成和签名验证处理可以有效地执行。

[0115] < 第二实施方式 >

[0116] 图 10 所示第一实施方式的验证处理不考虑其中当有些子内容被改变而剩余子内容未改变时用户允许内容的情况。因此, 这种实施方式将解释可以处理这种情况的方案。

[0117] 当找到作为步骤 S1006 中计算结果获得的其提要值不匹配于包括在聚合数据 909 中提要值的子内容时, 确定签名验证处理在图 10 的步骤 S1007 结束。但是, 即使在这种情况下, 如果两个提要值匹配的子内容或要处理的子内容存在且可以保证不改变, 则签名验证处理可以继续。

[0118] 因此, 在这种实施方式中, 即使当步骤 S1007 中的匹配结果为 NG 时, 从步骤 S1003 到 S1008 的验证处理也对包括在聚合数据 909 中的所有剩余标识符继续, 而不强制结束处理。然后, 作为验证结果, 返回未改变的子内容和改变的子内容列表。以这种方式, 可以通过计算机 103、打印机 104 等通知用户与各子内容改变存在 / 不存在关联的信息。以这种方式, 当有些子内容被改变而其它子内容未改变时, 用户可以允许内容。因此, 可以提供允许未改变的子内容被重新使用的机制。

[0119] < 第三实施方式 >

[0120] 这种实施方式将解释其中用户可以选择要签名的数据的情况。在以上实施方式中, 签名处理在签名信息生成处理 407 中执行, 该处理的细节已经利用图 8 进行了描述。在图 8 中, 整个数字文档数据作为要处理的数据进行了处理, 但签名处理不是通过选择文档数据的任何区域执行的。

[0121] 这种实施方式的特征在于在中间数字文档生成处理 405 和签名信息生成处理 407 之间提供了用于选择要签名的数据的新处理。在这种实施方式中, 这种处理将称为要签名的数据的选择处理。以下将描述要签名的数据的选择处理。

[0122] 在要签名的数据的选择处理中, 在纸质文档输入处理 404 中扫描的图像数据以图 6A 所示的格式显示在装置的屏幕上。在这种情况下, 用户可以利用如鼠标等的设备指针指定数据的矩形区域。例如, 用户可以利用设备指针指定描述“tour to visit...from Great Britain in1901”的区域。

[0123] 当以图 6B 所示的格式进行显示时, 有些矩形信息条 (602 至 606) 可以利用设备指针选择。由于这种矩形信息是可以作为已经内部保存的数据结构容易地处理的划分单元, 因此这种矩形信息的选择是对应于在选择后立即执行的签名信息生成处理 407 的处理。

[0124] 图 12 示出了其中两个划分区域 602 和 606 作为要签名的数据从图 6B 中选择出来

的例子。在图 12 中,所选的划分区域被高亮显示,从而提供允许用户容易地识别所选区域的屏幕结构。

[0125] 相反,用户可能常常想标记比在中间数字文档生成处理 405 中所划分区域窄的区域。例如,在有些情况下,更有可能在将来作为子内容被划分的区域比在中间数字文档生成处理 405 中所划分的区域窄。

[0126] 假定这种情况,即如图 13 所示,根据这种实施方式,所划分的区域可以在用户界面上划分成更精细的区域。如从这个例子中可以看到,比区域 606 窄的区域 1301 被选择并高亮显示。

[0127] 当允许指定更窄的区域时,期望区域(例如,图 13 中的区域 1301)的指定可以根据中间数字文档生成处理 405 中在步骤 S502 中的区域划分从用户接受。这种指定可以在用户利用设备指针指定期望区域时接受。这种技术对本领域技术人员是已知的,因此在本说明书中省略对其的具体描述。

[0128] 当在中间数字文档生成处理 405 中划分的区域可以进一步精细地划分,而且可以在签名信息生成处理 407 中用作要签名的数据时,可以提供对用户有更高自由度的要签名的数据的选择方法。

[0129] 应当指出,区域 1301 可以用作一个划分区域,而且区域 606 和 1301 之间的不同信息可以用作新的划分区域。在前一情况下,数字文档 411 的数据大小增加了,但处理容易。在后一情况下,需要新区域划分处理。

[0130] 如上所述,用户可以选择要签名的数据,而且可以执行签名信息生成处理。用户不仅可以事先指定所划分的矩形区域,而且可以指定作为要签名的数据的任意区域。

[0131] < 基于其它密码算法的实施方式 >

[0132] 在以上实施方式中,描述了基于公钥密码系统的加密处理(秘密转换)。但是,本发明可以容易地应用到基于密钥密码系统和 MAC(消息认证代码)生成方法的加密处理方法,而且本发明的范围包括其中通过应用其它密码算法实现以上实施方式的情况。

[0133] < 其它实施方式 >

[0134] 应当指出,本发明可以应用到包括单个设备的装置或者由多个设备构成的系统。

[0135] 此外,本发明可以通过向系统或装置直接或间接地提供实现了上述实施方式功能的软件程序、利用系统或装置的计算机读取所提供的程序代码、然后执行该程序代码来实现。在这种情况下,只要系统或装置具有程序的功能,实现的模式不需要依赖于程序。

[0136] 因此,由于本发明的功能是由计算机实现的,因此安装在计算机中的程序代码也可以实现本发明。换句话说,本发明的权利要求还覆盖了用于实现本发明功能的计算机程序。

[0137] 在这种情况下,只要系统或装置具有程序的功能,程序可以任何形式执行,如对象代码、由解释器执行的程序或提供给操作系统的脚本数据。

[0138] 可以用于提供程序的存储介质的例子是软盘、硬盘、光盘、磁光盘、CD-ROM、CD-R、CD-RW、磁带、非易失类型的存储卡、ROM、DVD(DVD-ROM、DVD-R 或 DVD-RW)。

[0139] 对于提供程序的方法,客户端计算机可以利用客户端计算机的浏览器连接到因特网上的站点,而且本发明的计算机程序或者程序的自动安装压缩文件可以下载到如硬盘的记录介质。此外,本发明的程序可以通过将构成程序的程序代码划分成多个文件并从不同

的站点下载文件来提供。换句话说,通过计算机向多个用户下载执行本发明功能的程序文件的 WWW(万维网)服务器也被本发明的保护内容覆盖。

[0140] 还有可能在例如 CD-ROM 的存储介质上加密并存储本发明的程序、向用户分发存储介质、允许满足特定要求的用户通过因特网从站点下载解密密钥信息并允许那些用户通过使用密钥信息解密加密的程序,由此程序安装到用户的计算机中。

[0141] 除了根据实施方式的上述功能通过由计算机执行读出的程序来实现的情况,运行在计算机上的操作系统等也可以执行全部或部分的实际处理,使得上述实施方式的功能可以由这种处理实现。

[0142] 此外,在从存储介质读取的程序写到插入计算机的功能扩展板或在连接到计算机的功能扩展单元中提供的存储器后,安装到功能扩展板或功能扩展单元的 CPU 等执行全部或部分的实际处理,使得上述实施方式的功能可以由这种处理实现。

[0143] 由于本发明的许多表面上非常不同的实施方式可以在不背离其主旨与范围的情况下进行,因此,应当理解除了在所附权利要求中定义的之外,本发明的实施方式不限于其特定的实施方式。

[0144] 尽管本发明已经参考示例实施方式进行了描述,但应当理解本发明不限于所公开的示例实施方式。以下权利要求的范围是要符合最宽的解释,从而包含所有这些修改与等效结构和功能。

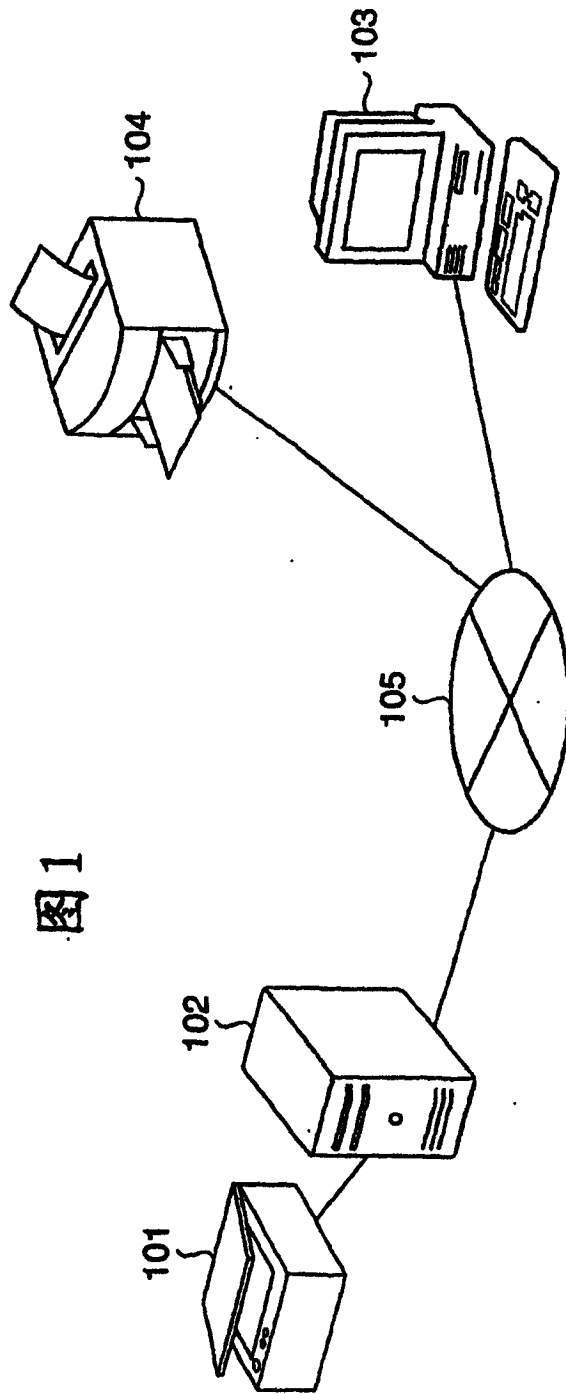


图 1

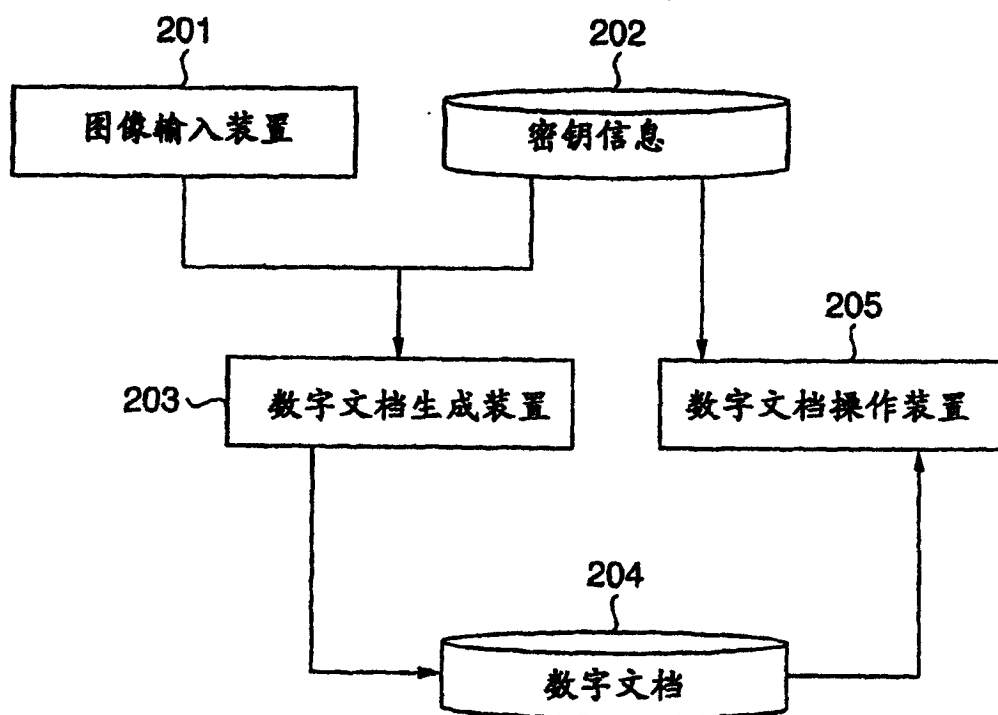


图 2

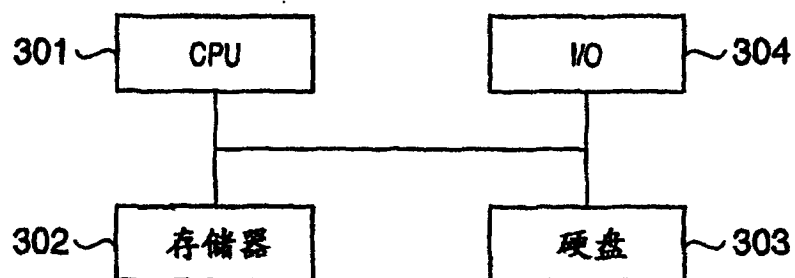


图 3

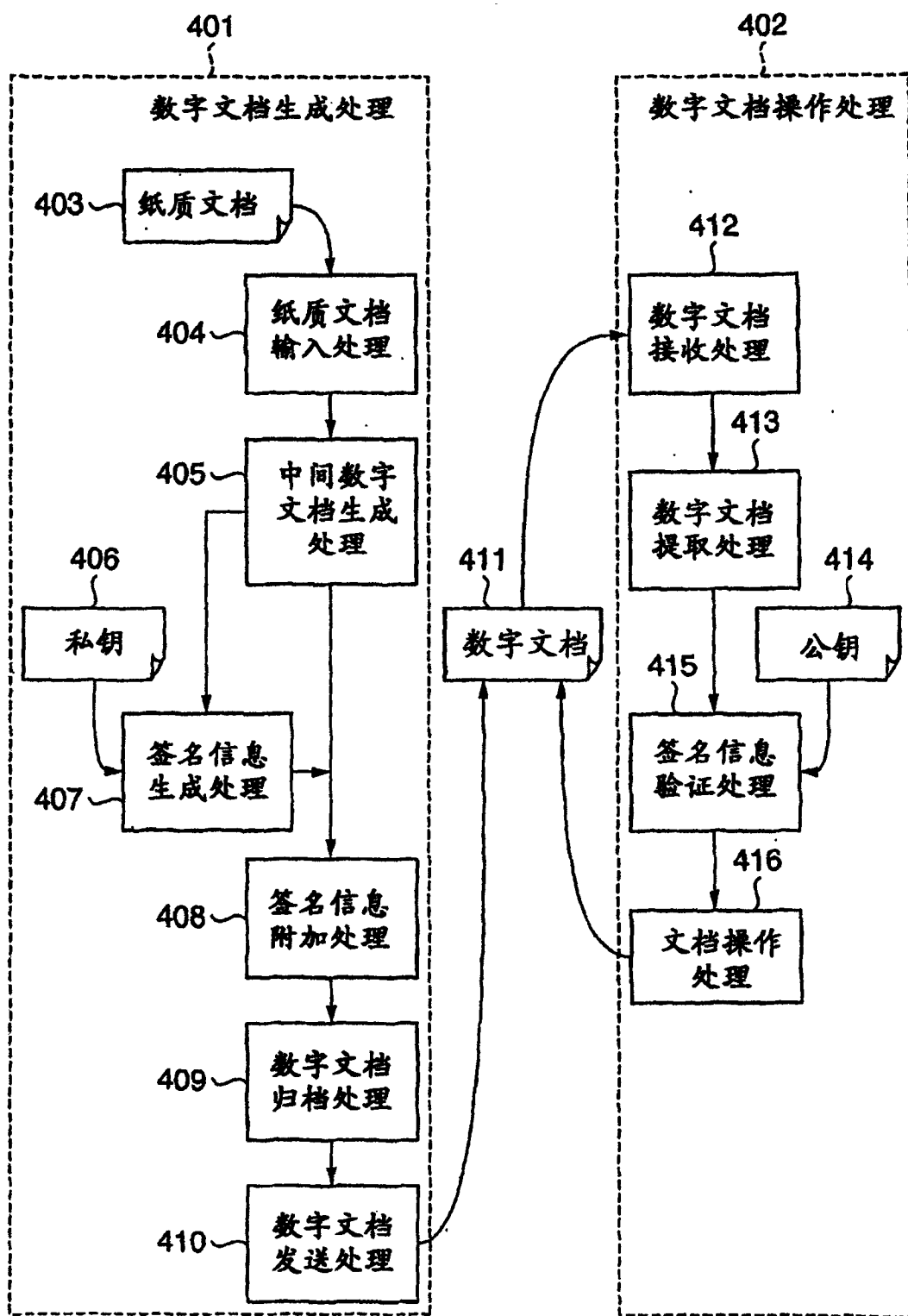


图 4

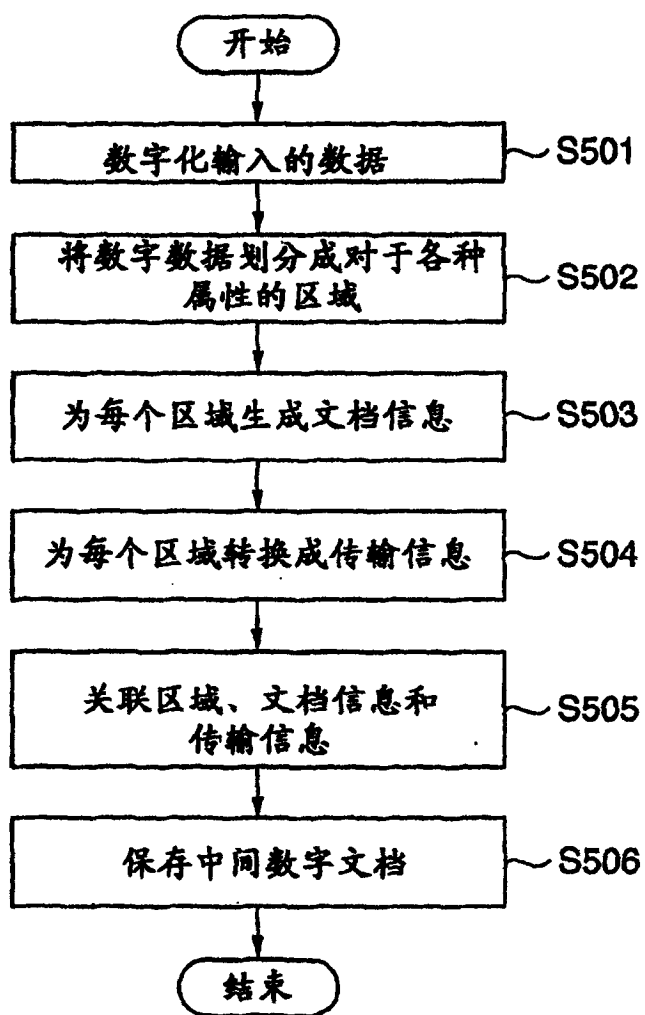


图 5

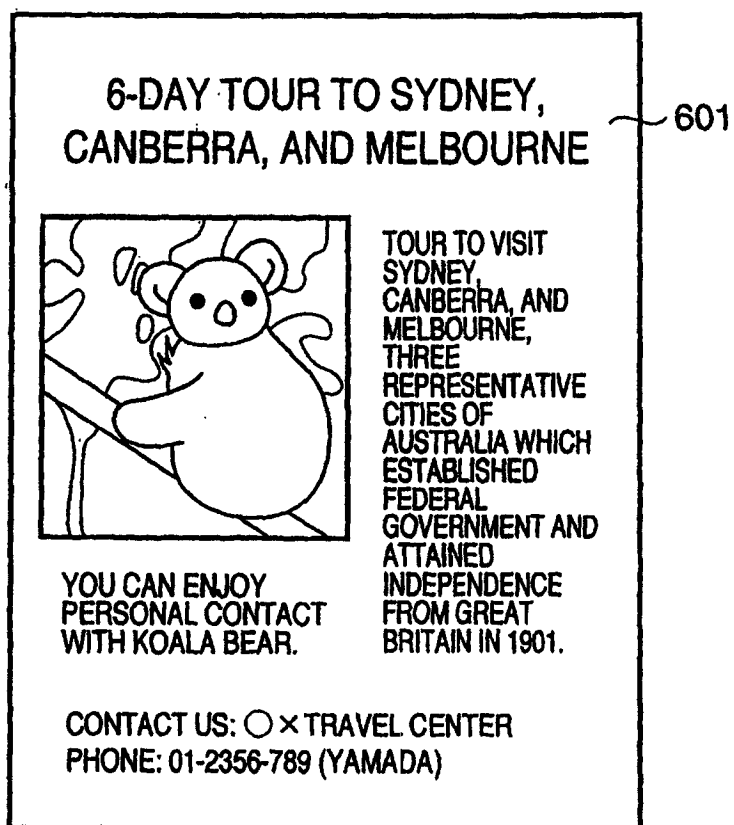


图 6A

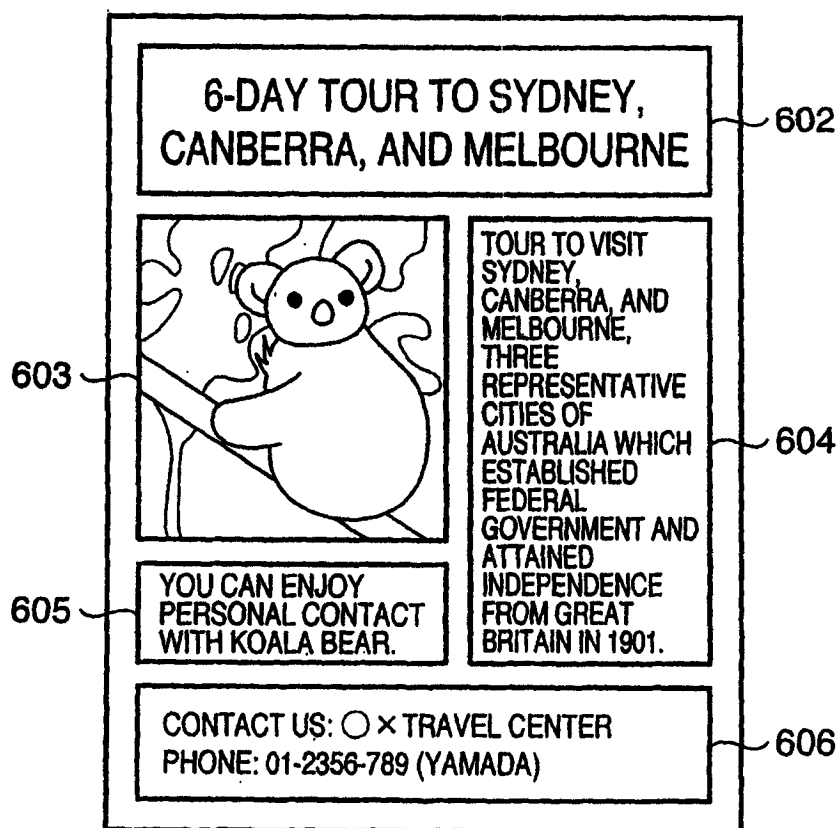


图 6B

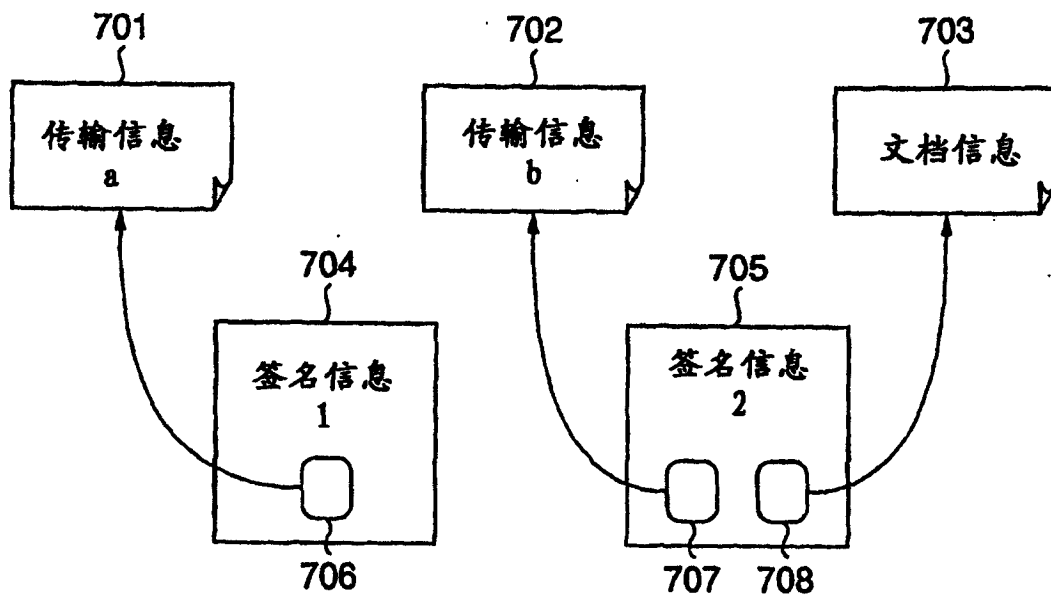


图 7A

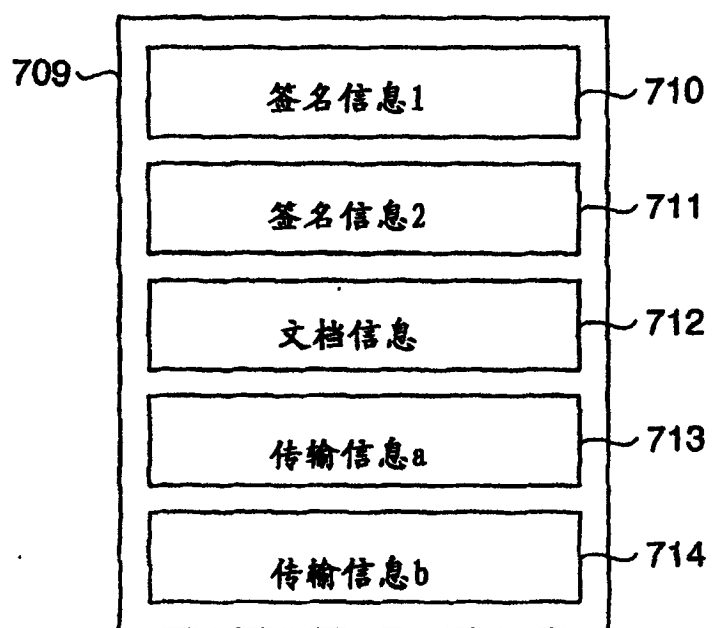


图 7B

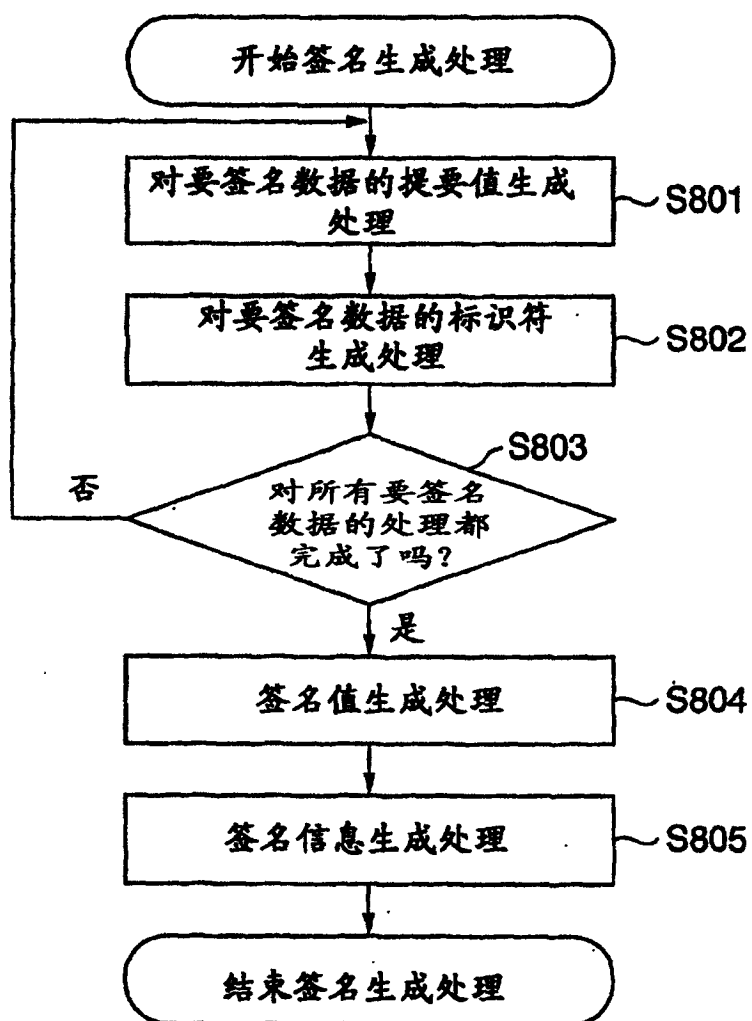


图 8

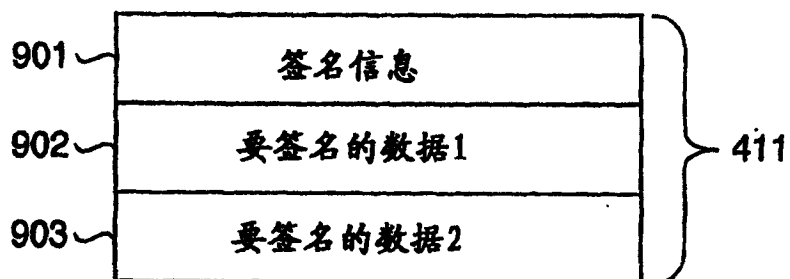


图 9A

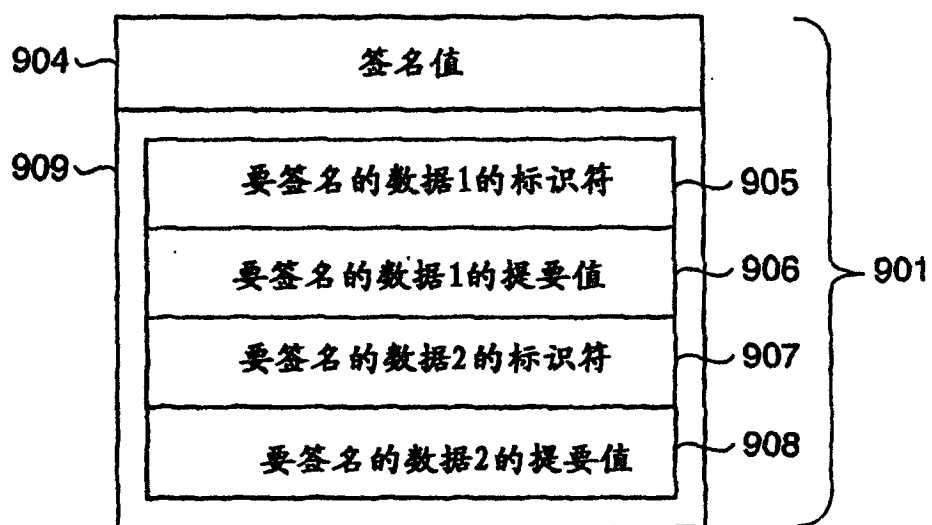


图 9B

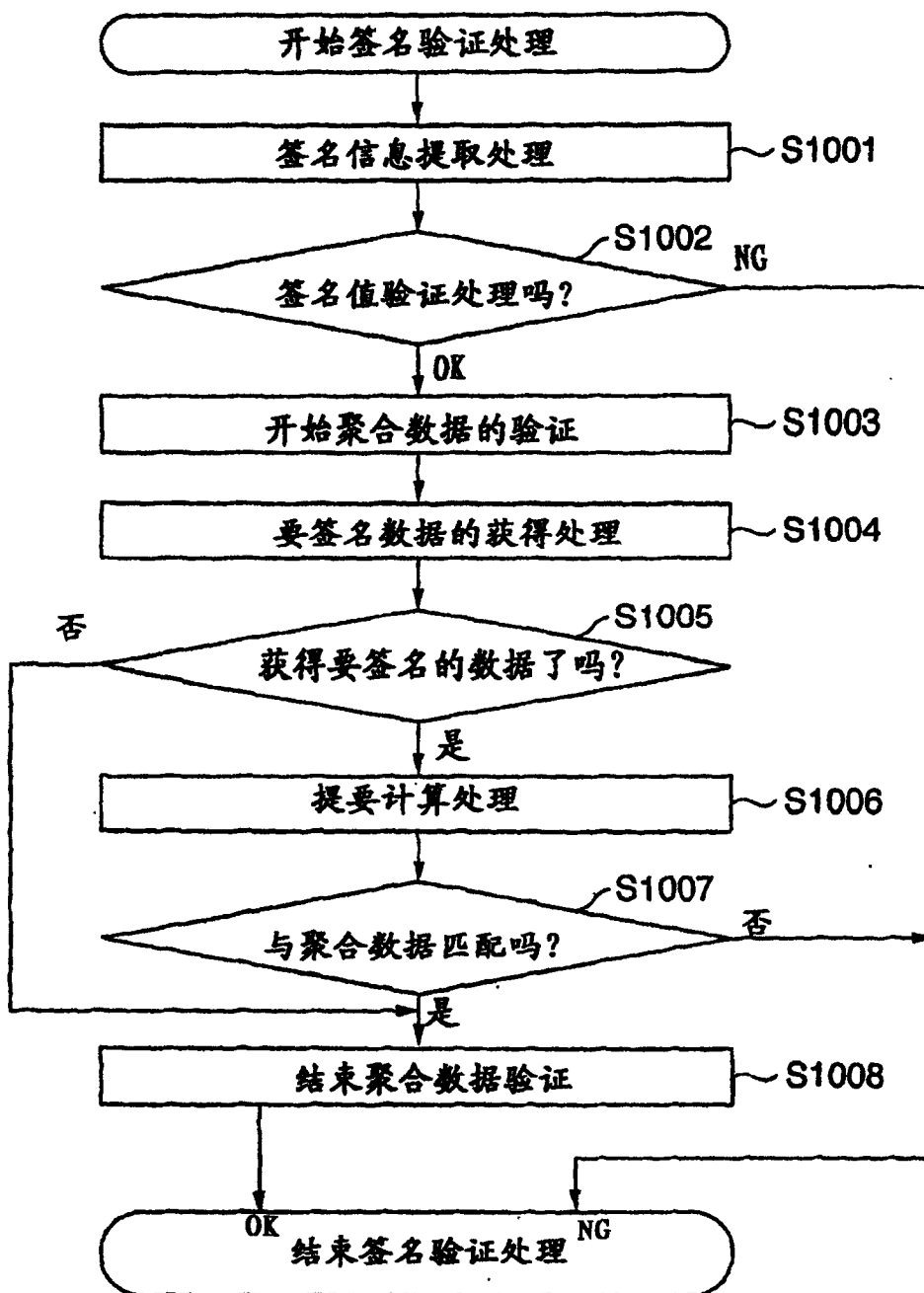


图 10

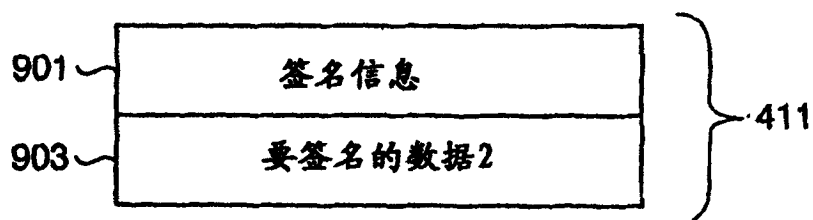


图 11A

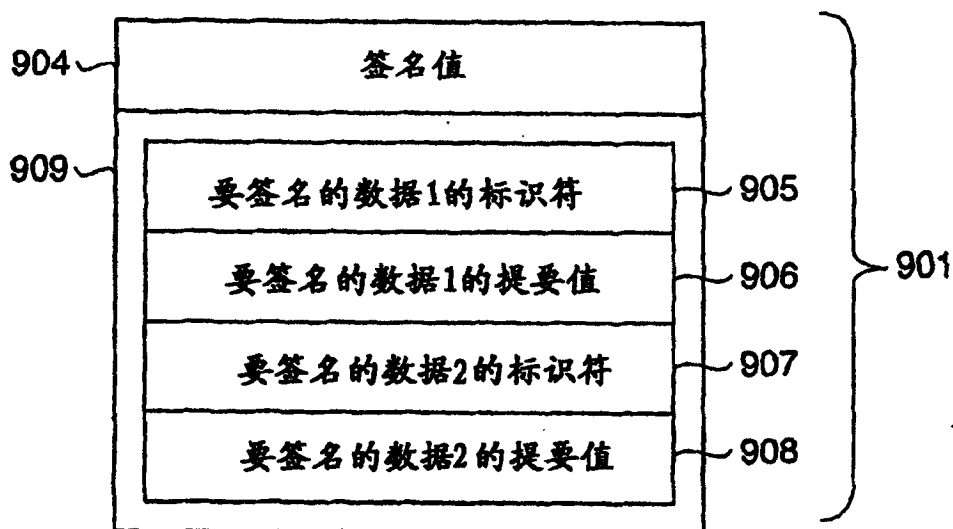


图 11B

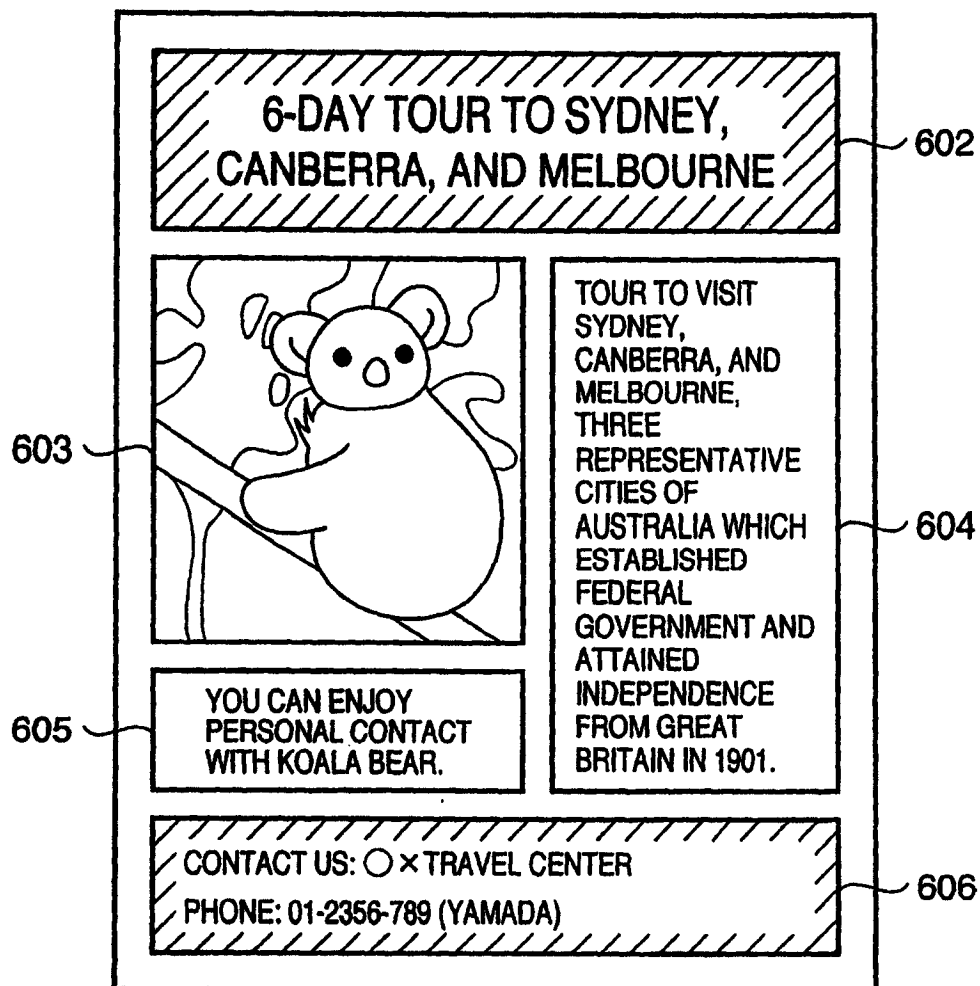


图 12

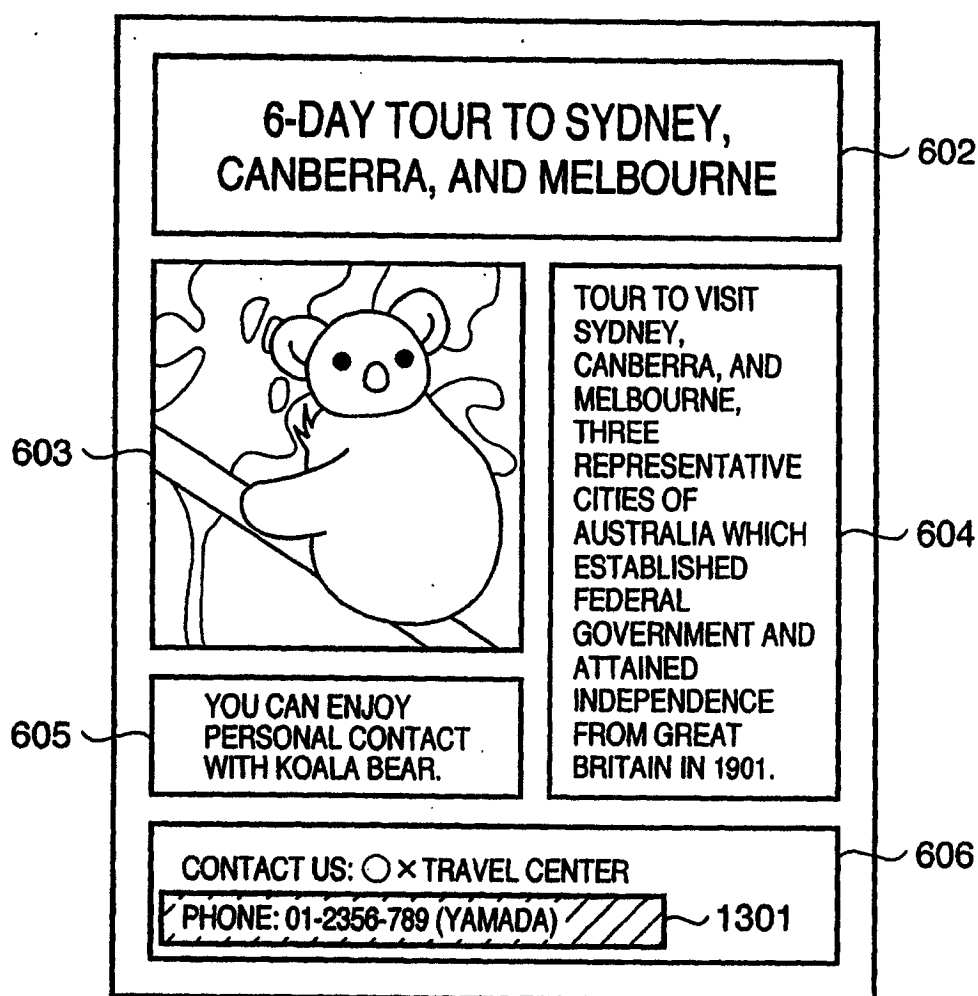


图 13

图14A

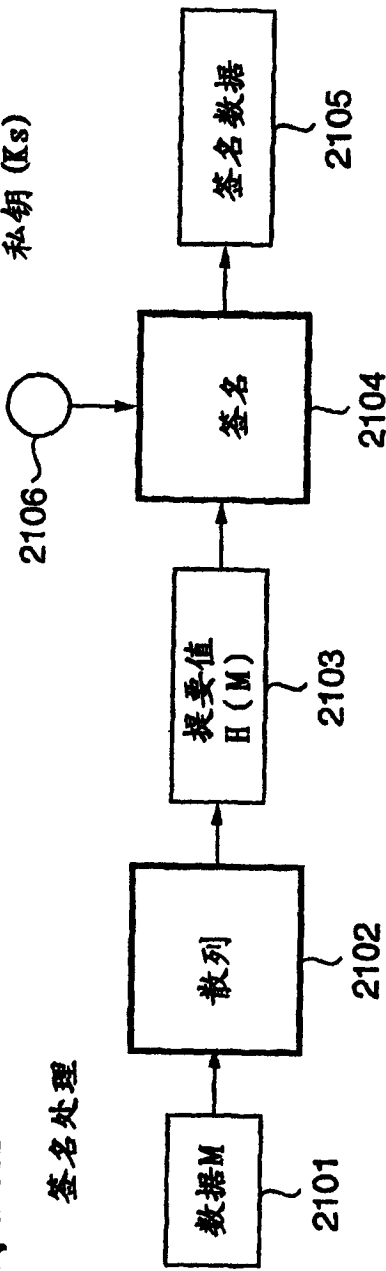


图14B

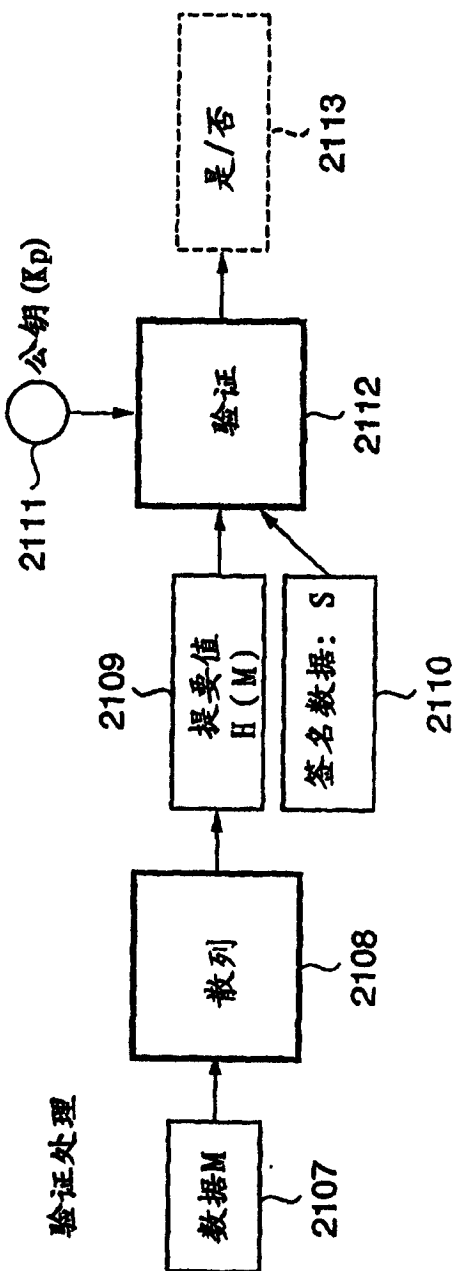
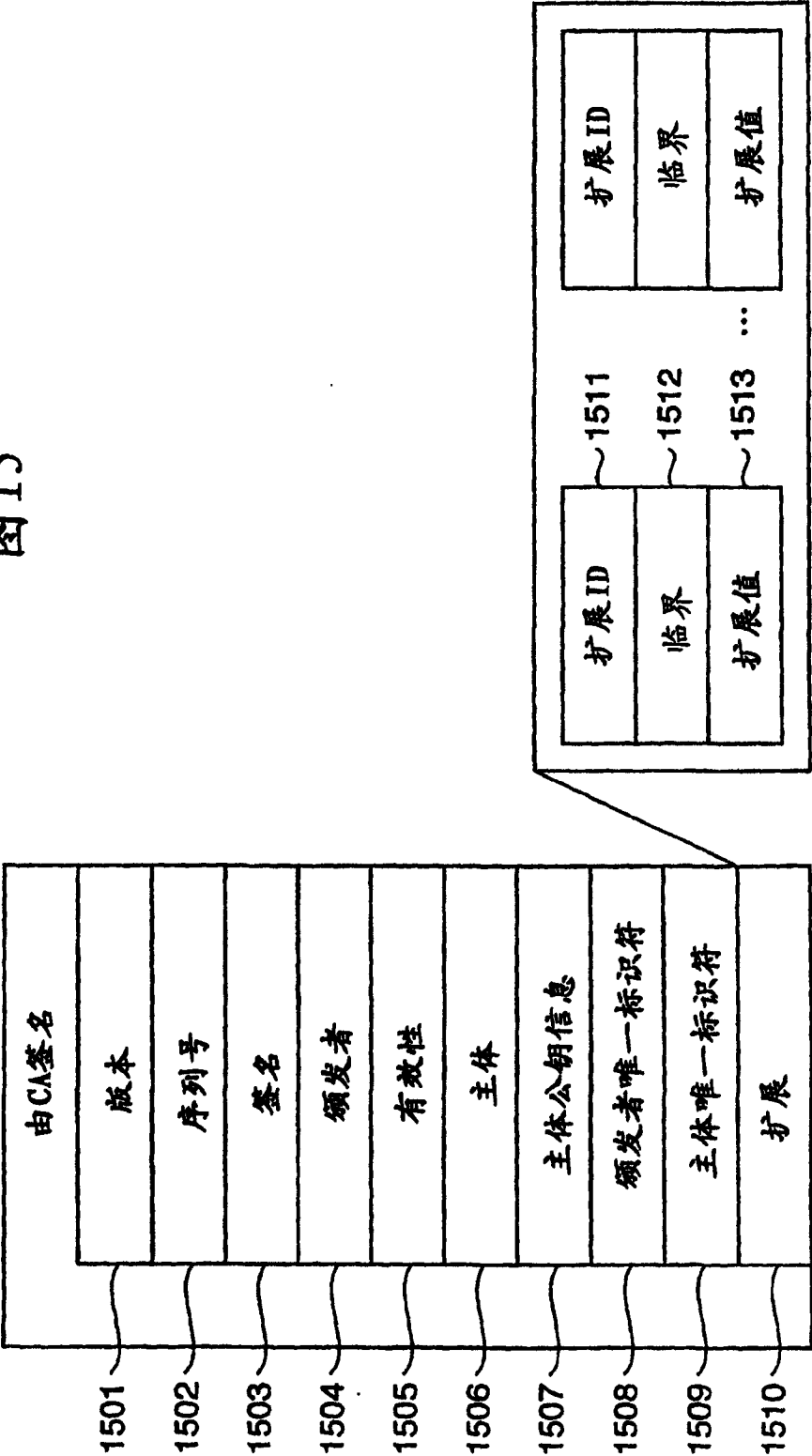


图15



X.509 v3格式