

基于 Merkle 哈希树的电力扰动数据 安全去重方法

李世明^{1†}, 卢建刚¹, 余志文¹, 郭文鑫¹, 汤健东²

(1. 广东电网有限责任公司电力调度控制中心, 广东 广州 510000; 2. 国电南瑞南京控制系统有限公司, 江苏 南京 210000)

摘 要:针对当前去重方法存在伪基和伪偏移量问题, 导致去重效果不佳, 提出了基于 Merkle 哈希树的电力扰动数据安全去重方法。使用收敛加密方法, 有效解决数据加密和去重矛盾, 应用 Merkle 哈希树得到电力扰动数据块指纹, 以此标记待去重电力扰动数据。利用 POW 协议证明拥有检测重复数据块。使用扰动状态似然代替对数似然, 避免电力异常扰动状态在去重标识过程中重复, 实现安全去重。由实验结果可知, 该方法最大基提取为 3000 B、最大偏移量为 6000 B, 均未超过实验设定阈值, 说明所提方法去重效果较好。

关键词: Merkle 哈希树; 电力扰动; 数据安全去重; 概率误差

中图分类号: TP309

文献标识码: A

Secure Deduplication Method for Power Disturbance Data Based Merkle Hash Tree

LI Shiming^{1†}, LU Jiangang¹, YU Zhiwen¹, GUO Wenxin¹, TANG Jiandong²

(1. Electric Power Dispatching Control Center of Guangdong Grid Co., Ltd., Guangzhou, Guangdong 510000, China;

2. Nari Technology Nanjing Control Systems Co., Ltd., Nanjing, Jiangsu 210000, China)

Abstract: In view of the problems of pseudo-base and pseudo-offset in the current de-weight method, the effect of de-weight is not good. This paper presents a safe method for removing power disturbance data based on Merkle hash tree. Convergence encryption method is used to effectively solve the contradiction between data encryption and deduplication. Merkle hash tree is used to obtain power disturbance data block fingerprint, which marks the power disturbance data to be de-duplicated. Duplicate data blocks are detected using POW protocol to prove ownership. The disturbance state likelihood is used instead of logarithmic likelihood to avoid repeating abnormal disturbance state in the process of deregistration and achieve safe deregistration. The experimental results show that the maximum base extraction of the method is 3000 B and the maximum offset is 6000 B, both of which do not exceed the threshold set by the experiment, indicating that the proposed method has a good effect of weight removal.

Key words: Merkle Hash tree; power disturbance; data security deduplication; probability error

在早期阶段, 电力干扰信号因其对电网的作用而备受关注, 且常伴有一定的畸变, 故被视为一种非侵入式的监测方法。随着国内外对其研究的不

断深入, 如何准确地检测出其对设备工作状态的影响就显得尤为重要。针对电力设备运行过程中大量存在的干扰信号, 结合其特征差异性, 研究具有

更高普适性的干扰信号辨识方法,可提高电力设备运行状态辨识的准确率。文献[1]提出了基于双层加密和密钥共享的去重方法,该方法利用聚合性质与对称性质,设计一类新的双重加密算法,用于更高安全级别的无效信息加密。采用门限秘密共享机制,将对外密码的加密密钥存储在多个密码密钥管理服务中,无须第三方参与,在安全密钥下实现电力扰动数据安全去重。然而,使用该方法虽然简单,但是无法实时检测短时间内的干扰;文献[2]提出了基于区块链的去重方法,该方法利用边界运算对电力数据进行预处理,通过构建一个安全存储链接以及一个可删除的验证链接,实现了扰动数据安全去重。然而,受实际环境噪声影响,其概率分布与高斯分布存有较大差异,且受电网动力扰动特性影响,在某种程度上影响了概率分布的测距,影响了该方法的准确度。为减少去重标记和密文运算代价,并在存取控制条件下有效实现数据去重,采用 Merkle 哈希树对电力扰动下的数据安全去重。该方法在 Merkle 哈希树结构下标记扰动数据,在云存储环境下搭建安全去重环境,利用 POW 协议证明实现数据安全去重。

1 基于 Merkle 哈希收敛加密的电力扰动数据去重标记

因为数据块的密码是从数据块中提取出来的,所以如果数据块处于一定的可选择区域,那么密码就会被泄露,敌人就可以在离线的情況下,用暴力攻破。如果采用常规的密码机制,则会因为密码不能被确认而被强行破译,而且需要进行冗余密码哈希运算,增加了运算代价。为此,研究 Merkle 哈希的收敛加密算法,解决现有密码算法中因密码不明确而带来的强力攻击者难以攻破的难题,降低密码算法中的冗余代价,从而提升电力扰动数据去重标记安全性。

1.1 收敛加密数据块标识产生

收敛加密算法能够保证在电力扰动数据去重时对每一组数据进行正确辨识,使相同数据块具有相同标识^[3]。构成收敛加密的数据块标识产生过程,如下所示:

$$Keyc(M) \rightarrow \alpha \tag{1}$$

式(1)中, M 表示以电力扰动数据块组成的明文; α 表示收敛密钥^[4]。

构成收敛密钥的加密密文产生过程如公式(2)所示:

$$X(\alpha, M) \rightarrow Y \tag{2}$$

构成输出加密密文的产生过程如公式(3)所示:

$$Z(\alpha, Y) \rightarrow M \tag{3}$$

电力扰动数据块标识的产生,以明文 M 为输入 X 参量,输出标识的过程如公式(4)所示:

$$T(M) \rightarrow R_M \tag{4}$$

通过上述公式,可得到电力扰动数据块标识。

1.2 电力扰动数据 Merkle 哈希去重标记

应用 Merkle 哈希树的收敛加密策略,首先通过上述步骤计算电力扰动数据块哈希值用于生成密钥,然后再次使用 Merkle 哈希树得到电力扰动数据块指纹,最后通过可信第三方处理密钥加密非重复电力扰动数据块,并表示待去重数据^[5]。

用如下公式表示 Merkle 哈希树的数据块指纹:

$$H_a = Hash(M) \tag{5}$$

公式(5)中, $Hash$ 为哈希函数,选用 SHA-1 或 SHA-256 等^[6]。标记过程与上传数据块表示去重同时进行:

首先利用数据块标识构建 Merkle 哈希树,如图 1 所示。

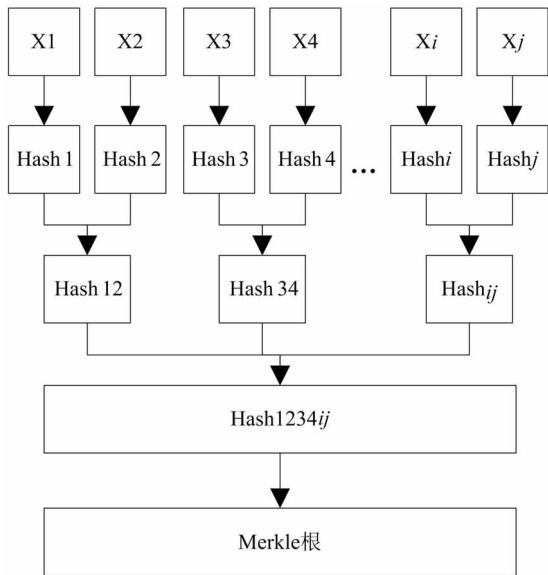


图 1 Merkle 哈希树结构示意图

然后构建一棵覆盖整个网络数据的哈希树,以达到高效地检测数据集的目的,在哈希函数中,根据单向特性对其内包含的子节点和它们需要的数据进行实时检测^[7-8]。

之后利用数据块私钥和标识对公钥进行一次循环的异或运算,得到数据块收敛密钥,利用该密钥对数据加密,从而产生数据密文^[9]。

最后利用 Merkle 哈希树的根节点对区块上的收敛密钥进行加密,即第一区块上的私有密钥对下一区块上的收敛密钥进行加密,就像 block 1 的私有密钥对 block 2 的收敛密钥进行加密一样^[10]。用户仅在 block 1 中保存收敛密钥的密文,并将该密钥及其他区块的密钥全部上传到云存储服务^[11]。通过云存储服务的数据区来建立 Merkle 哈希树,然后将这个根节点传回至用户,由此生成一个数据区块标识:

$$T(\text{block } 1) = \text{Hash}_{\text{SHA-256}}(\text{block } 1) \quad (6)$$

用户将式(6)所识别的数据块传送至云储存系统,为数据安全去重提供标识数据。

2 基于 Merkle 哈希标记的数据块安全去重

由于电力扰动数据分散,大量存取操作导致云存储效率低下。为此,借助云存储搭建并维护数据存储结构,为数据块重复检测提供方便网络。在安全网络中,对一个数据进行了检测,并且仅有一个副本被保存下来,其他的空间被保存为一个指针的数据结构。

指针数据结构指的是将去重数据块的收敛密钥加密文件上传与解密下载,并将其存储于云存储服务提供商实体中,形成指针数据结构。为了保证该结构安全性,使用流行度转换方法检验数据完整性,满足现代去重安全性。

2.1 云存储服务提供商实体检测重复数据块

用户将电力扰动数据文件分成 n 块数据块 $\{\text{block}_n\}$,将 n 块数据块标识 $T(\text{block } n)$ 发送给私有云,并计算数据块标识集:

$$f_n = \{T_1(T_0(\text{block } n), k_\varphi)\} \quad (7)$$

公式(7)中, k_φ 标识权限密钥,其中 φ 满足三元关系式:

$$t_\varphi = \begin{cases} 1 & \text{权限有效} \rightarrow \varphi \\ 0 & \text{权限失效} \rightarrow \varphi' \end{cases} \quad (8)$$

由私有云向用户传送标记集合,用户将标记集合上载到云存储实体,并进行数据块重复检测,设计的检测过程如图 2 所示。

在检测时,通过云存储实体检测重复数据块^[12]。如果重复,则利用如下 POW 协议证明所有权:

$$P | \text{Pow}_M | I(M) | \leq \epsilon \quad (9)$$

公式(9)中, $I(M)$ 标识部分与电力扰动原始数据相关的信息; ϵ 标识概率误差阈值^[13]。

云存储服务提供商实体在获取用户对数据的

授权验证结果后,通过向哈希树的根节点发送密文,并向该节点发送聚合密钥^[14]。

在加密数据块时,用户将一个汇集密钥的密文存储起来^[15]。如果没有得到有效授权,云存储服务提供商实体将停止上传操作并拒绝上传。

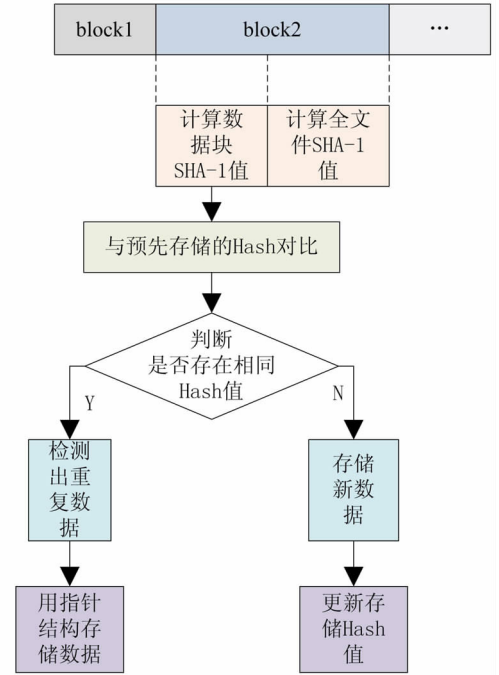


图 2 基于云存储服务提供商实体的数据块重复检测

如果不重复,私有密钥是利用 SHA-1 哈希 block 1 来生成私钥:

$$K_{\text{SHA-1}}(\text{block } 1) = \text{Hash}_{\text{SHA-1}}(\text{block } 1) \quad (10)$$

使用周期异或组合来与标识数据块和私有密钥相关联,构造一个收敛密钥:

$$\alpha(\text{block } 1) = T(\text{block } 1) \oplus K_{\text{SHA-1}}(\text{block } 1) \quad (11)$$

将数据去重检测结果作为输入数据,进行数据块收敛密钥加密文件上传与解密下载。

2.2 基于去重数据的指针结构存储

通过云存储服务提供商实体对重复数据块检测结果,使用收敛密钥对数据块加密后,设计如图 3 所示的文件加密上传流程,为得到一个指针数据结构提供基础文件。

由私有云通过加密的方式向用户发送收敛密钥加密密文,用户仅需要将该密文包上载至云端,就可以得到相应的数据块指针。

用户在获取数据时,必须先从多个云存储服务提供商实体上下载数据区块的密文及收敛密钥的密文。通过对用户权限认证,使用户能够进入相应的云存储服务提供商实体,获取数据块及汇合密

钥。按照如下方式,对数据文件解密:

用户首先利用 Merkle 哈希树的根节点译码 block 1 密钥,并把该密钥与整体密钥分开。然后通过私有密钥来译码数据块 block 2 的收敛密钥,并对数据块 block 2 的私有密钥进行分解,对数据块 block 3 的收敛性密文进行解码,依次类推,直至得到全部收敛密钥。最后采用一个收敛密钥对相应数据块进行加密,得到指针数据结构。

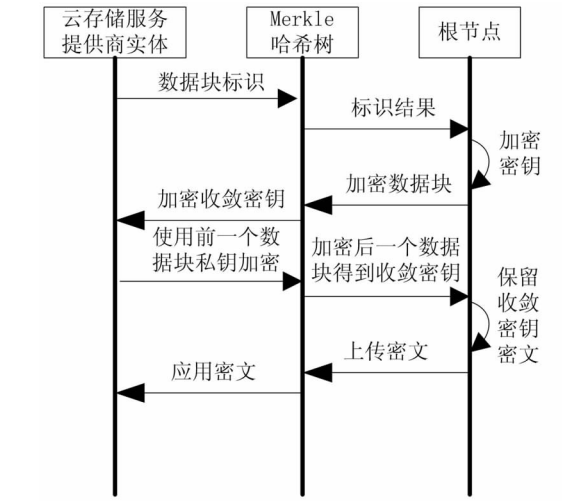


图 3 数据块收敛密钥加密文件上传流程

2.3 基于流行度转换的去重数据完整性校验

通过对数据块收敛密钥加密文件上传与下载,可实现数据去重。由于去重后的数据容易存在伪数据,所以必须对去重数据进行数据完整性校验,以防止用户恶意上传不实数据,从而造成去重数据的不完整。

在数据完整性校验时,使用扰动状态似然代替对数似然,避免电力异常扰动噪声状态在去重标识过程中重复。设定 m 扰动类别的去重标签标识标准:

$$T_n = \arg \min \sum_{i=1}^l |\ln f(G | \xi_i) - \ln f(G | \xi_n)| \tag{12}$$

公式(12)中, $\ln f(G | \xi_i)$ 表示 i 个测试数据 ξ_i 扰动状态似然; $\ln f(G | \xi_n)$ 表示 n 个测试数据 ξ_n 扰动状态似然; l 表示小波域数量。

通过去重标签标识标准,对如下三种异常扰动类型进行标识:

- (1)对快速电容器切换标识:电容器快速切换使得电力在小波分析范围内存在干扰似然脉宽。
- (2)对常规电容器切换标识:常规电容器使得电路除了产生瞬时高清晰度小波域外,还产生了持续扰动的高清晰度小波域。

(3)对功率电子器件陷波扰动标识:在高分辨率的小波域中有很大的干扰,但在很短时间里,电容切换会在很大区域中发生。

识别出三类不同类别干扰信息后,云存储服务提供商实体能够在流行度转换过程中,对其标志进行一致性检验,从而实现对用户所提交信息的真实性有效确认。在不依赖第三方服务器的情况下,用户可以将会聚密文上载至云存储服务提供商实体。

云存储服务提供商实体接收到文件后,验证标签一致性,如果:

$$T(F) = Tc(M_F^1) \tag{13}$$

公式(13)中, M_F^1 标识文件 F 的密文,则用 M_F^1 替换已经存在文件对应的密文,并将拥有权证明给用户。利用新生成的密文与已保存的密文进行比对,实现了对已上传信息的完整性确认。

3 实验

3.1 电力扰动数据集产生

电力设备在实际运行过程中,会产生如图 4 所示的扰动数据集。

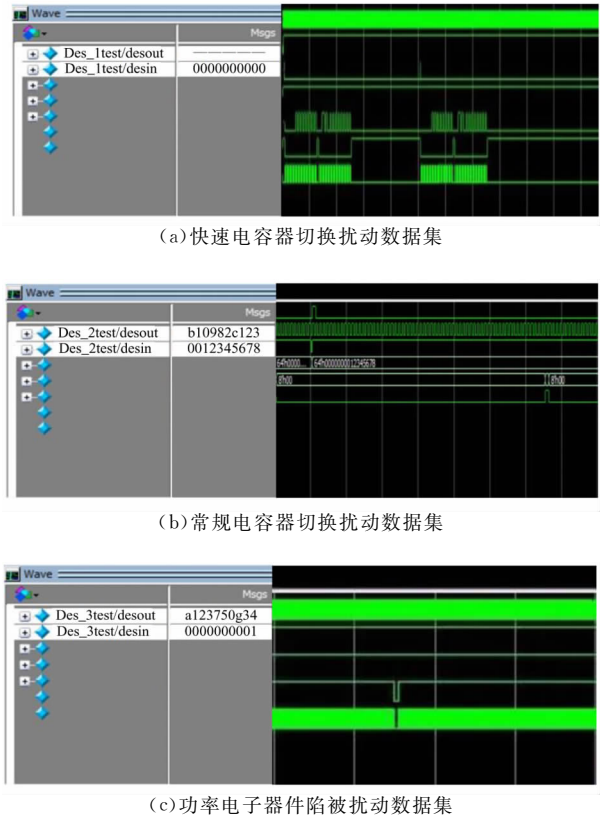


图 4 电力设备实际运行扰动数据集

由图 4 可知,该数据集由以下三种数据集

组成:

(1)快速电容器切换扰动数据集

一般情况下,当一台电容器从运行状态中脱离出来后,就必须等到它自己释放出来,否则就无法使用。然而,在要求功率电容器迅速投入运行时,将会引起电容失效,而在这种情况下,所生成的功率干扰数据就是功率负载迅速改变的数据集,即产生扰动功率数据集。

(2)常规电容器切换扰动数据集

在传统电容器切换过程中,存在着以等容量高电容替代低电容而无约束条件的情形,这种情形将导致输出电压值存在很大误差,即产生扰动电压数据集。

(3)功率电子器件陷波扰动数据集

陷波是指当功率电子器件在正常运行时,由于交流电源的转换而引起的一种周期性的电流波动现象,即产生扰动电流数据集。

3.2 实验数据分析

将数据提取基和提取偏移量作为去重指标,从上述三个数据集中随机抽选 50 个数据,通过图 5 展示每个数据提取基和偏移量。

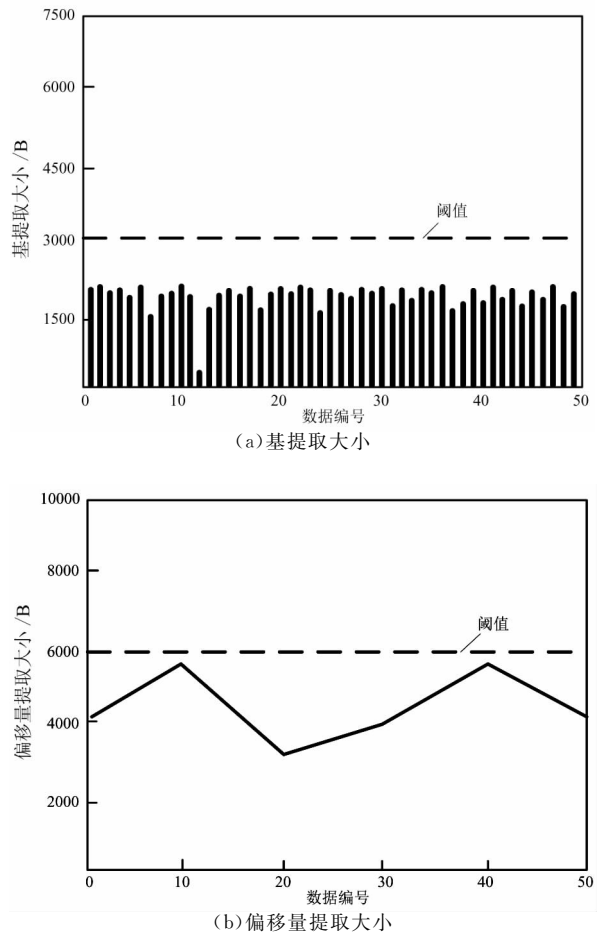


图 5 实验数据分析结果

图 5(a)、(b)中设置的阈值表示 50 个数据中均无重复数据需要处理,其中基提取阈值为 3000 B,偏移量提取阈值为 6000 B。

3.3 实验结果与分析

在实验过程中,比较 Merkle 哈希树、双层加密和密钥共享、区块链方法的扰动数据去重性能。

对于基提取指标,使用三种方法进行对比分析,结果如图 6 所示。

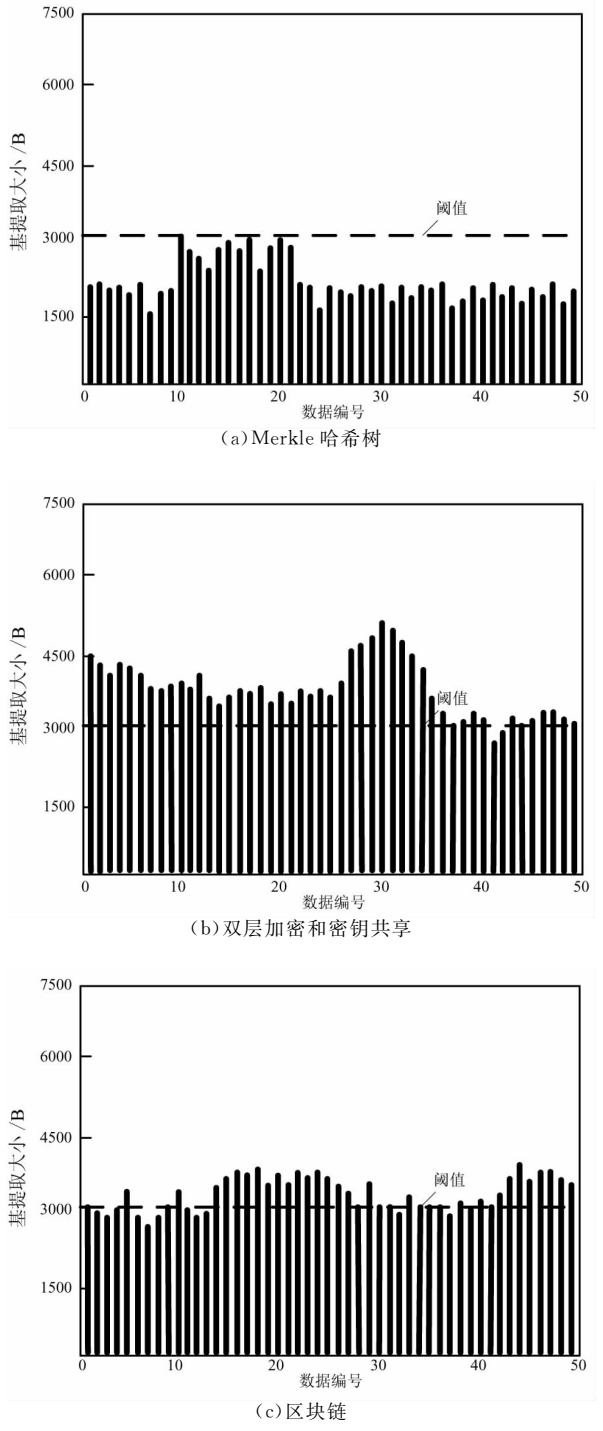


图 6 不同方法基提取大小

由图 6 可知,使用双层加密和密钥共享、区块链方法基提取均超过阈值,说明使用这两种方法没有先保存索引结果,出现了大量重复计算步骤,其对应的最大基提取分别为 5250 B、3800 B。使用基于 Merkle 哈希树的去重方法基提取在数据编号 10~20 时出现了与实际不一致的数据,但未超过阈值,说明使用该方法通过 Merkle 哈希树避免了重复大量计算,其对应的最大基提取为 3000 B。

对于偏移量提取指标,使用三种方法进行对比分析,结果如图 7 所示。

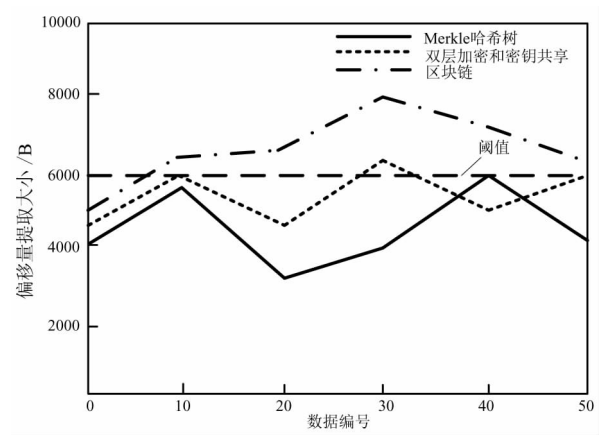


图 7 不同方法偏移量提取大小

由图 7 可知,使用双层加密和密钥共享方法偏移量提取时,部分提取量超过阈值,最大偏移量为 6500 B;使用基于区块链的去重方法偏移量提取时,绝大部分提取量超过阈值,最大偏移量为 8200 B;使用基于 Merkle 哈希树的去重方法偏移量提取时,所有提取量均未超过阈值,最大偏移量为 6000 B。

4 结 论

提出了基于 Merkle 哈希树的电力扰动数据安全去重方法,能够有效解决传统方法存在的去重效果不佳的问题,并得到如下结论:

- (1)利用 Merkle 哈希树对用户分享聚合密钥进行加密,保证密钥具有一定安全性。
- (2)通过云存储服务提供商,实现数据去重网络安全。
- (3)通过使用流行转换法检验去重数据完整性,实现电力扰动数据安全完整去重。

参考文献

- [1] 高文静,咸鹤群,程润辉. 基于双层加密和密钥共享的云数据去重方法[J]. 计算机学报,2021,44(11):2203—2215.
- [2] 张利华,曹宇,张赣哲,等. 基于区块链的微电网数据安全存储与删除验证方案[J]. 计算机工程与设计,2023,44(4):967—976.
- [3] 欧家祥,胡厚鹏,何沛林,等. 基于国密算法的电网 AMI 数据安全防护技术[J]. 西南师范大学学报(自然科学版),2023,48(3):17—24.
- [4] 哈冠雄,贾巧雯,陈杭,等. 无第三方服务器的基于数据流行度的加密去重方案[J]. 通信学报,2022,43(8):17—29.
- [5] 朱正春,杨知方,余娟,等. 面向小样本场景的数据驱动安全约束经济调度快速计算方法[J]. 中国电机工程学报,2022,42(12):4430—4440.
- [6] 王高洲,王惠剑,王聪,等. 基于 SM2 密码算法的电力数据安全接入方法[J]. 南京理工大学学报,2022,46(6):749—755.
- [7] 王梓,王治华,韩勇,等. 面向电力系统网络安全的多层协同防御模型研究[J]. 计算机工程,2021,47(12):131—140.
- [8] 杨挺,蔡绍堂,闫鹏,等. 面向电力物联网异构信息安全的多任务调度算法[J]. 电力系统自动化,2022,46(15):162—170.
- [9] 于海,李峰,霍英哲,等. 电力信息安全态势评估方法[J]. 科学技术与工程,2021,21(9):3642—3648.
- [10] 袁文勇,李秀广,李瑞峰,等. 基于无双线性对的可信云数据完整性验证方案[J]. 计算机应用,2022,42(12):3769—3774.
- [11] 王俊洪,张华赢,胡文曦,等. 电能质量扰动数据可用性评估与可视化软件设计[J]. 电网技术,2022,46(3):1109—1116.
- [12] 楚恬歆,张文海,瞿科,等. 小电流接地系统接地型早期故障扰动特征分析[J]. 电力系统保护与控制,2021,49(18):52—61.
- [13] 张涵,王程,毕天姝. 融合物理与数据知识的电力系统扰动后频率在线快速计算方法[J]. 电网技术,2022,46(11):4325—4335.
- [14] 李子康,刘灏,毕天姝,等. 基于数据驱动的鲁棒性电力系统扰动识别[J]. 中国电机工程学报,2021,41(21):7261—7275.
- [15] 苏寅生,李智勇,刘春晓,等. 基于小波—隐马尔可夫的波形异常扰动类型识别研究[J]. 电网与清洁能源,2021,37(4):53—59.