



基于密文策略属性基加密系统访问机制的缓存替换策略

陈建¹, 沈潇军¹, 姚一杨¹, 邢雅菲², 琚小明^{2*}

(1. 国网浙江省电力公司信息通信分公司, 杭州 310007; 2. 华东师范大学 计算机科学与软件工程学院, 上海 200062)

(*通信作者电子邮箱 xmju@sei.ecnu.edu.cn)

摘要:为提高基于密文策略属性基加密(CP-ABE)系统的数据缓存性能,针对CP-ABE加密的数据,提出一种有效的缓存替换算法——最小属性价值(MAV)算法。该算法结合CP-ABE加密文件的访问策略并统计高频属性值的个数,利用余弦相似度方法和高频属性值统计表来计算属性相似度;同时结合属性相似度和文件大小计算缓存文件的属性值价值,并替换属性值价值最小的文件。在与最近最少使用(LRU)、最不经使用(LFU)、Size缓存替换算法的对比实验中,针对CP-ABE加密后的数据,MAV算法在提高加密文件请求命中率和字节命中率方面具有更好的性能。

关键词:属性策略;缓存替换策略;密文策略属性基加密算法;加密数据;余弦相似度

中图分类号:TP393 **文献标志码:**A

Cache replacement strategy based on access mechanism of ciphertext policy attribute based encryption

CHEN Jian¹, SHEN Xiaojun¹, YAO Yiyang¹, XING Yafei², JU Xiaoming^{2*}

(1. Information and Communication Branch, Zhejiang Electric Power Company, State Grid Corporation of China, Hangzhou Zhejiang 310007, China; 2. School Computer Science and Software Engineering, East China Normal University, Shanghai 200062, China)

Abstract: In order to improve the performance of cache for encrypted data based on Ciphertext Policy Attribute Based Encryption (CP-ABE), an effective replacement algorithm named Minimum Attribute Value (MAV) algorithm was proposed. Combining the access mechanism of ciphertext in CP-ABE and counting the number of high frequency attribute values, the attribute similarity was calculated by using cosine similarity method and the table of high frequency attribute values; meanwhile, the attribute value of each cache file was calculated according to the attribute similarity and size of the encrypted file, then the file with the minimum attribute value was replaced. The experimental results prove that the MAV algorithm has better performance in increasing byte hit rate and file request hit rate than the algorithms of Least-Recently-Used (LRU), Least-Frequently-Used (LFU) and Size for encrypted data based on CP-ABE.

Key words: attribute policy; cache replacement policy; Ciphertext Policy Attribute Based Encryption (CP-ABE) algorithm; encrypted data; cosine similarity

0 引言

缓存技术是提高云存储数据的存取速度的一种重要方法,缓存技术的核心内容是缓存替换策略。针对缓存问题,目前所提出的缓存替换策略主要有:1)传统经典的缓存替换算法,包括最近最少使用(Least-Recently-Used, LRU)、最不经使用(Least-Frequently-Used, LFU)替换算法,这些传统的替换算法实现简单,但未考虑时间间隔、对象大小等因素的影响;2)基于对象大小的缓存替换算法,包括Size和GD-Size(Greedy Dual-Size)算法^[1],这些算法充分考虑了对象大小的影响因素,但并未考虑缓存对象的访问频率的影响因素^[2]。针对传统缓存替换算法的不足,结合不同的缓存数据的特点,一些新的缓存替换算法被提出,主要包括:根据数据的位置相关性提出的位置相关查询中基于最小访问代价的缓存替换方

法^[3],该方法对于位置相关性好的数据有良好的效果,对于位置相关性差的数据的作用却并不明显;根据缓存文件的权重值提出的基于PageRank的缓存替换策略^[4],该算法优于LFU算法,但是具有单个缓存系统的局限性;针对Web缓存具有延迟的特点提出的基于预测的Web缓存替换策略^[5],该策略相对于传统替换算法而言提高了Web缓存的命中率和字节命中率,但是该预测代价太大,算法过于复杂。

目前提出的缓存替换算法针对的都是明文数据,还没有一种是针对加密数据进行的缓存替换策略。本文充分考虑基于CP-ABE加密的数据的特点,结合基于CP-ABE加密数据的特有的访问策略,提出针对一种基于CP-ABE加密的数据的缓存替换算法——最小属性价值(Minimum Attribute Value, MAV)算法。该算法是一种低开销、高性能和适应性的算法,它结合基于CP-ABE加密的文件特有的访问策略,通

收稿日期:2017-04-20;修回日期:2017-06-23。 基金项目:国家电网科技项目(5211XT160008)。

作者简介:陈建(1965—),男,福建福州人,教授级高级工程师,主要研究方向:信息安全; 沈潇军(1972—),男,浙江绍兴人,中级工程师,主要研究方向:信息系统管理; 姚一杨(1984—),男,浙江杭州人,高级工程师,硕士,主要研究方向:信息安全; 邢雅菲(1993—),女,山东淄博人,硕士研究生,主要研究方向:云计算分布式缓存; 琚小明(1967—),男,浙江衢州人,副教授,博士,CCF会员,主要研究方向:嵌入式系统软件、计算机软件、互联网。



过统计加密文件的访问策略中的属性值,建立属性值统计表,计算属性权重,利用计算文本相似度方法中的余弦相似度算法计算加密文件的属性值相似度,同时考虑文件的大小因素,计算缓存中文件的属性值价值,替换属性值价值最小的文件。

1 基于 CP-ABE 访问机制的缓存问题

1.1 CP-ABE 属性策略的研究基础

基于 CP-ABE 加密的文件,每个文件都有一个访问策略,文件的访问策略与属性值集密切相关。在上传文件时,要先设计好相应的访问策略,才能进行文件的加密^[7]。访问策略被建立成一棵访问结构树,树的叶节点都是相应的属性值。解密时,用户首先要上传自己的属性,系统将要解密用户的属性值与访问策略中的属性值相比较,属性值可以匹配,就可形成相应的私钥,从而可以解密文件^[6]。有关 CP-ABE 属性策略的相关定义如下:

定义 1 属性。设 $P = \{p_1, p_2, \dots, p_n\}$ 为所有属性的集合,则每个用户的属性 Att_i 是 P 的一个非空子集, $Att_i \subseteq \{p_1, p_2, \dots, p_n\}$, 那么 N 个属性可用于鉴别 $2N$ 个用户。

定义 2 访问结构。访问结构 Tr 是全集 $\{p_1, p_2, \dots, p_n\}$ 的一个非空子集, $Tr \subseteq 2^{\{p_1, p_2, \dots, p_n\}} \setminus \{\emptyset\}$ 。 Tr 表示一个属性判断条件,在 Tr 中属性集合称作授权集,不在 Tr 中属性集合称作非授权集。

定义 3 访问结构树。用于描述加密文件的访问结构,结构树每个叶节点表示一个属性项,每一个非叶子节点表示某个门关系,例如 AND、OR 等门限关系。

1.2 基于 CP-ABE 访问机制的缓存问题

给定有限的加密文件集 F , 每个加密文件 $f \in F$ 。假设缓存的加密文件的请求序列为 $Q = f_1 f_2 \dots f_m$, 缓存的总容量为 C , 其状态表示为 S , 假定对于任一个加密文件 f_i 小于缓存容量。

$$\max \{f_i\} \leq C$$

当请求序列为 $Q = f_1 f_2 \dots f_m$ 时, S_0 表示为缓存的初始状态, S_k 表示为 K 时刻缓存的状态, f_k 表示 k 时刻被访问的文件,

$$S_k = \begin{cases} S_{k-1}, & f_k \in S_{k-1} \\ S_{k-1} \cup \{f_k\}, & f_k \notin S_{k-1}, C_f \geq f_k \\ \{S_{k-1} - D_k\} \cup \{f_k\}, & f_k \notin S_{k-1}, C_f < f_k \end{cases} \quad (1)$$

其中: D_k 表示的是被替换出的数据对象; C_f 表示的是剩余的缓存内的空间大小。当出现式(1)中的第三种情况时,表示缓存空间已满,这时就需要利用本文所提出的基于 CP-ABE 访问策略的缓存替换算法进行替换。以下是对本文算法的具体描述。

2 基于 CP-ABE 访问机制的缓存替换策略

本文所提出的基于 CP-ABE 访问机制的缓存替换算法——最小属性价值算法(MAV),要对加密文件访问策略中的属性值进行统计,进而对文件属性值价值进行计算,在缓存替换中,替换文件属性值价值最小的文件。

2.1 加密文件访问策略中的属性值统计

MAV 算法要对缓存文件的访问策略中的属性值进行统计。建立一张属性表 T , 容量为 C , $Att_e = \{a_1, a_2, \dots, a_q\}$ 是表 T 中的属性集合,属性统计表 T 分为三列,分别记载:序号、属性值、属性值个数。表 T 中的属性值的排列顺序根据包含有该属

性值的文件的被访问时间,按照最近使用的属性值排在最前面的原则进行排列。

假设现在缓存中存在 A, B, C, D, E 共 5 个文件,其访问顺序为 B, D, A, C, E , 这 5 个文件的访问策略中的属性值分别为, A : 华师大、软件学院、嵌入式、学生; B : 华师大、软件学院、嵌入式、老师; C : 华师大、软件学院、密码学、学生; D : 华师大、教育学院、学前教育、学生; E : 华师大、体育学院、健美操、学生。根据属性值的访问次数和属性值的访问频率,以上五个文件的属性值统计如表 1 所示。

表 1 属性值统计

Tab. 1 Statistics of attribute values

| 序号 | 属性值 | 个数 | 序号 | 属性值 | 个数 |
|----|------|----|----|------|----|
| 1 | 华师大 | 5 | 7 | 嵌入式 | 2 |
| 2 | 体育学院 | 1 | 8 | 教育学院 | 1 |
| 3 | 健美操 | 1 | 9 | 学前教育 | 1 |
| 4 | 学生 | 4 | 10 | 老师 | 1 |
| 5 | 软件学院 | 3 | : | : | : |
| 6 | 密码学 | 1 | | | |

当缓存中有文件被替换出去时,被替换出去的文件属性值在表 T 中相应的个数要减少,并且当存在某个属性的个数被减少为 0 时,将该属性值从表 T 中删除;当缓存中有新的文件被替换进来时,新文件的属性值的个数在表 T 中要相应增加,并且表 T 中属性的排列顺序要根据新的访问时间进行相应的位置的变动。

在上面的例子中,若最后通过 MAV 算法将 D 文件替换出去, F 文件被替换进来,其中 D 文件的属性值为:华师大、教育学院、学前教育、学生, F 文件的属性值为:华师大、软件学院、海量所、老师根据上面的原则,此时属性值统计表 T 发生相应的变化,新的属性值统计如表 2 所示。

表 2 变化后的属性值统计

Tab. 2 Statistics of changed attribute values

| 序号 | 属性值 | 个数 | 序号 | 属性值 | 个数 |
|----|------|----|----|-----|----|
| 1 | 华师大 | 5 | 6 | 健美操 | 1 |
| 2 | 软件学院 | 4 | 7 | 学生 | 3 |
| 3 | 海量所 | 1 | 8 | 密码学 | 1 |
| 4 | 老师 | 2 | 9 | 嵌入式 | 2 |
| 5 | 体育学院 | 1 | : | : | : |

2.2 文件属性值价值计算

借鉴数据挖掘中常用的“文本相似度”的概念,计算加密文件的“属性相似度”,同时考虑缓存文件的大小因素,计算文件属性值价值(File Attribute Value, FAV),将属性值价值作为缓存中文件替换的标准。本文采用余弦相似度的方法计算加密文件的属性相似度,余弦相似度用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异的大小,余弦值越接近 1,就表明夹角越接近 0° ,也就是两个向量越相似^[7]。

在属性表 T 中,将表 T 看作为一种文本,表 T 中的属性值看作为关键词,将属性值的个数作为属性的权重,从而得出属性频向量 $\vec{X} = (x_1, x_2, \dots, x_n)$, 其中,每一个分量都对应属性表 T 中的一个属性 a_i , 分量值 x_i 为属性 a_i 的个数;同理,将单个加密文件看作为一种文本,加密文件访问策略中的属性值看作为关键词,计算属性值的个数作为属性的权重,得出另一个属性频向量 $\vec{Y} = (y_1, y_2, \dots, y_m)$, 其中,每个分量对应该文件访问策



略中的一个属性值 p_i ,分量值 y_i 同样为属性 p_i 的个数。这里要求向量 \tilde{X} 和向量 \tilde{Y} 中 $n = m$,但是在表 T 的属性集合 $Ate = \{a_1, a_2, \dots, a_q\}$ 和单个文件的属性集合 $P = \{p_1, p_2, \dots, p_k\}$ 中, $q \neq k$ 且 $q \gg k$ 。为了进行相似度计算,令 n 根据 m 的大小,在表 T 中从序号1到 q 依次对应,从而使得 $n = m$ 。根据余弦相似度计算公式,得出属性相似度计算公式为:

$$Sim(X, Y) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}} = \frac{a \cdot b}{\|a\| \times \|b\|} \quad (2)$$

根据余弦相似度的性质,计算出的属性相似度越接近1,就表明夹角越接近 0° ,也就是两个文本越相似。以2.1节例子中的文件 A 为例,由表 T 得出向量 $\tilde{T} = (5, 1, 1, 4)$,由文件 A 得出向量 $\tilde{A} = (5, 3, 2, 4)$,计算 $\tilde{A} = (5, 3, 2, 4)$ 与 $\tilde{T} = (5, 1, 1, 4)$ 的属性相似度为:

$$\cos(A) = \frac{5 \times 5 + 3 \times 1 + 2 \times 1 + 4 \times 4}{\sqrt{5^2 + 3^2 + 2^2 + 4^2} \times \sqrt{5^2 + 1^2 + 1^2 + 4^2}} = \frac{46}{\sqrt{43} \times \sqrt{54}} \approx 0.955 \quad (3)$$

得出的相似度接近1,表明文件 A 与表 T 具有较高的相似度。下面结合缓存文件的大小因素,描述计算文件属性值价值的具体过程。

设 FAV_i 为文件的属性值价值,初始值设为0, $P_i = \{p_1, p_2, \dots, p_k\}$ 为缓存文件 F_i 的属性值集合, $Size_i$ 表示文件 F_i 的大小,根据式(2),结合传统的Size算法的策略,得到 FAV_i 的计算公式如下:

$$FAV_i = \frac{\sum_{i=1}^k (x_i \times y_i)}{\sqrt{\sum_{i=1}^k (x_i)^2} \times \sqrt{\sum_{i=1}^k (y_i)^2} \times Size_i} = \frac{Sim(X, Y)}{Size_i} \quad (4)$$

缓存替换过程中,分别计算缓存中文件和即将被替换进来的新文件的属性值价值,替换 FAV_i 值最小的文件。

3 仿真实验与分析

本章给出MAV算法与传统经典的LRU算法、Size算法、LFU算法在不同条件下的比较结果。

3.1 实验环境

本次实验硬件使用的是Dell台式机,具体配置为:CPU Intel Core Duo i5 主频3.1 GHz;内存4 GB;硬盘1 TB。

软件环境采用的是:Windows 8操作系统;网络仿真软件OPNET Modeler10。

本次实验主要使用网络仿真软件OPNET Modeler10进行仿真模拟,利用C语言对MAV算法,以及传统的LRU算法、Size算法、LFU算法进行编写。实验的数据来源于浙江电力局提供的相关数据集。在实验中,设置总共可以访问的文件数目为1000个,文件大小在1 MB~1024 MB内随机选取,在测试中保证每种算法的总访问文件的次数为345686。

3.2 实验结果

为了进行对比,本实验在相同的实验环境下,与文献[9]中的传统经典的缓存替换算法LRU、LFU和Size的数据进行

对比。同时,对本文所提出的MAV算法进行实验,分别计算其在不同缓存大小下的加密文件请求命中率和字节命中率,以及针对不同文件大小的加密文件请求命中率和字节命中率。本文设置的缓存大小为:1%、3%、5%、10%、20%、30%、40%、50%,设置的文件大小为:32 MB、64 MB、128 MB、256 MB、512 MB、1024 MB。

图1为各种算法在不同缓存大小下的字节命中率。从图1可以看出,当缓存大小小于20%时,本文提出的MAV算法的字节命中率高出LRU算法和Size算法,略低于LFU算法;当缓存大小大于20%以后,MAV算法的字节命中率和加密文件命中率都高于其他三种算法。

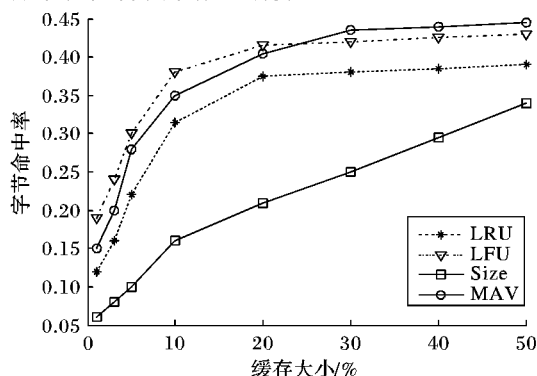


图1 缓存大小对字节命中率的影响
Fig. 1 Effect of cache size on byte hit rate

从图2所示的各种算法在不同缓存大小下的加密文件命中率可看出:当缓存大小小于16%时,MAV算法的加密文件命中率高于LRU算法和Size算法,略低于LFU算法;当缓存大小大于16%以后,MAV算法的加密文件命中率都高于其他三种算法。

从图3所示的各种算法在不同缓存文件大小下的字节命中率可看出:对于不同大小的加密文件,在字节命中率方面,

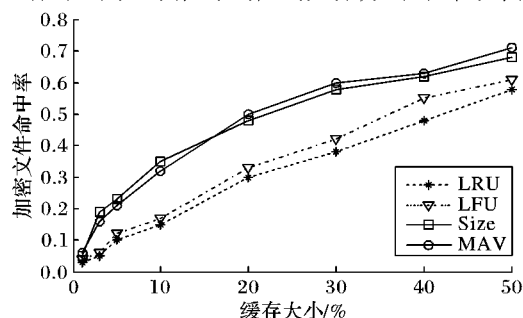


图2 缓存大小对加密文件命中率的影响
Fig. 2 Impact of cache size on the hit rate of encrypted files

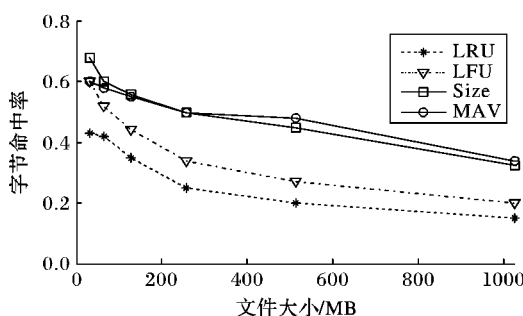


图3 文件大小对字节命中率的影响
Fig. 3 Effect of file size on byte hit rate



MAV 算法明显高于 LRU 算法和 LFU 算法,对于小文件 MAV 算法的命中率略低于 Size 算法,但对于大文件 MAV 算法具有明显的优势。

从图 4 所示的各种算法在不同缓存文件大小下的加密文件命中率可看出:对于不同大小的加密文件,在加密文件命中率方面,MAV 算法同样明显高于 LRU 和 LFU 算法,也同样高于 Size 算法,这样的优势对于大文件更加明显。

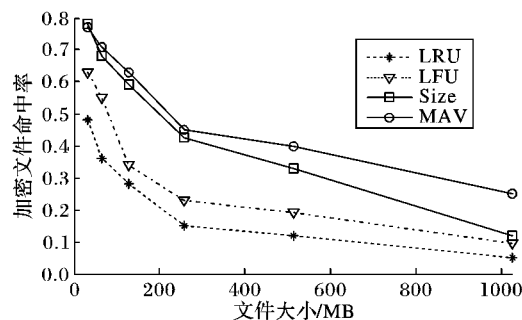


图 4 文件大小对加密文件命中率的影响

Fig. 4 Impact of file size on encrypted file hit rate

4 结语

本文提出的缓存替换算法,分析了基于 CP-ABE 加密的文件的特点,结合其访问策略中的属性值,充分考虑了文件的访问频率、访问时间问题,建立了属性值统计表,根据余弦相似度的计算方法计算文件属性相似度,并且综合 Size 算法中文件的大小因素,计算出文件的属性值价值。在实验部分,分别考虑了在不同缓存大小下的加密文件命中率和字节命中率,以及在不同文件大小下的加密文件命中率和字节命中率。实验数据表明,针对基于 CP-ABE 加密的数据,本文提出的 MAV 算法可以有效地提高缓存文件命中率和字节命中率。

本文所设计的缓存替换策略仅考虑了单个缓存的环境,今后的工作将集中在分布式缓存下加密数据的缓存策略的研究。

参考文献 (References)

- [1] 刘磊,熊小鹏.最小驻留价值缓存替换算法[J].计算机应用,2013,33(4):1018-1022. (LIU L, XIONG X P. Least cache value replacement algorithm [J]. Journal of Computer Applications, 2013, 33(4): 1018-1022.)
- [2] 秦秀磊,张文博,魏峻,等.云计算环境下分布式缓存技术的现状与挑战[J].软件学报,2013,24(1):50-66. (QIN X L, ZHANG W B, WEI J, et al. Progress and challenges of distributed caching techniques in cloud computing [J]. Journal of Software, 2013, 24(1): 50-66.)
- [3] 卢秉亮,梅义博,刘娜.位置相关查询中基于最小访问代价的缓存替换方法[J].计算机应用,2011,31(3):690-693. (LU B L, MEI Y B, LIU N. Cache replacement method based on lowest access cost for location dependent query [J]. Journal of Computer Applications, 2011, 31(3): 690-693.)
- [4] 肖敬伟,赵永祥.基于 PageRank 的缓存替换策略[J].信息技术,2016(6):107-110. (XIAO J W, ZHAO Y X. Cache replacement strategy based on PageRank [J]. Information Technology, 2016(6): 107-110.)
- [5] 石磊,孟彩霞,韩英杰.基于预测的 Web 缓存替换策略[J].计算机应用,2007,27(8):1842-1845. (SHI L, MENG C X, HAN Y J. Web replacement policy based on prediction [J]. Journal of Computer Applications, 2007, 27(8): 1842-1845.)
- [6] 程思嘉,张昌宏,潘帅卿,等.基于 CP-ABE 算法的云存储数据访问控制方案设计[J].信息安全,2016(2):1-6. (CHENG S J, ZHANG C H, PAN S Q, et al. Design on data access control scheme for cloud storage based on CP-ABE algorithm [J]. Netinfo Security, 2016(2): 1-6.)
- [7] 彭敏,黄佳佳,朱佳晖,等.基于频繁项集的海量短文本聚类与主题抽取[J].计算机研究与发展,2015,52(9):1941-1953. (PENG M, HUANG J J, ZHU J H, et al. Mass of short texts clustering and topic extraction based on frequent itemsets [J]. Journal of Computer Research and Development, 2015, 52(9): 1941-1953.)
- [8] 肖敬伟.基于数据挖掘的缓存替换算法研究[D].北京:北京交通大学,2015:1-66. (XIAO J W. Cache replacement algorithms based on data mining [D]. Beijing: Beijing Jiaotong University, 2015: 1-66.)
- [9] 陈勇.基于 P2P 的内容分发网络及缓存替换算法研究[D].成都:电子科技大学,2008:1-79. (CHEN Y. Research on P2P-based content distribution network and cache replacement algorithm [D]. Chengdu: University of Electronic Science and Technology of China, 2008: 1-79.)
- [10] GOMAA W H, FAHMY A A. A survey of text similarity approaches [J]. International Journal of Computer Applications, 2013, 68(13): 13-18.
- [11] HARATY R A, ZEITOUNY J. Rule-based data mining cache replacement strategy [J]. International Journal of Data Warehousing & Mining, 2015, 9(1): 56-69.
- [12] DAS S, AAMODT T M, DALLY W J. Reuse distance-based probabilistic cache replacement [J]. ACM Transactions on Architecture & Code Optimization, 2016, 12(4): 33.
- [13] GAST N, VAN HOUTD B. Transient and steady-state regime of a family of list-based cache replacement algorithms [C]// SIGMETRICS 2015: Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems. New York: ACM, 2015: 123-136.
- [14] LEE M K, MICHAUD P, SIM J S, et al. A simple proof of optimality for the MIN cache replacement policy [J]. Information Processing Letters, 2015, 116(2): 168-170.
- [15] LI L, CHEN X, JIANG H, et al. P-CP-ABE: parallelizing ciphertext-policy attribute-based encryption for clouds [C]// Proceedings of the 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. Piscataway, NJ: IEEE, 2016: 575-580.
- [16] ZOU J, ZHANG Y. Research of the application of CP-ABE algorithm in ABAC model under cloud environment [J]. Journal of Computational Information Systems, 2015, 11(1): 261-268.

This work is partially supported by the National Grid Technology Project (5211XT160008).

CHEN Jian, born in 1965, senior engineer. His research interests include information security.

SHEN Xiaojun, born in 1972, engineer. His research include information system management.

YAO Yiyang, born in 1984, M. S., senior engineer. His research interests include information security.

XING Yafei, born in 1993, M. S. candidate. Her research interests include cloud computing distributed cache.

JU Xiaoming, born in 1967, Ph. D, associate professor. His research interests include embedded system software, computer software, Internet.