BLAKE抗线性化差分分析安全性研究

毛 明[1] 贺 强[1,2] 曾绍昆[1,2] 张 珺[3]

(北京电子科技学院信息安全系 北京 100070)[1] (西安电子科技大学通信工程学院 西安 710071)[2]

(北京邮电大学电子工程学院 北京 100876)[3]

摘 要 基于模加、循环及异或运算的系统(ARX系统)通常被认为是设计密码算法的重要基础。在SHA-3第二轮候选算法中,BLAKE等杂凑函数基于该系统设计。通过对BLAKE-32中的模加运算进行线性化,分析了初始差分在各轮运算过程中的扩散情况,得出了初始状态字存在差分时各轮运算中差分的扩散特征。在此基础上,研究了多次加法运算的线性化逼近概率,并分析了线性化逼近方法对分析算法的有效性。研究表明,线性化情况下,部分初始差分字在BLAKE-32中的扩散效果并不理想,可以将BLAKE算法局部线性化后进行差分攻击。

关键词 杂凑函数,ARX,BLAKE算法,线性化,差分攻击

中图法分类号 TP309 **文献标识码** A

Security Analysis of Resistance against Differential-linear Attack on BLAKE-32

MAO Ming[1] HE Qiang[1,2] ZENG Shao-kun[1,2] ZHANG Jun[3]

(Department of Information Security, Beijing Electronic Science and Technology Institute, Beijing 100070, China)[1]

(School of Telecommunication Engineering, Xidian University, Xi'an 710071, China)[2]

(School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijng 100876, China)[3]

Abstract Hash function BLAKE is one of candidates for the second round SHA-3 competition, which is based on modular addition, rotation and XOR that is called as ARX system commonly. It is a common belief that the mixture of the three operations gives a good primitive in designing cryptographic algorithm. By replacing modular addition with XOR, this paper researched linearization of ARX system in BLAKE-32 algorithm, then, analyzed differential diffusibility of the algorithm after linearization and exploited some diffusion characteristics, furthermore, researched the probability of linear approximation of addition, and analyzed its validity on the algorithm. The result shows that differential spreading does not satisfy the designers' declaration. On account of ARX in BLAKE, differential attack can be applied on BLAKE by local linearizing its core function.

Keywords Hash function, ARX, BLAKE algorithm, Linearization, Differential attack

杂凑函数也称消息摘要函数、hash函数等,它是将任意不定长度的消息串通过映射变换成固定长度输出的函数。杂凑函数是密码学的一个重要分支,广泛应用于包括数字签名、数据完整性验证等信息安全领域。一个安全的杂凑函数应该具有抗差分攻击等各种攻击的能力。差分分析是研究密码算法安全性的一种重要分析方法,常用于分析基于ARX(Addition-XOR-Rotation)结构设计的密码算法。其中,MD5[1,2],SHA-1[3]等具有某些ARX结构特征的一系列MD结构杂凑函数的破解都利用了差分分析的思想。

在NIST征集的SHA-3候选算法中,有14种算法进入第二轮评估。在这14处算法中,BLAKE[4],Skein[5],Cube-Hash[6]等算法混合使用了异或、模加、循环3种运算,我们称之为基于ARX结构的杂凑函数。通常认为,只要算法的轮数达到一定数量,混合采用3种运算是设计安全杂凑函数的一种重要手段。然而,到目前为止,还没有理论能够证明基于ARX结构的杂凑函数一定安全。

针对BLAKE算法,Li Ji[7]等人利用消息置换的特点,进行了2轮的自由起始原象攻击,并给出了该攻击所需的计算时间复杂度。但迄今为止,还没有3轮以上原象攻击的公开结果。Guo jian, J. P. Aumasson[8]等人对部分G函数中的模加运算用异或运算替换后,采用差分分析方法,得到了4轮BLAKE-32运算的近似碰撞。在Guo jian等人的分析中,严格设置明文差分在4的倍数位置上,例如差分为0x88888888的形式,使差分扩散不经过循环7的步运算。该分析方法存在较大缺陷,无法针对更多轮的运算进行有效攻击。当前,对该算法的分析工作还没有取得更进一步的进展。

密码算法的应用必须经过严密的安全性分析,才能使信息安全得到保证。因此,本文将在不控制差分位置和差分路径的前提下研究线性化差分分析对BLAKE算法安全性的影响。提出了一种新的G函数线性化模型,分析了初始差分在

到稿日期:2010-08-05 返修日期:2010-11-14

毛 明(1963-),男,教授,硕士生导师,主要研究方向为信息安全、密码学,E-mail:maoming@besti.edu.cn;贺 强(1982-),男,硕士生,主要研究方向为信息安全;曾绍昆(1986-),男,硕士生,主要研究方向为信息安全;张 珺(1987-),女,硕士生,主要研究方向为电路与系统。

各轮的扩散情况，总结了差分扩散特征，进一步分析了线性化差分攻击对分析该算法的有效性。研究表明，部分消息字存在差分时，其差分扩散效果并不理想，理论上达不到位扩散的要求。对杂凑函数的 ARX 结构局部进行线性化，然后实施差分攻击，是分析算法的一种较为有效的方法。

1 G 函数线性化

Bernstein[9]分析了具有 ARX 结构的 Salsa20 算法，表明 ARX 系统在一定条件下可以转化为 AX 系统。Paul 和 Preneel[10]进一步发现了 AX 结构的系统化分析方法。对于 AR 系统，尽管理论上等价于 ARX 系统，但在运算数量相同的情况下，AR 系统的安全性没有 ARX 系统高[11]。因此，消除 ARX 系统中的任何一个运算符，都将影响以该结构为基础设计的杂凑函数的安全性。

通常，对于 ARX 系统构造的密码算法，常用的分析方法是先对 ARX 系统线性化，然后进行差分分析。所谓系统线性化，就是指在系统方程中，为了尽可能减小(避免)二进制数各比特位运算的相互影响，在满足结果的前提下，用线性运算取代非线性运算的过程。差分攻击是指不考虑状态变量的具体取值，而研究仅当其差分存在时对该算法进行攻击的一种方法。

BLAKE[4]算法继承了传统杂凑函数的一些设计思想。该函数分为 3 部分：初始化、轮运算、压缩。其中，轮运算中的 G 函数采用了典型的 ARX 结构，也是该算法的核心。根据消息分组长度和摘要值长度不同，BLAKE 算法分为 4 种形式：BLAKE-28，BLAKE-32，BLAKE-48，BLAKE-64，其原理相同。BLAKE-32 是 BLAKE 算法家族中的代表，也是本文的研究重点，其分组消息长度是 512-bit，初始值和摘要值均为 256-bit。在运算过程中，明文输入 m_i、初始状态 h_j、链路变量 v_i 和摘要值 h_j^r 均以 32-bit 长的字表示，其中 $i\in\{0,1,\cdots,15\}$，$j\in\{0,1,\cdots,7\}$，$r\in\{1,2,\cdots,10\}$。完整的 BLAKE-32 算法共有 10 轮消息置换运算，各轮的置换规则不同。每轮运算有 8 个 G 函数($G_0,G_1,\cdots,G_7$)，各 G 函数运算规则相同，所不同的是参与运算的参数。其中，前 4 个 G 函数(G_0,G_1,G_2,G_3)的输出值作为后 4 个 G 函数(G_4,G_5,G_6,G_7)的输入值，前 4 个函数并行运算，后 4 个函数也如此。每个 G 函数的参数包括 4 个状态变量、2 个消息字、2 个常数。算法原理见文献[4]。

为了研究链路状态字在各轮的差分扩散情况，我们对 G 函数进行线性化。记第 $s(s\in\{0,1,\cdots,7\})$个 G 函数在第 r 轮的运算为 $G_s^r(a,b,c,d)$，把模 2^{32} 加运算用 XOR 运算替换，线性化 G 函数(见图 1)。

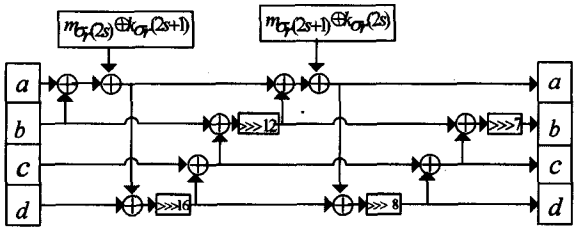

图 1 线性化 G 函数

图中，k 表示常数，σ 为明文消息字和常数置换规则。

10 轮计算后，如果不考虑盐值，则摘要值的计算公式为 $h_j^{10}=h_j\oplus v_j\oplus v_{8+j}$。这里及后面的分析过程，我们都没有考虑盐值的因素。

2 BLAKE-32 的线性化差分分析

对 BLAKE-32 中的 G 函数进行线性化后，本节研究当消息字和(或)初始状态字的 1-bit 和 2-bit 存在差分时，该函数减小轮数运算后的差分扩散情况。

2.1 1-bit 明文差分

对 G 函数线性化后，不管状态值如何，仅考虑差分值。假设明文只有 1-bit 差分，并遍历消息字的差分位置，执行 5 轮 BLAKE-32 运算，存储计算结果，得到各轮差分的扩散情况(见表 1)。

表 1 1-bit 差分各轮扩散数(单位：bits)

明文字差分	1 轮	2 轮	3 轮	4 轮	5 轮
m_0	65	182	236	263	261
m_1	50	195	254	252	275
m_2	65	186	244	264	264
m_3	50	197	250	268	249
m_4	65	181	259	252	243
m_5	50	191	247	243	244
m_6	65	182	242	252	256
m_7	50	197	252	270	247
m_8	14	147	249	264	273
m_9	4	126	251	256	242
m_{10}	14	117	223	246	257
m_{11}	4	119	249	257	273
m_{12}	14	127	234	251	246
m_{13}	4	140	269	273	245
m_{14}	14	138	228	260	248
m_{15}	4	117	229	257	225

根据各轮差分扩散的结果，我们总结出：在线性化情况下，当只有 1-bit 差分时，BLAKE-32 的差分扩散具有以下特征：

1)在任意一个消息字中，无论差分位置如何，各轮的输出都产生相同的差分数，只是具体运算过程中各链路状态的差分路径随输入差分的位置而相应变化。

2)遍历 16 个消息字差分后，我们发现，除当 m_5 存在差分时外，在差分扩散位数相对稳定(即约一半的比特数存在差分)以前，第奇数个消息字存在差分时差分的扩散速度不比相应偶数个消息字存在差分时的慢。

3)多数情况下，前 3 轮运算的差分扩散速度最快。第 4 轮运算结束后，差分数基本达到算法设计的差分扩散要求(m_5 存在差分时除外)。

4)当 m_8,m_{10},m_{12},m_{14} 存在差分时，链路差分扩散速度相对平稳，即雪崩效应较差；当 m_{10} 存在差分时，经过 4 轮线性运算，差分扩散数量才达到设计必须满足的汉明重量；而当差分位在消息字 m_5 中时，5 轮线性化运算后的差分汉明重量还达不到 256-bit，是差分扩散速度最慢的一个消息字。

分析表明，当只存在 1-bit 明文差分时，并非在每个位置所引起的差分扩散位数都相同，且扩散速度也有很大差异。即使经过 3 轮运算，也不能使任意 1-bit 的变化影响状态字中一半以上的比特位。例如 m_5,m_{10},m_{12} 存在 1-bit 差分时，并没有导致 3 轮运算的输出差分扩散到全部比特位数量的1/2。因此，位扩散和混淆的效果在线性化运算情况下达不到设计要求。

2.2 2-bit 输入差分

根据 G 函数内部运算中消息置换的特点，在前 4 个 G 函数中设置消息字和其对应的初始变量差分（对应初始变量是指上述 G 函数方程的变量 a）。通过实验，得到了 6 轮 BALKE-32 运算的差分扩散情况（见表 2）。

表 2 2-bit 差分各轮扩散数（单位：bits）

初始差分对	1 轮	2 轮	3 轮	4 轮	5 轮	6 轮
(m_0, v_0)	0	14	133	239	238	254
(m_2, v_1)	0	4	139	276	257	252
(m_4, v_2)	0	65	182	238	252	246
(m_6, v_3)	0	50	197	250	253	264

观察表中各轮线性化运算的差分结果，发现当 m_2，m_6 有差分时，其扩散速度较快；m_2，v_1 存在差分时，两轮运算后，只引起 4-bit 的链路变量差分，差分扩散最少。当差分位在 m_0，v_0 时，链路状态变量差分扩散速度最慢，6 轮运算后，才达到理想的扩散数量。并且，如果执行 9 轮线性运算，得到摘要值的差分数最少。

对于 3-bit 及以上初始差分的情况，多次线性化后的运算结果显示：其差分扩散性与非线性情况下的差分扩散性相比，没有差异，均能在 2 轮运算后达到理想的扩散效果。

3 线性化差分分析的有效性

Bernstein[9] 及 Paul[10] 等人的研究结果表明，简化 ARX 系统对于分析基于该结构的密码算法具有重要意义。对具有该结构的算法，第 2 节中详细分析了把 ARX 系统线性化为 XR 系统后链路状态变量差分在各轮运算过程中的扩散情况。该分析基于对各轮 G 函数的模加运算进行线性化，它表明了 BLAKE 函数的位扩散并不理想。通过分析消息字差分和（或）初始状态字差分在各轮运算中的扩散情况，可了解该结构压缩函数的位扩散效果和差分扩散路径，并有助于在分析中选取差分消息字。该结果基于把所有加法运算用异或运算取代而得到，而每一个模加运算的线性化都以概率来实现。

假设变量 x 用二进制表示时的长度为 n-bit，记 $x=(x_0, \cdots, x_{n-1})$，x_i 指 x 的第 $i(i\leqslant 0\leqslant n-1)$ 位，其值为 0 或 1。x 的汉明重量是指在变量 x 中 x_i 为 1 的数量，记为 $wt(x)$。如果一对变量 x 和 x'，$x\oplus x'=\alpha$，则称 x 与 x' 存在差分，且差分为 α；同理，记变量 y 和 y' 的差分为 β。其中，差分 α，β 的值不依赖于变量的取值。那么，一个模加运算的线性化概率逼近可根据如下定理得到。

定理[12] $p_r\{(x'+y')\oplus(x+y)=\alpha\oplus\beta\}=2^{-wt((\alpha|\beta)<<1)}=p_2$

假设 i 个加法运算相对独立，且满足上述定理，则根据此概率，我们很容易得到如下推论。

推论 $p_r\{\sum_1^i(x'+y')\oplus\sum_1^i(x+y)=\sum_1^i\alpha\oplus\sum_1^i\beta\}=p_2^i$

证明：当只有一个模加运算时，线性化概率已在文献[13]证明。如果对多个独立的模加运算进行线性化，则满足最终结果的概率就是各单个运算线性化逼近概率之积。假设各独立运算的线性化概率均相同，则推论必然成立。

在完整的 BLAKE-32 运算中，共有 480 次模加运算，根据定理的计算可知，差分位置在链路变量的高位或者低位时，线性化逼近概率较高。如果设置最高位差分，单步线性化逼近甚至以概率 1 发生。但每个 G 函数都有 4 次循环移位和多次模加运算，由推论可知，随着运算次数的增多，差分逼近概率将急剧降低。

如果把其 ARX 结构简化为 AX 结构，并假设加法运算的最高位没有进位，则 $X<<<r\equiv 2^r\cdot X(\mathrm{mod}\ 2^n-1)$[11] 的概率为 1/2。当加法运算个数为 q 时，其概率为 2^{-q}。由此可知，在 5 轮以上的 BLAKE 算法中，把加法和循环运算转换为只有加法的运算，所需计算复杂度至少为 2^{240}。而该算法即使转换，仍然还存在异或循环结构，即 ARX 形式。因此，在 BLAKE 算法中，并不能通过把 ARX 结构简化为 AR 结构来有效攻击 BLAKE 函数。即使能够化为 AR 结构，由于计算复杂度太高，对算法安全性也没有多大影响。

如果把算法中的 ARX 结构线性化为 XR 结构，遍历 BLAKE-32（其它 BLAKE 算法类似）不超过 4-bit 的输入差分，观察各轮链路状态输出差分的扩散情况，我们发现多数情况下，3 轮运算后的链路状态变量中有接近或超过一半数量的比特位存在差分。而根据线性化逼近概率公式，如果把 BLAKE 算法中的 ARX 结构完全线性化为 XR 结构，由于模加运算次数多，并且多次的循环置乱使得线性化逼近概率极低，因此，完全线性化不会降低攻击算法的计算复杂度。而本文第 2 节中的分析结果显示，线性化 G 函数后，各消息字引起的差分数不相等，即具有不均匀性；某些消息字存在差分时，差分扩散速度也不理想。由此可知，在线性化情况下，该函数消息字差分的雪崩效应并不理想，并且由推论可知，无论多少次线性化，都不会使其概率逼近为 0。因此，一定可以有条件地线性化部分 G 函数。

分析表明，对于 BLAKE 算法，还没有办法把 ARX 结构完全转化为 AX 结构。对于完整轮数的运算，也不能完全用异或运算代替模加运算。但对于减少轮数的 BLAKE 算法，对部分 G 函数的 ARX 结构局部线性化后进行差分攻击，是分析该算法安全性的一种较为有效的方法。

结束语 在密码算法中，很多算法基于 ARX 结构设计。本文以 SHA-3 第 2 轮候选算法中采用该结构的 BLAKE 算法为例，对具有典型代表的 BLAKE-32 中的模加运算用异或运算替换，即把 ARX 结构线性化为 XR 结构。首先分析了该函数中链路变量的差分扩散性，然后分析了对 ARX 系统局部线性化后进行差分攻击的可行性和有效性。设计者声称，两轮运算后，比特位能够达到理想的扩散效果。但我们的分析结果表明，即使 3 轮运算后的差分能够扩散为理想的效果，也并非每一比特的扩散结果都相同。并且，当部分输入状态字存在差分时，经过多轮线性化运算后，差分扩散还达不到理想的汉明重量。对于采用 ARX 系统的 BLAKE 算法，研究显示，如果不对差分位置做特殊限制，可以对 G 函数局部线性化后进行超过 4 轮的差分攻击。下一步将继续研究如何设置差分，并根据差分扩散路径来设置线性化 G 函数的模加运算位置和数量，以及在此基础上进行差分攻击的有效轮数。

参 考 文 献

[1] Wang Xiao-yun, Yu Hong-bo. How to break and other hash functions[C]// EUROCRYPT 2005. Advances in Cryptology. Berlin: Springer-verlag, 2005: 19-35

[2] 梁杰. MD5-Hash 函数的安全性分析[D]. 上海：上海交通大学，2007

[3] Wang Xiao-yun, Yu Hong-bo. Finding collisions in the Full SHA-1[C]// Crypto 2005: Advances in Cryptology. Berlin:

Springer-verlag,2005:17-36

[4] Aumasson J P, Henzen L, Meier W, et al. Sha-3 proposal blake [EB/OL]. http:// 131002. net/blake/blake. pdf,2009-08-11

[5] Bernstein D J. CubeHash specification(2. b. 1)[EB/OL]. http://ehash. iaik. tugrazat/wiki/CubeHash,2009-08-11

[6] Ferguson N, Lucks S, Schneier B, et al. The Skein hash function family[EB/OL]. http://www. skein-hash. info/sites/default/files/skein1. 1. pdf,2009-12-20

[7] Li Ji, Xu Liang-yu. Attacks on Round-reduced BLAKE [EB/OL]. http://ehash. iaik. tugraz. at/wiki/The _ SHA-3 _ Zoo, 2009-11-08

[8] Aumasson J P, Guo Jian, Knellwolf S, et al. Differential and invertibility properties of BLAKE[EB/OL]. http://ehash. iaik. tugraz. at/wiki/BLAKE,2010-05-15

[9] Bernstein D J. The Salsa 2 0 encryption function [EB/OL]. http://cr. yp. to/snuffle. html,2010-02-23

[10] Paul S, Preneel B. Solving systems of differential equations of addition[C]// ACISP'05. Berlin: Springer-verlag,2005:75-88

[11] Khovratovich D, Nikoli I. Rotational Cryptanalysis of ARX [EB/OL]. http://www. skein-hash. info/sites/default/files/ARX. pdf,2010-04-15

[12] Brier E, Khazaei S, Meier W, et al. Linearization Framework for Collision Attacks: Application to CubeHash and MD6(Extended Version)[C]// ASIACRYPT 2009: Advances in Cryptology. Berlin: Springerverlag,2009:560-577

(上接第 57 页)

另外,本文选择对分布式系统而言具有代表性的两种攻击方法,分析探测攻击和中间人攻击。

1)探测攻击

敌意方通过试探的方法来获取私密的策略或属性内容。由于本方案采取交互加解密措施,如果敌手随机提交请求,基于隐藏证书的协商流程,对方将无法猜测到我方的策略。由于我方在发送信息时已经采取了加密措施,即使敌手有非法获取的正确请求方式也无法解密,除非对方有正确的信任证书,这一点确保了敌方无法采取探测攻击。

2)中间人攻击

敌手可能通过证书伪造来发起中间人攻击。在本文的方案中,双方传输的敏感证书和策略都已经进行了基于属性的加密,只具有理论上进行破解的可能性。

结束语 本文提出了基于 ABE 的隐藏证书扩展模型,详细描述了扩展模型的体系架构、系统构造、双方信任协商过程、多方信任协商过程、性能分析、安全性分析和实际应用场景示例。在开放的环境如 Internet 中,与陌生方进行信任协商时,经常是基于请求方的属性特性,而不是请求方的身份。因此,相对基于 IBE 的隐藏证书而言,基于 ABE 的隐藏证书有着更为广泛的应用。此外,由于 ABE 技术能灵活地实现一对多的加密特性,因此基于 ABE 的隐藏证书不仅能用于双方信任协商,还能用于多方信任协商。由于在 ABE 体系中,加密及解密都是基于集合及阈值,因此发送方和接收方都可以保证自己的具体信息不被泄漏。而且,由于每个用户只有一个属性集证书,在各发证机构采用不同的随机多项式进行发布的情况下,共谋攻击无法实施,从而消除了很多隐患。

同时,我们也清楚地认识到,本文提出的基于 ABE 的隐藏证书扩展模型是单认证中心架构的,在实际应用中还有一定的局限性。在后续研究中,基于多认证中心的 ABE 隐藏证书技术将是研究的主要方向[10,11]。另外,鉴于现实世界中各类属性一般都是分层的,基于分层属性 ABE 的隐藏证书技术将是另外一个研究方向[12-14]。

参 考 文 献

[1] Winsborough W, Seamons K, Jones V. Automated Trust Negotiation[C]// Proceeding of DARPA Information Survivability Conference and Exposition. ACM Press,2000:156-182

[2] Winsborough W H, Jacobs J. Automated trust negotiation in attribute-based access control[C]//DARPA Information Survivability Conference and Exposition. Proceedings of Volume 2. April 2003:252-257

[3] Li N H, Winsborough W H, Mitchell J C. Distributed credential chain discovery in trust management[C]// Proc. of the 8th ACM Conf. on Computer and Communications Security. New York: ACM Press,2001:156-165

[4] Boneh D, Franklin M. Identity-based Encryption from Weil Pairing[A]// Kilian J CRYPTO 2001[C]. Berin: Springer-Verlag, 2001:213-229

[5] Holt J E, Bradshaw R, Seamons K E, et al. Hidden credentials [C]//Proceedings of 2nd ACM Workshop on Privacy in the Electronic Society. ACM Press,2003:1-8

[6] Bradshaw RW, Holt J E, Seamons K E. Concealing complex policies with hidden credentials[C]// ACM Conf. on Computer and Communications Security. New York: ACM Press, 2004: 146-157

[7] Frikken K, Atallah M, Li J T. Hidden access control policies with hidden credentials[C]// ACM Workshop on Privacy in the Electronic Society. New York: ACM Press,2004:27-28

[8] Sahai A, Waters B. Fuzzy Identity Based Encryption[C]// Advances in Cryptology-Eurocrypt. volume 3494 of LNCS. Springer,2005:457-473

[9] Goyal V, Pandey O, Sahai A, et al. Attribute-based Encryption for Fine-grained Access Conrol of Encrypted Data[C]// ACM Conference on Computer and Communications Security (ACM CCS). 2006

[10] Chase M. Multi-authority Attribute-based Encryption [C]// TCC. volume 4392 of LNCS. Springer,2007:515-534

[11] Chase M, Chow S S. Improving Privacy and Security in Multi-authority Attribute-based Encryption[C]// CCS '09: Proceedings of the 16th ACM Conference on Computer and Communications Security. New York, NY, USA: ACM,2009:121-130

[12] Li J, Wang Q, Wang C, et al. Enhancing attribute-based encryption with attribute hierarchy[C]// 4th International Conference on Communications and Networking in China (Chinacom). To appear ACM MONET,2009

[13] Gentry C, Silerberg A. Hierarchical ID-based Cryptography[A]// Zheng Y ASICCRYPT 2002[C]. Berlin: Springer-Verlag,2002: 548-566

[14] Horwitz J, Lynn B. Toward Hierarchical Identity-based Encryption[A]//Knudsen L EUROCRYPT 2002[C]. Berlin: Springer-Verlag,2002:466-481