文章编号:1001 - 9081(2005)12 - 2770 - 02

公钥基础设施中证书撤销机制的研究与改进

徐成强,朱方金,史清华
(山东大学 计算机科学与技术学院,山东 济南 250061)
(qianggemen@ mail. sdu. edu. cn)

摘 要:从证书序列号出发,引用位标识指针将证书序列号缩减,以减少证书撤销列表(Certificate Revocation List,CRL)所需空间,提高 CRL 的查询速度,并成功构造了一棵新型撤销证书查询树。该树既继承了证书撤销树(Certificate Revocation Tree,CRT)证明一个证书的状态(是否被吊销)不需要整个 CRT,而只与其中部分相关路径有关的优点,又克服了 CRT 在更新时几乎需要对整个树重新计算的缺点。该树在更新时仅需计算相关部分路径的数值,加速了撤销树的更新速度。

关键词:公钥基础设施;证书撤销机制;证书撤销列表;证书撤销树
中图分类号: TP393.08 **文献标识码**:A

Research and improvement of certificate revocation mechanism in PKI

XU Cheng-qiang, ZHU Fang-jin, SHI Qing-hua

(*Department of Computer Science and Technology, Shandong University, Jinan Shandong* 250061, *China*)

Abstract: To decrease the storage space and improve the search velocity of CRL(Certificate Revocation List), a bit pointer was used to shorten the certificate number of it. And a new certificate recocation tree was proposed, which could keep the good properties of CRT(Certificate Revocation Tree) that is easy to check or prove whether a certificate is revoked or not, the check only need the related path values but not the whole CRT values. The new tree also could overcome the disadvantage of CRT that any update will cause the whole CRT to be computed, so it accelerate the speed of the CRT update.

Key words: Public Key Infrastructure (PKI); certificate revocation mechanism; Certificate Revocation List (CRL); Certificate Revocation Tree(CRT)

0 引言

证书撤销是公钥基础设施(Public Key Infrastructure, PKI)的一个重要组成部分,传统证书撤销方式采用的是证书撤销列表(Certificate Revocation List,CRL)发布机制。CRL 是一种包含撤销信息的证书列表的签名结构,它的发布有一定的周期性,基于这个特点可以通过缓存 CRL 来提高查询机制的性能,CRL 缓存使得客户可以在离线时验证证书。但 CRL 还存在着许多需要改进之处:CA(Certificate Authority)发布 CRL 列表的周期难以确定;CRL 存储库性能,尤其是如何有效降低大规模 PKI 系统中存储库峰值负荷;CRL 的增大使验证周期长等。具体地说,终端实体的数目、撤销的概率、已颁发证书的验证周期、证书序列号的大小等都会影响 CRL 大小[1]。本文对证书序号进行调整来改善 CRL 的存储、查询以及撤销树更新等问题,并成功构造了一棵新型撤销证书查询树,以加快查询和更新速度。

1 对证书序号的调整

CRL 只存储撤销证书序号及证书颁发者即可表示唯一的使用者身份[1]。X.509 在定义上对序号的格式并有没明确地说明。证书序号是递增且唯一的,此序号用以对证书进行认证、查询等。考虑到一个企业或区域中 CA 可以颁发的证书的数量,要求具有足够的字节长度来表示最大证书的序号,这样在 CRL 或证书库中所存储的具有较少有效位的序号必然造成存储空间的浪费(此时,无效字节、字的值为零)。也正是由于证书序号(最大值)而导致了 CRL 列表和证书库存储证书号效率的低下。这里从存储证书号所用的有效比特位的角度出发,将文献[4]中的字节动态增长的方法应用于此,进行无用位的删除,在读取证书序号时按照所规定的由若干比特位组成的元组进行读取,通过对大量整数数据的实测能够提高约 25%[4] 以上的查询速度。因为证书序号均为整数,与文献[4]的研究非常类似,且本研究又做了一些有利改进,所以有理由相信,此方法应用于 CRL 可以得到更加明显的效果。

1.1 位标识指针的定义

定义 令每字节的最低比特位(或多个比特位)作为一个标识指针。值为 1,表示该字节并非为最低字节;值为 0,表示该字节为最低字节。

举例如下:

(1) 517 的二进制表示:00000000-00000000-00000010-00000101

加入位指针去掉无效位后为:00000101-00001010

规定一:当某一个数字的高字节为 0 时,高字节为无效字节,将其删去。

(2) 17891328 的表示为:00000001-00010001-00000000-00000000

加入位指针去掉无效位后为:00000011-00100011-

收稿日期:2005 - 06 - 07;**修订日期**:2005 - 08 - 23 **基金项目**:山东省科技厅基金资助项目(003090309)

作者简介:徐成强(1980 -),男,山东济南人,硕士研究生,主要研究方向:网络信息系统; 朱方金(1971 -),男,山东平原人,讲师,主要研究方向:网络信息系统; 史清华(1967 -),男,山东济南人,副教授,主要研究方向:网络信息系统.

00000001

规定二:当最低字节为 0 时,最低字节为无效字节,只将其删去即可。

1.2　位指针意义

位指针的值只表示该字节后有字节,即该字节是否为最低字节。位指针并不表示数值中实际值的大小。在具体应用中每字节中实际使用的有效位数可以灵活定义,如定义字段 5 有效比特(3 比特标识指针),6 有效比特(2 比特标识指针)等。在实际使用中,数值的表示与加入位指针后的表示需要从高到低字节进行一定的编码转换。逆变换的规则为从低到高逐字段进行。引入位指针进行证书序号变换后存储空间比以前减少了,相应的 CRL 查询速度也有一定的改善。

1.3　CRL 中首尾标识法

在 CRL 中并不存放作废证书的全部内容,只存放作废证书的序列号[1]。若某些撤销证书号为连续号码则不必逐一列出所有证书号码,只需给出首序号和尾序号即可。若某些撤销证书号不连续但是介于其间的证书已经过期,则仍可以用此方法缩减 CRL 的长度。

例如:241-267 表示自号码 241 开始到 267 为止所对应的证书均已撤销,而中间的证书如 251,260 是过期的。

1.4　编码转换的开销

在以上讨论中,可知对于证书序号的操作涉及到编码转换工作。重新编码需要的时间取决于计算机的 CPU 速度、内存,缓存等,而本研究中解码及其之后的查找耗时远小于重新编码之前的耗时[4],所以本研究是可取的。本应用中证书序号的变动使得 CRL 的长度减小,系统读取资料的时间明显变短,系统 I/O 次数也远小于原机制。

2　一种新型的证书撤销树(CRT)

2.1　CRT

Kocher 提出了证书撤销树(Certificate Revocation Tree,CRT)方案[3],主要目的是证书验证者能得到一个证书状态(有效还是吊销)的简短证明。CRT 是 Hash 树,其叶节点对应于撤销的证书,但是其更新成本过高。更新是指最新被吊销的证书号需要加入 CRT 中,或原来 CRL 中被吊销证书因到达其失效期而自动失效者需从 CRT 中删除。假如有一个证书新被吊销,则 CRT 树几乎全部需要重新计算,因为其采用了证书号的完全顺序排列。在文献[2,5]中采用了 2-3 树的 CRT,但是需要经常调整树的平衡,在大规模 PKI 中撤销的证书较多,树的深度便成为不得不考虑的问题。文献[6]给出了一种 CRTBSHT 树,它仍是一棵二叉树,存在与 2-3 相似的问题,且其结构中绝大多数节点需给出有效证书区间。

2.2　新型 CRT

为了提高对已撤销的证书序号的查询速度,在此构造一棵新型的撤销树,称为定路径撤销树,每一个撤销证书的序号对应于树中的一个固定路径和节点。

对于一个 32 位的证书序号来说,如果将每一比特位看作一个深度层次则一个 32 字长的证书序号对应的树的深度最高为 28(去掉 4 个标识指针)。此处,可以动态地采取将 2 个比特位、3 个比特位等作为一个深度层次单元,从而树的深度最高为 14 或 8 等,具体使用情况视 PKI 域的规模及策略而定。例如以 2 个比特看作一个层次单元,为更直观理解作以下映射:

00:A 01:B 10:C 11:D

如证书号为:00010001 01100100,去掉无效位和位指针后:01-00-00-11-00-10,映射为树中路径:BAADAC,如图 1。

树中节点对应的结构为:SID|Data|*p|hash。其中 SID 为符号标识(A,B,C,D);Data 为有效位标识,值为 1,说明此证书已撤销,值为 0 说明未撤销;*p 为对应的指向该节点儿子的指针组;hash 值为该节点的所有儿子节点值和该节点 SID 及 Data 值的哈希函数值,对于根节点还需 CA 加入时戳后进行签名。

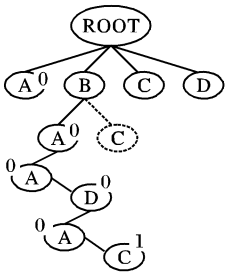

图 1　映射为树中的路径示意

这棵新型的撤销树因为树中节点的儿子具有一定的顺序性,所以能够很容易地给出一个证书的有效与否的简单证明,其证明可见后面的操作部分。

当此 CRT 在一个规模的 PKI 中高度不合适时,可以进行相应的动态变换,如图 1 中是以 2 个比特位作为一个高度层次单元,可以变为 3 个比特位的单元,也可以变换成一个比特位的层次单元,即 CRT 成为二叉树。

2.3　树更新

树的更新主要有两种操作:< Insert,e >,< Remove,e >[5]。对于插入,若树中已存在对应的节点,则将此中间节点的 Data 值更新为 1,表示该节点对应证书为撤销;若不存在对应的节点,则按照证书号固定路径进行添加。删除操作,若节点为中间节点,则将此节点的 Data 值更新为 0;若此节点为叶节点,则删去此节点,并同时判断其父节点。若父节点无其他儿子且 Data 值为 0,则将此父节点删去,再判断父节点的父节点,依次进行。

对每次更新操作无需对整个树的全部节点重新计算,只需计算相应节点到根节点路径中的节点即可。即若 32 位证书序号以 2 比特作为深度单元,每次更新操作需重新计算的节点不会超过 14 个;若以 3 比特作为深度单元,每次更新重新计算的节点不会超过 8 个,这样维护 CRT 只需很小的开销。

3　PKI 中各实体主要操作

3.1　CA

CA 通过签名一个包含证书序号、用户名及其公钥的消息来生成证书;在具有证书之后 CA 根据要撤销的证书序号构造一棵 CRT。CA 向 Directory 发送要撤销证书在 CRT 中的路径及对 CRT 根节点值和时戳的签名。CA 将 insert 和 delete 操作影响到的路径进行签名(含根结点的值)传送给 Directory 进行更新。

3.2　Directory

当 Directory 收到来自 CA 的证书撤销信息时,自动创建一个 CRT 并验证根结点的值;当 Directory 收到 CA 对 CRT 更新的消息时更新树中的相关节点并重新计算相关路径及节点值;当有 user 询问某证书的状态时,Directory 按照证书号路径搜索 CRT,如找到对应节点则回答包含自该证书节点到根节点的路径和相关节点值(这可作为证书撤销的证据);若无相关节点则回答自该证书所在路径的最深节点的其他儿子到根结点所在路径和相应儿子及根节点的值。

例如图 1 中询问节点为虚线节点 C 时,因为 C 节点对应的证书有效所以此时返回给用户的路径为虚线左边的 A 节点到根结点的路径中各节点及其相关儿子节点的值。回答为

(下转第 2782 页)

文档依次提取出 5、10、15 个关键词时，与传统 tf * idf 权重方法的性能比较。

表 1 与传统方法的比较

每篇文档提取词数		5	10	15
p	本文方法	0.4624	0.5032	0.5396
	传统方法	0.4235	0.4426	0.4713

由表 1 可以看出，在提取的关键词数为 5、10、15 时，本文方法比传统方法在准确率上分别提高了 9.2%、13.7% 和 16.6%。这说明，词语的平均位置对于标示该词是否为关键词有着较大的影响。也就是说，如果一个词在文档中的平均位置靠前，则该词是一个潜在的关键词。

实际上，该算法的一个优点是标示关键词的特征项是可以扩展的。如果存在其他可能的特征项，那么我们同样可以按照上述算法对其进行离散化处理以及计算分类概率。可以依次增加这样的特征项，观察提取结果的好坏，从而确定一个特征项可标示关键词的合理性。

(2)图 2,3 给出了在训练集大小不同时，从测试集的每篇文档依次提取出 5、10、15 个关键词时的性能比较(根据上述的两套标准)。

图 2,3 表明，在训练集的文档数达到 30 之后，提取出的关键词的准确率基本上趋于平稳。这就意味着只要选择一个相对较小的训练集即可保证比较好的提取性能。这样，在原始训练语料没有赋予关键词的情况下，只要进行少量的人工阅读和赋予关键词的工作。

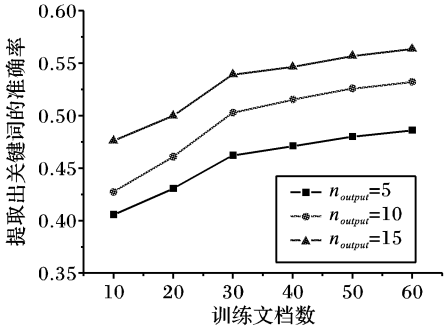

图 2 不同训练集大小下的性能比较(严格匹配)

在训练集小于 30 篇文档时，准确率的提高比较明显。我们认为，这主要是在离散化时，较少的特征值造成数据稀疏的现象，因而使断点无法准确判定出来，而这势必会大大影响模型的准确性和有效性。

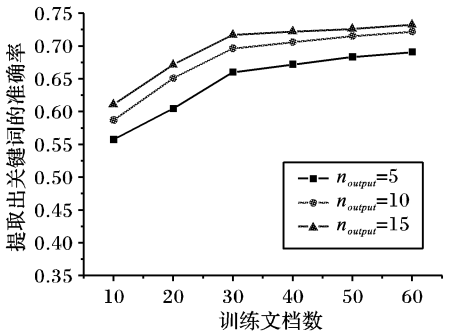

图 3 不同训练集大小下的性能比较(近似匹配)

4 结语

本文提出了一种基于朴素贝叶斯方法的中文关键词提取算法。与传统的方法相比，其性能有了明显的提高。另外，本算法的训练集在达到一个相对较小的数量后，即可获得比较稳定的性能。同时注意到，在该算法中，标示关键词的特征项是可以扩展的，可以通过增加类似的特征项，并观察提取结果的好坏，从而确定该特征项是否可用作标示关键词的一个指标。

参考文献：

[1] WITTEN IH, PAYNTER GW, FRANK E, *et al*. KEA: Practical Automatic Keyphrase Extraction[A]. Proceedings of ACM Digital Libraries Conference[C]. 1999. 254 - 255.

[2] MATSUO Y, ISHIZUKA M. Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information[J]. International Journal on Artificial Intelligence Tools, 2004, 13(1): 157 - 169.

[3] FRANK E, PAYNTER G, WITTEN IH, *et al*. Domain-Specific Keyphrase Extraction[A]. Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence(IJCAI-99)[C]. Stockholm, Sweden, Morgan Kaufmann, 1999. 668 - 673.

[4] KAGEURA K, UMINO B. Methods for automatic term recognition: A review[J]. Terminology, 1996, 3(2): 259 - 289.

[5] LEE-FENG CHIEN. PAT-Tree-Based Keyword Extraction for Chinese Information Retrieval[A]. Proceedings of the ACM SIGIR[C]. 1997. 50 - 58.

[6] FAYYAD UM, IRAN KB. Multi-interval discretization of continuous-valued attributes for classification learning[A]. Proceedings of IJCAI[C]. 1993. 1022 - 1029.

(上接第 2771 页)

C-A-D-A-A-B-ROOT 及相关儿子值(这可作为证书有效性的证据)。

3.3 用户

用户首先验证证书上 CA 的签名和有效期。用户向 Directory 查询一个证书，当收到来自 Directory 的回答时，用户验证 CA 对根结点的签名和路径节点的 Hash 值。

4 结语

当前对证书撤销机制的研究仍在继续，现存方案各有优缺点。文中所讨论的对证书序列号的变动在理论上应能减少 CRL 的存储空间，加快 PKI 领域中对 CRL 的查询的速度，同时也降低了 LDAP 服务器的负担；所提出的新型 CRT 完全可以给出证书有效、撤销状态的简短证明，每次搜索树时间复杂度为 O(树深度)，最坏情况下树为直线型。

参考文献：

[1] 关振胜. 公钥基础设施 PKI 与认证机构 CA[M]. 北京: 电子工业出版社, 2002.

[2] AHO AV, HOPCROFT JE, ULLMAN JD. Data Structures and Algorithms[M]. Addison-Westey, 1983.

[3] KOCHER P. On certificate revocation and validation[A]. Proceedings of International Conference on Financial Cryptography, Volume 1465 of Lecture Notes in Computer Science[C]. Berlin: Springer-Verlag, 1998. 171 - 177.

[4] WILLIAMS E, ZOBEL J. Compressing integers for fast file access[J]. The Computer Journal, 1999, 42(3).

[5] NAOR M, NISSIM K. Certificate revocation and certificate update[J]. IEEE Journal on Selected Areas in Communications, 2000, 18(1): 561 - 170.

[6] 王尚平, 张亚玲, 王育民. 证书吊销的线索二叉排序 Hash 树解决方案[J]. 软件学报, 2001, 12(09): 1343 - 1350.