DOI:10. 20033/j. 1003-7241. (2026)03-0167-04

一种基于用户隐私的分布式多数据中心分级储存方法

唐远富[1], 陈远扬[1], 谢耀恒[2]

(1. 国网湖南省电力有限公司,湖南 长沙 410000; 2. 国网湖南省电力有限公司电力科学研究院,湖南 长沙 410000)

摘　要:为提升分布式存储的数据完整性和隐私保护能力,提出一种基于用户隐私的分布式多数据中心分级储存方法。通过构建最大权值节点机制与二叉树结构,动态划分负载节点,实现数据均衡分配与高效读取。以最低负载为分类单位,引入数据中心分级存储流程,利用闭合序列模式动态检测与调控负载均衡。结合敏感数据标记与权限控制,设计基于空间向量的索引密钥机制,以 BINDER 层级模型为基础,配置分布式多数据的层次结构,实现多层级数据封装与安全传输。实验结果表明所提方法可在 10 s 内完成较高负载的数据存储,且在 10 次攻击过程中,数据的保留完整度保持在 90%以上。实验验证了该方法在复杂网络环境中的可行性与优越性,为大规模数据的安全存储提供了理论依据与技术路径。

关键词:用户隐私;分布式数据;中心分级;储存方法;负载均衡;索引密匙;层次模型

中图分类号: TP393. 08　　**文献标志码**: A　　**文章编号**: 1003-7241(2026)03-0167-04

Hierarchical storage method for distributed multi data centers based on user privacy

TANG Yuanfu[1], CHEN Yuanyang[2], XIE Yaoheng[1]

(1. State Grid Hunan Electric Power Corporation Limited, Changsha 410000, Hunan, China;
2. State Grid Hunan Electric Power Corporation Limited Research Institute, Changsha 410000, Hunan, China)

Abstract: To enhance data integrity and privacy preservation capabilities of distributed storage, this paper proposes a hierarchical storage method for distributed multi-data centers based on user-privacy. By constructing a maximum weight node mechanism combined with a binary tree structure, the proposed approach dynamically partitions load nodes to achieve balanced data distribution and efficient access. Taking the minimum load as the classification unit, a hierarchical storage workflow is introduced to dynamically monitor and regulate load balancing using closed sequence patterns. Furthermore, by integrating sensitive data labeling with access control, an index key mechanism based on spatial vectors is designed. Grounded in the BINDER hierarchical model, the method configures the hierarchical architecture of distributed data, thereby realizing multi-level data encapsulation and secure transmission. Experimental results demonstrate that the proposed method can accomplish high-load data storage within 10 seconds and maintain a data retention integrity of over 90% even after 10 consecutive attacks. The feasibility and superiority of the scheme in complex network environments are validated, providing both theoretical foundations and technical pathways for the secure storage of large-scale data.

Keywords: user privacy; distributed data; center classification; storage method; load balancing; index key; hierarchical model

不同的网络平台中,用户浏览数据呈现不间断发展趋势,因此在网络中产生了恶意软件,以资源消耗和隐私窃取为主。由于网络中存有大量的隐私性个人信息,对用户来说,数据的安全和私密性更加重要,在网络平台中需要有一个强大的保护用户隐私的存储方法。传统的集中式存储在数据爆发式增长的背景下暴露出扩展性差、运营成本高、系统间连通困难等问题,基于此,分布式存储模式应运而生,可对数据进行高效管理,提升隐私数据的防护效果。目前分布式存储系统已被衍生出了多个种类,如 Google 在 2003 年发布了 GFS 系统用户存储海量检索数据,这是分布式存储中真正部署应用的较大项目之一。而后,还衍生处理超算中的 Lustre、云平台中广泛应用的 Ceph 等。该种分布式存储方式在一定程度上实现了隐私数据的高效管理,但还存在一定不足,故国内外学者纷纷开始研究,寻找更好的信息存储改进方案。文献[1]中提出了开源反汇编存储程序,在静态分析过程中,对数据进行非信任检查,对于敏感信息给定访问权限,监测数据流转中是否存在危险特征,在此基础上对数据实现存储,但静态检测只适用于已知的数据信息,只能保证提供给用户的数据信息是安全的,无法实时保证用户的隐私数据存储。文献[2]中展开了动态分析技术的应用,对数据传输过程中的非信息行为进行拦截,对存在敏感源的数据进行

收稿日期:2025-11-23;**录用日期**:2025-12-05

基金项目:国网湖南省电力项目(SGHN0000HWJS1900586)

作者简介:唐远富(1984—),男,博士,高级工程师,研究方向:数据资产运营管理。

引用本文:唐远富, 陈远扬, 谢耀恒. 基于用户隐私的分布式多数据中心分级储存方法[J]. 自动化技术与应用, 2026,45(3):167-170. (TANG Yuanfu, CHEN Yuanyang, XIE Yaoheng. Hierarchical storage method for distributed multi data centers based on user privacy[J]. Techniques of Automation and Applications, 2026,45(3):167-170.)

标记,实现敏感数据的动态追踪存储,但也只能在发生隐私泄露时给用户提出提醒,难以满足安全存储要求。本文在考虑用户隐私的前提下,设计分布式多数据中心分级存储方法,为保证用户的隐私安全提供理论支持。

1 数据存储负载节点划分

以读取速度和负载平衡为基础,均衡负载对海量数据的存储起着重要影响,在一定程度上发挥关键作用。通过最大权值节点的圈定,对协同存储中的数据分级结构进行划分。数据的存储最终目的是数据应用,即在数据库中的信息,能够具有独立的存储空间外,还能够在网络中实现自由流转[3]。通过递归的数据结构,在高并发数据中,以最优节点作为均衡负载节点,将多个处理器中的数据进行整合,以负载节点分配对应数据存储,具有实现不同处理器的数据传输的存储。

将负载均衡代入数据分配任务,找出负载较轻的数据节点,为后续的数据存储提供最优策略[4]。通过二叉树作为约束条件,设置 Z 为数据节点, left$[Z]$ 表示 Z 的左孩子, right$[Z]$ 表示 Z 的右孩子, size$[Z]$ 为以 Z 为根节点的节点量,则满足特性

$$\begin{aligned}\text{size}[\text{right}[Z]] \geqslant \max\{&\text{size}[\text{left}[\text{left}[Z]]],\\ &\text{size}[\text{right}[\text{left}[Z]]]\}\end{aligned} \tag{1}$$

$$\begin{aligned}\text{size}[\text{left}[Z]] \geqslant \max\{&\text{size}[\text{right}[\text{right}[Z]]],\\ &\text{size}[\text{left}[\text{right}[Z]]]\}\end{aligned} \tag{2}$$

式中, size$[Z]$ 对负载的排序较为重要,是获取负载最优值的关键,设定式(1)和(2)的性质为对称状态,则可以划分为

$$\text{size}[\text{left}[\text{left}[Z]]] > \text{size}[\text{right}[Z]] \tag{3}$$

$$\text{size}[\text{right}[\text{left}[Z]]] > \text{size}[\text{right}[Z]] \tag{4}$$

式中表示 Z 的右孩子和左孩子,均可满足特性,能够实现节点的左右旋转,对数据信息匹配存储结构[5]。式(4)表示只对数据进行右旋转,即在不满足左旋的前提下,对执行维护操作,使其保持有最高权值节点,实现数据存储结构的对应匹配[6]。

2 设定数据中心分级存储流程

相较于普通存储流程,分布式数据存储流程更复杂,以多结构存储类型进行数据分级后,会存在多个数据空间,而空间内对应负载不一致,会影响整体数据的存储效率。为避免某些区块内的负载过高或者过低,影响整体数据的存储效率,需通过检测模块获取数据负载[7]。在分布式网络环境中,对数据进行区块划分,可将其看作为序列挖掘过程,设定在互联网中存在有 X 独立节点,均存在于无共享的存储结构中[8]。将主站点设置为 C_0,其余站点作为从站点 $C_V(V=1,2,\cdots,X-1)$, 其序列关系为 NM_V, 所有从站点上的序列数据库表示为 $B=\text{NM}_1 \cup \text{NM}_2 \cup \cdots \cup \text{NM}_{V-1}$。通过闭合序列模式对数据存储流程进行概念描述,则存在有

$$\sum_{V=1}^{X-1} A.\sup_V < \sum_{V=1}^{X-1} \min - \sup D_V \tag{5}$$

$$\sum_{V=1}^{X-1} A.\sup = A.\sup \tag{6}$$

$$\sum_{V=1}^{X-1} \min - \sup D_V = \sum_{V=1}^{X-1} \min - \sup \times |\ \text{NM}_V\ | \tag{7}$$

式中, A 表示站点中的局部序列模式。在分布式环境中,站点 C_V 含有 A 的元数据总数,表示为局部支持数,记作 $A.\sup_V$, 也称作分布式环境中的全局支持数,即 $A.\sup = \sum_{V=1}^{X-1} A.\sup$,可以设定为 $\min - \sup$ [9-10]。B 和 NM_V 中元数据的个数为 B 和 NM_V 的大小,分别记作 $|B|$ 和 $|\text{NM}_V|$。$\min - \sup \times |\ \text{NM}_V\ |$ 为站点 C_V 的局部最小支持阈值, $\min - \sup \times |B|$ 为全局最小支持阈值,分别记作 $\min - \sup D_V$ 和 $\min - \sup D$。

在序列 A 中若 $A.\sup_V \geqslant \min - \sup \times |\ \text{NM}_V\ |$, 则称 A 为站点 C_V 的局部元数据存储模式[11]。当 $A.\sup \geqslant \min - \sup \times |\ \text{NM}_V\ | = \min - \sup \times (|\ \text{NM}_1\ | + |\ \text{NM}_2\ | + \cdots + |\ \text{NM}_{V-1}\ |)$ 成立,则 A 为全部元数据存储模式。因此对于站点 C_V 来讲,在数据存储时以序列模式完成分级流程设定,将其对应在各个分块中,并加设用户隐私作为存储条件。

3 考虑用户隐私构建索引密钥

为保证数据的安全存储,减少用户的隐私暴露问题,通过空间向量为检索基础,以考虑用户隐私作存储条件,构建数据的索引密钥[12-14]。针对分布式数据中的特征向量,获取数据存储相似性为

$$\text{consine}(H_I, G) = \frac{\sum_{J=1}^{K} \boldsymbol{L}_{IJ}\boldsymbol{G}_J}{\sqrt{\sum_{J=1}^{K} \boldsymbol{L}_{IJ}^{\ 2} \times \sum_{J=1}^{K} \boldsymbol{G}_J^{\ 2}}} \tag{8}$$

式中,存储特征为 G, 其转化向量为 $\boldsymbol{G}_J$。数据为 H_I。在流转时间 K 中,其特征向量为 $\boldsymbol{L}_{IJ}$。数据存储相似性为。在高精准度要求下,以单个密钥检索 $\text{consine}(G, H_I)$ 为基础,其加密频数为

$$Q_{\text{WE}} = 1 + \ln(\text{WE}) \tag{9}$$

式中, ln 表示单个密钥的显示次数。单个密钥为 WE。加密频数为 Q_{WE}。在同类型数据中,以全局序列为基础,各组数据的相对加密相似性,则为

$$R_T = \ln \frac{Y}{Y_T} \tag{10}$$

式中, R_T 为单组词汇的加密相似性。Y 为同类数据中的数据总量。Y_T 为数据出现频次,则单组数据与单个密钥的加密程度成反比趋势。T 为标识数据频次。以此引入安全索引 $\breve{\boldsymbol{H}}_U$ 进行加密,获取数据在存储过程中的相似度为

$$\begin{aligned}\breve{\boldsymbol{H}}_U \cdot \breve{\boldsymbol{G}} &= \{\boldsymbol{O}_1^P \boldsymbol{H}_U'',\ \boldsymbol{O}_2^P \boldsymbol{H}_U'''\} \cdot \{\boldsymbol{O}_1^{-1}\boldsymbol{G}'', \boldsymbol{O}_2^{-1}\boldsymbol{G}'''\} =\\ &\tilde{\boldsymbol{H}}_U \cdot \tilde{\boldsymbol{G}} + \sum_{I=1}^{S} \chi_I = \text{consine}(\boldsymbol{H}_U \cdot G) + \sum_{I=1}^{T} \chi_I\end{aligned} \tag{11}$$

式中，$\boldsymbol{S}$ 为随机比特向量中的一组随机向量，安全存储请求为 $\breve{\boldsymbol{G}}$，按照向量要求将其划分为 $\breve{\boldsymbol{G}} = \{\boldsymbol{O}_1^{-1}\boldsymbol{G}'', \boldsymbol{O}_2^{-1}\boldsymbol{G}'''\}$，其中 $\boldsymbol{O}_1$ 和 $\boldsymbol{O}_2$ 为随机加密矩阵，大小为 $(|\boldsymbol{P}| + \boldsymbol{T}) \times (|\boldsymbol{P}| + \boldsymbol{T})$。此时 $\boldsymbol{P}$ 对应数据的存储长度，$\boldsymbol{T}$ 为随机比特向量中的一组随机向量，其特征值的随机值设定为 $\boldsymbol{\chi}_I$，且 $\boldsymbol{I} \in [1, T]$。原始的数据存储向量为 $\boldsymbol{G}$，在长度下扩充为 $\tilde{\boldsymbol{G}}$。索引加密请求为 $\breve{\boldsymbol{H}}_U$，表示为 $\breve{\boldsymbol{H}}_U = \{\boldsymbol{O}_1^P\boldsymbol{H}_U'', \boldsymbol{O}_2^P\boldsymbol{H}_U'''\}$，原始数据加密索引 $\boldsymbol{H}_U$，在扩充后表示为 $\tilde{\boldsymbol{H}}_U$，可分解为 $\{\boldsymbol{H}_U'', \boldsymbol{H}_U'''\}$。根据同类型数据的加密过程，基于层次模型对分布式数据进行中心分级存储。

4　基于层次模型分级存储数据

在构建数据索引密钥的基础上，将中心分开存储过程看作为一个多层次结构，引入 BINDER 分层模型建立层次关系，按照适配原理实现数据的分级存储。由于该模型自带配置层和通信层，能在 BINDER 驱动中匹配两端数据库，通过本地服务器中传输的数据，自动搭建传输和存储框架，并按照特定通路完成服务端的分级存储，层次关系见图 1 所示。

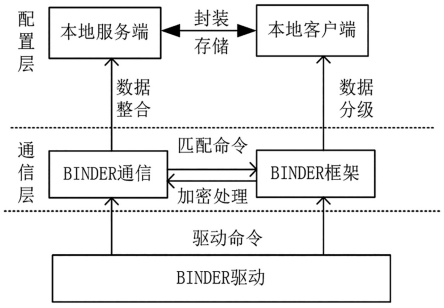


图 1　分层模型下数据存储的层次关系

Fig. 1　Data storage hierarchy in the layered model

由图 1 可知，整体模型中 BINDER 驱动是实现数据存储的重要基础，数据库通过封装处理，将驱动命令转化为存储命令，实现本地数据传输的加密处理，对上文中的密钥进行代入，保证数据中用户的信息隐私安全。在该框架下，可对本地服务器进行多次封装，并通过数据重复调用和存储，为数据的存储和交互提供技术支持。

以 BINDER 分层模型为基础，对分布式的多数据进行分级处理，选择 CLIENT-SERVICE 存储模式，以代理形式执行数据存储命令。划分层级模型中的 NATIVE 交互层，针对分布式数据传输对象，实现对应各自服务主体的存储，具体见表 1 所示。

以表 1 中内容所示，通过 BINDER 层级模型连接本地服务器，在各个层级对应的功能下，与本地数据进行转化，按照不同的数据节点实现接收的传输，完成分布式数据分级存储。至此，在考虑用户隐私的前提下，通过构建存储流程和划分存储框架，实现分布式多数据的中心分级存储方法设计。

表 1　分级交互数据存储模式

Tab. 1　Interactiveata storagem odel with hierarchical structure

分层	对象	功能
BINDER	实体数据	将 B-BINDER 与其连接
BINDER-PROXY	代理数据	实现与 BP-BINDER 的连接
B-BINDER	本地数据	1. 接收 BINDER-IPC 2. 生成 BP-BINDER
BP-BINDER	数据节点	1. 接收和传递 BINDER-IPC 2. 实现 BINDER-PRC
BINDER-IPC	数据代码	在发送过程中将其传输给 BINDER
BINDER-PRC	数据转化	在接收过程中将其传输给 B-BINDER

5　实验测试分析

5.1　搭建实验环境

为验证上文提出的储存方法具有使用性能，在 100 M 的局域网中搭建测试环境，以 4 组 PC 机作为分布式站点，在 1 组 DELL 工作站中接收传输数据，并完成分布式数据的中心分级存储。整个测试既需要满足数据存储的效率，也需要保证数据存储的安全度，PC 机配置为 16 GB，主站配置为 128 GB。利用 VISUALC+8.0 程序构建操作系统，以随机采样法对数据进行收集，并将本文方法与多目标存储和 DSP 存储进行比较。各组 PC 机在不同支持度阈值变化时，产生的数据量见表 2 所示。

表 2　分布式 PC 机数据量/组

Tab. 2　Data volume of distributed PCs

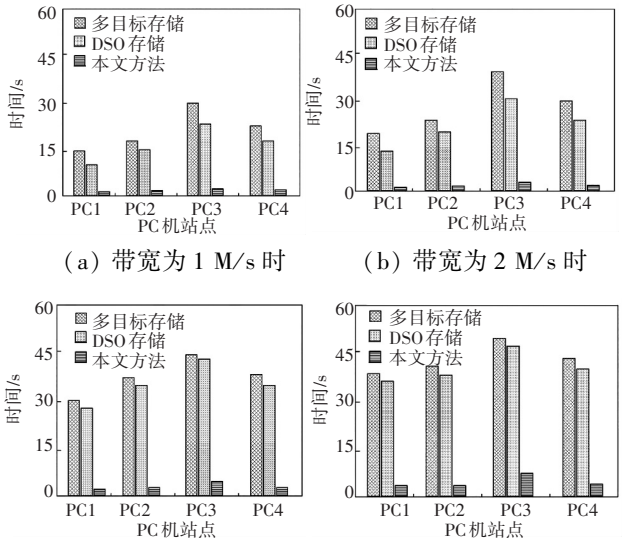
支持度阈值/%	PC1	PC2	PC3	PC4
0.25	200	500	1 000	800
0.75	300	700	1 200	900
1.25	400	900	1 400	1 000
1.75	500	1 100	1 600	1 100
2.25	600	1 300	1 800	1 200
2.75	700	1 500	2 000	1 300
3.25	800	1 700	2 200	1 400

根据表中内容所示，在不同支持度阈值下各组 PC 机的数据产生量逐渐增加，其中 PC1 和 PC3 站点中分别为最少数据和最多数据的分级中心。为保证不同站点在运行时，各自的数据能够按照其所属类型进行对应存储，并在较短时间内完成节点数据的整合，将测试分为两个阶段。第一阶段检测存储效率，以数据传输量为基础，分别在不同网络带宽下，比较各组方法在不同 PC 机中的应用效果。第二阶段验证存储过程中，各组 PC 机数据的抗攻击程度，即验证不同方法的存储安全性。

5.2　存储效率检测

在对应的存储空间中，若数据的运行负载过高，则会影响数据的存储效率，按照上文中搭建的分布式局域网结构，分别连接选择的三组存储方法。以不同的网络宽带限

制作为变量,在每组 PC 机最大数据负载下,完成存储效率测试,结果见图 2 所示。



(a) 带宽为 1 M/s 时 (b) 带宽为 2 M/s 时

(c) 带宽为 4 M/s 时 (d) 带宽为 8 M/s 时

图 2 不同带宽限制下数据存储效率对比结果

Fig. 2 Data storage efficiency under different bandwidth constraints

根据图 2 中内容所示,数据的传输负载越小,即数据量越小,所需的存储时间就越少。而越低的带宽条件限制,对数据的存储影响越小,随着带宽限制的增加,不同方法对各组 PC 机数据的存储时间有所增加。对于本文方法来看,其数据的存储时间变化较小,以 PC3 最大数据负载为例,其在带宽限制为 8 M/S 时,存储时间仍可以保持在 10 s 以内,而两组传统方法所需时间均超过 35 s,综合说明本文方法更加有效。

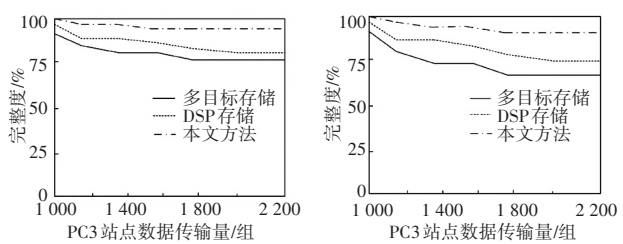
5.3 存储安全检测

在上文测试结果中验证了新方法的有效性,以此为基础,直接选择 PC3 为测试对象,在不同攻击强度下验证三组方法的数据存储完整度,检验公式为

$$\mathrm{mmn}(x)=\frac{\varepsilon(x)}{\mathrm{full}(x)}\times 100\% \tag{12}$$

式中, $\mathrm{mmn}(x)$ 为数据的完整度, $\varepsilon(x)$ 为遭受攻击后数据的保存量, $\mathrm{full}(x)$ 为原始数据传输量。

以完整度表示数据存储过程的安全程度,结果见图 3 所示。



(a) 1 次攻击中存储完整度 (b) 10 次攻击中存储完整度

图 3 不同攻击强度下数据存储结果

Fig. 3 Data storage results under different attack intensities

根据图 3 中内容所示,在不同攻击强度下各组方法的数据存储完整度存在差异,当数据传输量越小时,数据的保留程度越高,即使受到了一定程度的攻击,也能在小范围内保证数据的完整性。而随着数据传输量的增加,在高强度攻击下,数据的保留程度有所降低。本文方法应用在设定的两组攻击强度下,不论数据传输量如何变化,均可保证数据的完整度在 90%以上,而两组传统方法在 10 次攻击中,数据保留的完整度分别为 75%和 70%,说明本文方法更具有应用价值。

6 结论

在移动通信技术的快速发展势态下,移动终端向智能化转变,人们有了获取信息更加便利的渠道,而海量数据的产生和流转过程中,对用户自身的隐私情况产生了巨大威胁。为保证数据存储的安全性,本文在考虑用户隐私的前提下,设计了一个新的存储方法,并在实验测试中论证了其有效性。网络作为信息传输载体,其中存在的攻击类型多种多样,由于时间限制,本文没有对具体的类型进行描述,仅以攻击次数作为对照,具有一定不足之处,后续研究中会加深研究程度,为网络的数据安全存储提供理论支持。

参考文献

[1]原迪. 软件定义数据中心网络中自适应路由技术[J]. 电子产品世界, 2022, 29(9): 74-76, 87.
[2]李秀艳, 刘明曦, 史闻博, 等. 基于云存储的动态组共享数据完整性验证方案[J]. 计算机工程与设计, 2022, 43(6):1510-1519.
[3]刘家丞, 吴江, 刘鹏远, 等. 面向负荷特征分析的地理分布式协同聚类方法[J]. 电力系统自动化, 2022, 46(15):112-120.
[4]方世敏, 朱建华. 高效区块链用户隐私数据网故障溯源算法研究[J]. 现代电子技术, 2022, 45(2):162-166.
[5]韩飞, 张葛祥. 基于超混沌系统的网络用户隐私信息加密仿真[J]. 计算机仿真, 2021, 38(12):295-298, 405.
[6]耿秀丽, 王著鑫. 考虑用户兴趣分析的差分隐私方案推荐[J]. 计算机应用研究, 2022, 39(2):474-478.
[7]陈昊, 张嵩, 吕途. 智能系统用户隐私意识与隐私保护意愿研究[J]. 情报理论与实践, 2022, 45(2):168f175.
[8]唐镇, 胡勇华, 陆浩松, 等. 基于弱约束指派的 DSP 寄存器偶对分配算法研究[J]. 计算机科学, 2021, 48(增刊 1):587-595.
[9]刘期烈, 陈澄. 基于区块链的 V2G 匿名身份认证方案[J]. 计算机工程, 2021, 47(11): 22-28.
[10]贺敬伟, 程伟华, 张世杰. 基于 Kubernetes 调度算法的动态负载均衡方法研究[J]. 自动化技术与应用, 2025, 44(9):138-142.
[11]谭琪, 张凤荔, 张志扬, 等. 社交网络用户影响力的建模方法[J]. 计算机科学, 2021, 48(2):76-86.
[12]唐远富, 陈远扬. 基于数据驱动模型的电力大数据关联规则挖掘研究[J]. 自动化技术与应用, 2024 , 43 (12): 110-113, 162.
[13]李姣军, 蒋扬, 邱天, 等. 基于压缩感知的 OFDM 稀疏信道估计算法[J]. 重庆理工大学学报(自然科学), 2021, 35(4):117-122.
[14]黄振, 王涛, 邵志敏,等. 基于区块链技术的电力共享数据动态溯源方法[J]. 自动化技术与应用, 2025, 44(3):110-114.