DOI:10.12158/j.2096-3203.2022.03.006

基于区块链与数据湖的电力数据存储与共享方法

曾飞，杨雄，苏伟，肖小龙，易文飞
（国网江苏省电力有限公司电力科学研究院，江苏 南京 211103）

摘　要：针对电力物联网边设备之间、云主站平台营配调各系统之间的数据存储和共享的需求，提出了一种基于区块链与数据湖的电力数据存储与共享方法。首先，设计边缘层分布式电力数据存储架构，通过环签名和 CryptoNote 协议对边设备存储节点间的数据交互进行加密，实现双方身份认证，并基于区块链智能合约实现电力数据安全存储系统中节点间的数据共享。然后，基于数据湖与智能合约构建营配调平台间的数据共享和访问控制模型，只在链中存储数据的哈希值，而将数据存储在数据湖中，以实现数据跨平台跨域安全共享和访问。最后，搭建实验平台对所提方法进行论证，结果表明，所提方法的最高存储延迟时间短，并且吞吐量和安全性也较高，具有良好的应用前景。

关键词：区块链；电力物联网；分布式存储；环签名；智能合约；数据湖
中图分类号：TM769　　**文献标志码**：A　　**文章编号**：2096-3203(2022)03-0048-07

0　引言

电力数据的安全高效存储共享是电网业务稳定可靠服务用户的重要保障[1—2]。随着电力物联网建设的推进，如何进行海量数据存储和共享越来越受到关注[3—4]。具体而言，云主站平台中营配调各系统间的数据需要存储和共享访问，边设备和边设备之间需要进行数据共享访问，以支持营配调融合的发展趋势，打造大电网战略。因此，边设备和边设备之间存储的数据如何进行安全高效共享访问以及原先营配调各个系统的“数据孤岛”如何打通都是亟须解决的难题[5—6]。

相关数据存储和共享的研究主要分集中式和分布式。文献[7]基于 Hadoop 平台对电网大数据进行了优化存储，提出了哈希桶存储方式，实现相关联数据的集中存储和多源配用电数据规范化集成，以缩短数据查询和分析耗时。文献[8]提出基于 Spark 平台的弹性分布数据集概率模型，可有效处理可再生能源的大规模样本数据。文献[9]基于动态实时优先级调度算法，提出一种数据处理任务控制调度模型，以优化数据中心的数据处理能力。但是上述方法的数据在传输和存储过程中对集中式存储中心服务器造成的访问压力巨大，吞吐量易受通信瓶颈制约[10]。

分布式存储技术将数据信息分散存储在多个独立设备中，提升了数据存取效率[11]。其中，区块链技术是分布式存储的重要研究方向，其存储节点通过维护一个不可篡改的公共账本以实现数据共享[12—13]。而数据湖以原始数据和可用于分析的形式存储所有数据，是一种安全的辅助存储库。文献[14]利用区块链技术取代集中式服务器，执行存储交互验证并借助无证书密码学进行审验，但该做法会导致数据访问延时过大。文献[15]提出基于人工智能驱动的网络框架，建立相互信任的数据共享框架，结合监督学习和基于智能合约的细粒度数据访问控制，确保数据共享环节中的隐私性，但该方法容易导致数据存储冗余度增加[16]。文献[17]提出基于区块链的共享存储系统模型，在不涉及第三方的情况下为用户提供隐私保护。文献[18]以电动汽车和充换电站为分布式存储单元，在存储数据的同时实现通用数据与资源的共享，并确保全部数据信息不被篡改。以上方法对区块链在电力系统的应用做了一些尝试，但未探讨数据存储安全和存取延时方面的平衡。

因此，对于电力物联网边设备和边设备之间、云主站平台营配调各系统间的数据存储和共享需求，结合区块链与数据湖技术的良好特性，文中提出了基于区块链与数据湖的电力数据存储与共享方法。在边缘侧设计分布式电力数据存储架构，通过环签名和 CryptoNote 协议加密边设备存储节点间的数据，利用区块链智能合约实现数据共享。构建基于数据湖与智能合约的数据共享和访问控制模型，解决营配调不同平台之间数据难共享和难访问的问题。

1　电力物联网的数据管理架构

电力物联网采用云边端架构，如图 1 所示。对

收稿日期：2021-12-25；修回日期：2022-03-09
基金项目：江苏省自然科学基金资助项目(BK20210056)

于边缘层设备,利用分布式存储数据库与区块链作为底层支撑,实现数据存储与共享。同时,将边缘层数据信息汇聚后上传至云层,云层基于改进数据湖模型进行数据共享与访问。

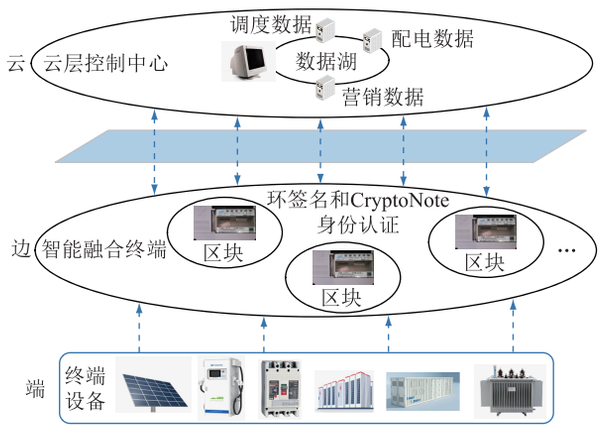


图 1　电力物联网系统架构

Fig.1　System architecture of power Internet of Things

(1) 边缘层。汇聚智能电表、低压故障指示器、智能单元等终端设备实时采集的电力数据,进行分布式存储与共享。针对电力物联网边缘层中边设备之间存储数据无法进行安全高效共享访问的问题,提出分布式电力数据存储架构,利用分布式存储数据库与区块链作为底层支撑,并且通过环签名和 CryptoNote 协议加密边设备存储节点间的数据,利用区块链智能合约实现系统节点间的数据共享。

(2) 云层。对于电力物联网云主站中营配调各系统间存在"数据孤岛"的问题,基于数据湖与智能合约的数据共享和访问控制模型,在数据湖存储的基础上优化区块链中的智能合约模块,实现营配调不同平台之间数据共享与访问。

2　边缘层分布式电力数据存储共享

2.1　分布式电力数据存储架构

文中提出的分布式电力数据存储模型将区块链技术应用到数据存储的实际场景中,其具体架构如图 2 所示。分布式电力数据存储模型分为 2 个模块,分别为分布式存储数据库和区块链,其中存储数据库支撑区块链的存储,而区块链支撑存储数据库的安全防护。电力用户通过端设备将存储请求传至边设备,如智能融合终端、物联代理装置等,每个边设备拥有若干个存储节点,组合成数据聚合器,边设备收到存储请求后将需求传至各自的存储节点,其过程记录于区块链中。

文中采用脱链存储,只在区块体中存储用户名、用户地址和存储记录的信息,而将采集的原始

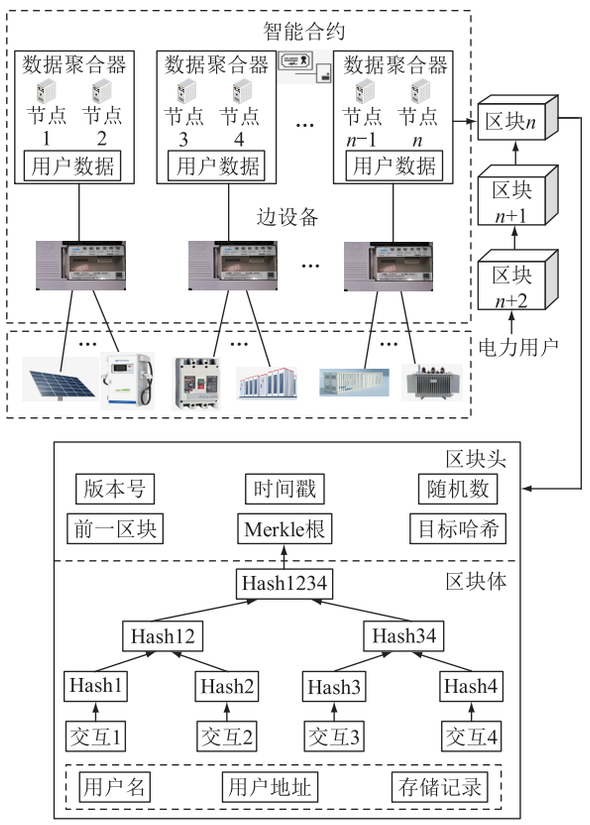


图 2　电力数据分布式存储框架

Fig.2　Distributed storage framework for power data

数据存储在边设备数据聚合器的存储节点中,在存储时进行加密,加密存储过程如下。

第 1 步:电力用户向边设备提出存储需求,选择其中任一节点存储数据,且一定时间(如 1 min)内只能向该节点发送存储请求。

第 2 步:存储节点收到请求后回复存储响应给发送方,并提供存储服务次序号,待发送方确认后即可进行数据存储。

第 3 步:存储节点上传存储记录到链上的区块中,每个区块均由区块头和区块体组成。

第 4 步:用户对存储过程进行信用评价,系统根据结果评估该存储节点性能。

2.2　节点间身份认证

在分布式电力数据安全存储系统中,采用 CryptoNote协议对边设备存储节点间的数据交互进行加密,并进行双方身份认证。CryptoNote 提供环签名和密钥图像,如果多次使用同一签名,则接收者将拒绝该存储交互[19]。CryptoNote 协议的交互过程如下:首先,根据接收者的随机数据 (A,B) 与发送者的随机数据 r 相关联得出签名值 P。

$$P = H_s(rA)G + B \tag{1}$$

式中:H_s 为加密散列函数;G 为基点。然后,接收者的随机数据由发送者通过安全通道询问接收者获

得,接收者使用其私钥 (a,b) 检查收到的存储交互数据,并获得认证值 P' 为:

$$P' = H_s(aR)G + b \tag{2}$$

式中:R 为认证权重。最后,接收者可以回复对应的一次性密钥 x 为:

$$x = H_s(aR) + b \tag{3}$$

由于接收者收到的消息与一次性密钥关联,因此此协议的数据存储交互对于攻击者而言是不可追踪的。

CryptoNote 协议收发每一方都有唯一的父密钥,用于提交业务,并为每个业务生成一个隐身地址[20],密钥从预先生成的门限置换函数获得。隐身地址的生成过程如图 3 所示。

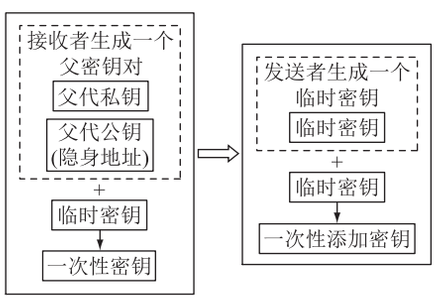


图 3 隐身地址的生成过程

Fig.3 Generation process of stealth address

首先,接收者生成父密钥对,与临时密钥组成一次性密钥并发布,发布的密钥称为隐身地址。然后,发送者接收该密钥,与其生成的临时密钥组合成新的一次性添加密钥,即新的一次性存储交互地址。可见,存储交互地址会随着存储交互而进行不断更新,很难被攻击者捕捉[21—22]。CryptoNote 中使用了隐身地址协议,为区块链网络中的用户提供更强的隐私性。同时,存储系统中的边设备和用户采用环签名对信息进行签名认证,认证后方可传输。签名协议的过程定义如下[23]:

(1) $\varphi_{\text{Sign}}(m,P_s,P_1,P_2,\cdots,P_n)$,签名由每个消息成员 m 的公钥 $(P_s,P_1,P_2,\cdots,P_n)$ 组成,并与签名者的私钥 P_s 相关联以产生签名 φ_{Sign}。

(2) $f_{\text{Verify}}(m,\varphi_{\text{Sign}})$,验证签名 φ_{Sign},由所有可能的签名者的公钥以及消息 m 组成,输出结果是真或假。

签名者可以直接生成环签名,而无需组管理器参与。签名者计算对称密钥 S 作为待签名消息 m 的哈希值,即 $S = h(m)$,创建环签名组需要每一方的公钥,签名者可以选择要在存储交互中使用的签名数,以提供模糊的签名者。假设存储数据由 $\omega = l_i n + r_i$ 组成,该数据的签名 $f_i(\omega)$ 可表示为:

$$f_i(\omega) = \begin{cases} l_i n + d_i q_i & n \leqslant 2^b \\ \omega(l_i + 1) & n > 2^b \end{cases} \tag{4}$$

式中:d_i 为发送者的签名;q_i 为发送者产生的随机数;l_i 为接收者产生的随机数;n 为签名者数量;r_i 为校验码。

由式(4)可以看出,$f_i(\omega)$ 是{0,1}上的置换函数,这也是一个单向门限函数,可基于成员公钥的加密构造一组包含 n 个签名者的环签名:

$$R_s = f_1 \oplus f_2 \oplus \cdots \oplus f_n \tag{5}$$

发送者使用环签名为存储交互的消息签名,而无需组管理器参与,因此攻击者无法判断真实发送者对于相应存储交互的身份。

2.3 节点间数据共享

分布式电力数据安全存储系统中,节点间的数据共享通过区块链智能合约实现。智能合约的脚本类型包含了锁定与解锁 2 种,锁定脚本限定了共享数据的输出约束,而解锁脚本则限定了其运行标准。基于智能合约的数据共享流程总结如下。

(1) 共享访问请求。数据共享的请求 R 由数据存储节点 N_p 发出,由存储节点 N_q 接收,并且 R 中需要具备访问地址、时间与频次等信息。存储节点 N_q 针对 N_p 设置约束条件 C_0,并将该条件通过私钥 S_{PID} 传至邻近集合器 B_j:

$$\begin{cases} N_q \rightarrow N_p : R = f_{\text{PK}_{N_p}}(C_0 \| C_{N_p} \| t) \\ N_p \rightarrow B_j : M = f_{\text{PK}_{B_j}}(C_0 \| S_{\text{PID}} \| P_{N_p} \| t \| C_{N_q}) \end{cases} \tag{6}$$

式中:C_{N_p},C_{N_q} 分别为数据存储节点 N_p,N_q 的证书;P_{N_p} 为公钥;f_{PK} 为实体加密信息;M 为存储的数据;t 为时间戳。

(2) 执行智能合约。B_j 验证信息后根据设置的约束条件锁定脚本,并采用 P_{N_p} 加密环签名。

(3) 发送共享数据。若 N_q 与 N_p 处于同一个数据聚合器中,可通过 B_j 直接获得共享数据;若 N_q 与 N_p 不在同一聚合器中,则通过合约方式传至相邻的 B_{j+1} 获得共享数据。

(4) 指定访问数据。N_p 接收到数据后,经过私钥解密方可进行数据访问。

3 云层基于数据湖的共享与访问模型

随着电力物联网建设的推进,营配调融合以及大电网搭建逐步被提上日程,因此,营配调各类业务数据将不单只局限于各自部门内的系统进行共享,还将涉及电力物联网云层不同平台之间的数据共享。而数据湖是大数据应用中的一种数据共享方式,特别适用于跨平台间的数据共享,其本质是

一种数据管理的思路,可以存储不同规模、不同结构、不同类型的数据,包括不同量级的结构化、半结构化和图片文本等非结构化数据,允许各业务方通过访问工具和框架来访问数据而不必迁移,大大节省定义数据结构和转换的时间,使得跨平台、跨领域的数据分析能够低成本、高效率实现。因此,采用数据湖可以打破国网等单位数据中心“数据孤岛”的闭塞,在各数据仓库之间建立连接,允许营配调各类业务访问数据,解决电网公司传统数据仓库的痛点。但是数据湖的数据共享缺乏安全防护,因此,文中提出基于数据湖与智能合约的数据共享和访问控制模型,只在链中存储数据的哈希值,而将数据存储在数据湖中,兼顾了营配调不同平台之间的数据共享和安全性。

3.1 数据湖数据共享

数据湖中数据传送方和接收方两者的共享准则自主实行,并不存在统一的信任规则。利用智能合约能够让营配调用户具备数据访问与运用模式的控制权,同时采用分开的虚拟机执行相应的智能合约,用户不能随意篡改最终的结果。而数据湖能够保质保量存储所有类型的数据,并且分析处理数据的成本更低且速度快,这得益于其共享机制,如图 4 所示。

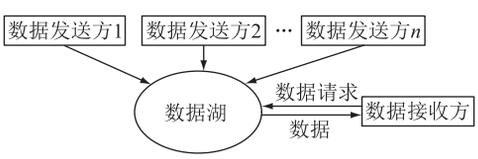


图 4 数据湖中数据共享机制

Fig.4 Data sharing mechanism in data lake

提出的数据共享与访问模型只在区块链中存储数据的哈希值,将加密后的数据存于数据湖对,营配调用户通过智能合约进行数据访问。

3.2 数据湖访问控制

营配调不同用户在采用智能合约对数据湖进行访问时,其访问控制和数据使用具有平等地位。基于智能合约的数据湖访问管理模型如图 5 所示。

该模型包含端设备和边设备、区块链中的智能合约及数据湖中的营配调数据三部分。现有电力系统中通常广泛采用纵向加密设备,在接入网交换机和数据网路由器之间增加纵向加密设备,并在设备中添加隧道和访问控制策略,所有与访问请求都将经由纵向加密设备放行方能进入调度数据中心。每个端设备和边设备均加设加密芯片,入湖的数据均须通过加密芯片进行加密,同时也解密收到数据信息;电网公司营配调用户通过区块链智能合约和

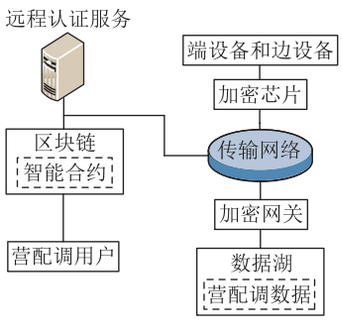


图 5 数据湖中访问控制模型

Fig.5 Access control model in data lake

远程服务认证进行数据访问;数据湖之前加设加密网关,对入湖、出湖的数据进行加解密,从而确保数据湖数据访问的安全性。其中智能合约模块将数据的哈希值存放在区块链中,但加密的原数据存放在数据湖中。具体而言,在设备进行数据存放时,须提供设备的合约地址、唯一识别号以及要写入的数据,并将其组合成哈希映射列表,列表中存储与该设备对应的所有数据信息,数据写入时必须验证对应的合约地址和唯一识别号,以确保只有合约设备才能进行数据存放操作。而在数据读取时,要求访问设备发送数据读取请求,其中包含用户设备合约地址和唯一识别号组成的哈希映射,数据湖检查该哈希映射,只有已注册的用户设备才能获得访问权限。

4 实验结果与分析

实验主要采用延迟时间、吞吐量和安全性作为性能指标来评估所提方法。使用 35 台主机构建所提的存储体系,系统内所有主机硬件设置为 64 GB 内存和英特尔 i7 处理器。边设备使用笔记本电脑进行模拟,并且利用分布式框架相互连接形成区块链中数据存储节点,便携式电脑硬件设置为 32 GB 闪存和英特尔 i5 处理器。

4.1 存储延迟分析

影响数据存储的参数之一是块大小,每个区块都有自身的大小,具体取决于业务类型。设置块大小的目的在于防止攻击,例如拒绝服务攻击。同时,块大小也会影响确认的时间长度,通常来说,块越大,确认所需的时间就越长。因此,块的大小与延迟时间之间存在相关性。块大小对存储时间的影响如图 6 所示,其中,业务的块大小在 0~350 KB 范围内变化,以便确定接收该块所需的时间。

图 6 中,10%,25%,50%,80%代表业务的占比,数值越高,代表数据块越大。从图 6 中的变化趋势可以看出,数据块越大,延迟时间越长。

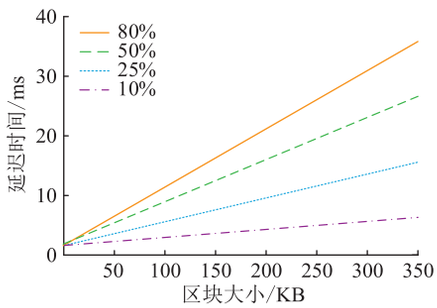
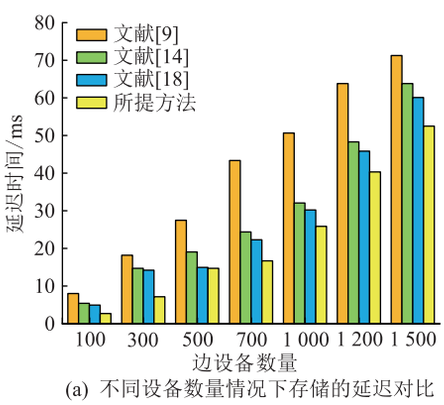


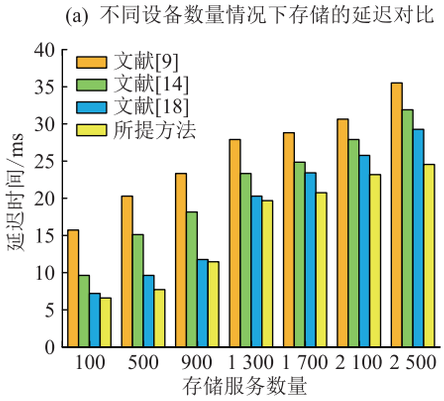
图 6　块大小与存储延迟时间之间的关联

Fig.6　Correlation between block size and storage delay time

延迟时间作为评价存储方法的重要指标之一，有必要对所提方法在这一方面的性能进行验证，其与文献[9]、文献[14]和文献[18]中方法的延迟时间对比结果如图 7 所示。



(a) 不同设备数量情况下存储的延迟对比



(b) 不同存储服务数量情况下存储的延迟对比

图 7　不同存储方法的延迟对比

Fig.7　Delay comparison of different storage methods

由图 7 可以看出，随着边设备和服务数目的增加，存储延迟时间也在不断增加，但是相比于其他对比方法，所提方法的延迟时间最短。这是由于所提方法采用分布式存储，避免了集中存储的拥塞，并且使用区块链和改进智能合约去中心化，进一步缩短了存储延迟时间。文献[9]中采用集中处理的存储方式，延迟时间最长；文献[14]采用基于区块链的数据存储方式，文献[18]采用三层分布式存储架构，均能够在一定程度上缩短延迟时间。

4.2　吞吐量分析

吞吐量是衡量存储系统性能的重要标准之一，其定义为单位时间内存入或读取的信息量。以电力数据作为存储服务，将所提方法与文献[9]、文献[14]及文献[18]进行对比，吞吐量对比结果见图 8。

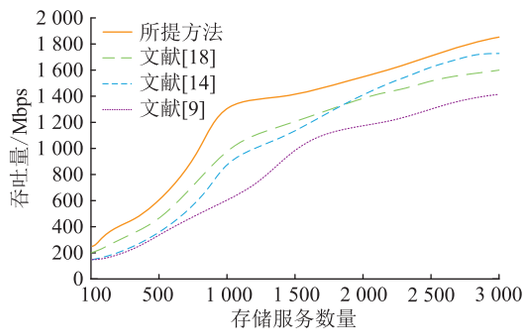


图 8　不同存储方法的吞吐量对比

Fig.8　Throughput comparison of different storage methods

由图 8 可以看出，所提方法的吞吐量高于其他对比方法，超过了 1 800 Mbit/s。这是由于所提方法采用区块链分布式存储架构，并采用环签名组，用户或者边设备可以根据自身需求寻找合适的存储节点，进一步避免数据拥塞，提高了吞吐量。所提方法可视为文献[14]和文献[18]方法的结合并进行优化，因此性能优于两者。文献[9]采用集中存储，大量的电力数据对中心存储器的要求太高，无法满足存储容量需求，因此吞吐量较小。

4.3　安全性分析

所提方法采用环签名、CryptoNote 协议等一系列加密手段，同时使用数据湖提高数据共享的安全性，因此安全性是电力数据存储性能的一个重要评价指标。将抵御外界攻击等非正常数据存储行为占比作为安全性的分析标准，具体指标为成功防止非正常数据存储次数与所有非正常数据存储次数的比值。所提方法与文献[9]、文献[14]、文献[18]的安全性对比如图 9 所示。

由图 9 可以看出，相比于其他方法，所提方法的安全性最高，能够抵御最多的外界网络威胁，即使文件大小达到 1 024 KB，其安全性也不低于 80%。文件大小的增加意味着所需存储的数据更多，也伴随着更高的安全风险，因此，电力大数据对存储系统的安全性提出了更高的要求。文献[9]采用集中存储方式，一旦中心处理器遭到攻击者篡改就会严重威胁其数据隐私性，因此安全性能不高，并且随着文件的增大，安全性能快速下降。文献[14]采用区块链存储方式，文献[18]采用分布式存储方式，均在一定程度上保证了存储的安全性能，但是缺乏

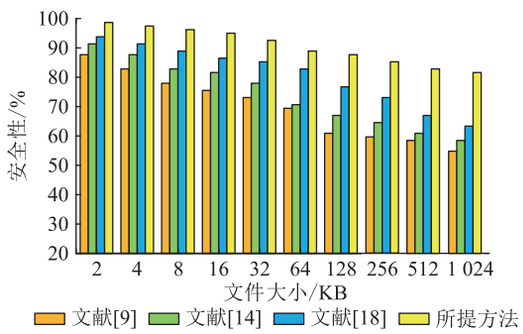


图 9 不同方法的存储安全性对比

Fig.9 Comparison of storage security of different methods

相应的加密算法,因此安全性能还有很大的提升空间。

5 结语

随着电力物联网电力数据的迅猛增长,集中式存储方式已无法满足时延、安全性等要求,为此,提出一种基于区块链与数据湖的电力数据存储与共享方法。实验结果表明,相比于其他对比方法,所提方法的存储延迟时间更短,吞吐量更高,且安全性能更佳,能够满足电力数据存储对安全性与高效性的要求。

下一步研究将考虑共享存储的模式容量,同时还要考虑区块链的可扩展性,以应对快速增长的电力数据。电力物联网的快速发展必将导致电力数据飞速增长,存储系统须具备很好的扩展性。

本文得到江苏省电力试验研究院有限公司科技项目"基于区块链的安全可信新一代配电自动化体系研究"资助,谨此致谢!

参考文献:

[1] HU C,DENG Y H. Aggregating correlated cold data to minimize the performance degradation and power consumption of cold storage nodes[J]. The Journal of Supercomputing,2019,75(2):662-687.

[2] KOIWA K,ISHII T,LIU K Z,et al. On the reduction of the rated power of energy storage system in wind farms[J]. IEEE Transactions on Power Systems,2020,35(4):2586-2596.

[3] WIDJONARKO W,SOENOKO R,WAHYUDI S,et al. Prediction of power characteristic curve on small scale compressed air energy storage by using regression analysis[J]. International Energy Journal,2019,19(2):89-100.

[4] 江秀臣,盛戈皞. 电力设备状态大数据分析的研究和应用[J]. 高电压技术,2018,44(4):1041-1050.
JIANG Xiuchen,SHENG Gehao. Research and application of big data analysis of power equipment condition[J]. High Voltage Engineering,2018,44(4):1041-1050.

[5] WANG Y C,LOU S H,WU Y W,et al. Flexible operation of retrofitted coal-fired power plants to reduce wind curtailment considering thermal energy storage[J]. IEEE Transactions on Power Systems,2020,35(2):1178-1187.

[6] 黄天恩,郭庆来,孙宏斌,等. 模型-数据混合驱动的电网安全特征选择和知识发现关键技术与工程应用[J]. 电力系统自动化,2019,43(1):95-101,208.
HUANG Tianen,GUO Qinglai,SUN Hongbin,et al. Hybrid model and data driven concepts for power system security feature selection and knowledge discovery:key technologies and engineering application[J]. Automation of Electric Power Systems,2019,43(1):95-101,208.

[7] 王林童,赵腾,张焰,等. 配用电大数据多源集成及存储优化方法[J]. 高电压技术,2018,44(4):1131-1139.
WANG Lintong,ZHAO Teng,ZHANG Yan,et al. Multi-source integration and storage optimization method for big data of power distribution and utilization[J]. High Voltage Engineering,2018,44(4):1131-1139.

[8] YANG Y,YU J,YANG M F,et al. Probabilistic modeling of renewable energy source based on Spark platform with large-scale sample data[J]. International Transactions on Electrical Energy Systems,2019,29(3):e2759.

[9] 王德文,刘庭辉. 电力全业务统一数据中心突发性数据处理任务调度方法[J]. 电力系统自动化,2018,42(8):177-184.
WANG Dewen,LIU Tinghui. Scheduling method for burst data processing task in full-service unified data center of electric power system[J]. Automation of Electric Power Systems,2018,42(8):177-184.

[10] MAJIDPOUR M,NAZARIPOUYA H,CHU P,et al. Fast univariate time series prediction of solar power for real-time control of energy storage system[J]. Forecasting,2018,1(1):107-120.

[11] 陶琼,王德顺,叶季蕾,等. 考虑储能配置模式的多数据源融合分布式光伏发电并网接纳分析方法[J]. 高电压技术,2018,44(4):1093-1098.
TAO Qiong,WANG Deshun,YE Jilei,et al. Capacity analysis of distributed photovoltaic generation integrated into power grid considering energy storage configuration mode based on fusion of multiple data sources[J]. High Voltage Engineering,2018,44(4):1093-1098.

[12] ESFANDYARI M J,HAIRI YAZDI M R,ESFAHANIAN V,et al. A hybrid model predictive and fuzzy logic based control method for state of power estimation of series-connected Lithium-ion batteries in HEVs[J]. Journal of Energy Storage,2019,24:100758.

[13] 韩平平,张祥民,丁明,等. Hadoop 数据存储分析技术在风电并网系统中的应用[J]. 电力系统及其自动化学报,2018,30(1):43-50.
HAN Pingping,ZHANG Xiangmin,DING Ming,et al. Application of data storage and analysis technology of Hadoop to wind power grid-connected system[J]. Proceedings of the CSU-EPSA,2018,30(1):43-50.

[14] LI R N,SONG T Y,MEI B,et al. Blockchain for large-scale

Internet of Things data storage and protection[J]. IEEE Transactions on Services Computing,2019,12(5):762-771.

[15] ZHANG G Z,LI T,LI Y,et al. Blockchain-based data sharing system for AI-powered network operations[J]. Journal of Communications and Information Networks,2018,3(3):1-8.

[16] 李新鹏,高欣,阎博,等. 基于孤立森林算法的电力调度流数据异常检测方法[J]. 电网技术,2019,43(4):1447-1456.
LI Xinpeng,GAO Xin,YAN Bo,et al. An approach of data anomaly detection in power dispatching streaming data based on isolation forest algorithm[J]. Power System Technology,2019,43(4):1447-1456.

[17] RAHMADIKA S,RHEE K H. Towardprivacy-preserving shared storage in untrusted block chain P2P networks[J]. Wireless Communications and Mobile Computing,2019,2019:6219868.

[18] 王冠男,杨镜非,王硕,等. 考虑 EV 换电站调度和区块链数据存储的电网分布式优化[J]. 电力系统自动化,2019,43(8):110-116,182.
WANG Guannan,YANG Jingfei,WANG Shuo,et al. Distributed optimization of power grid considering dispatching of electric vehicle battery swapping stations and data storage of blockchain[J]. Automation of Electric Power Systems,2019,43(8):110-116,182.

[19] ROUINDEJ K,SAMADANI E,FRASER R A. A comprehensive data-driven study of electrical power grid and its implications for the design,performance,and operational requirements of adiabatic compressed air energy storage systems[J]. Applied Energy,2020,257:113990.

[20] 陈思光,杨熠,黄黎明,等. 基于雾计算的智能电网安全与隐私保护数据聚合研究[J]. 南京邮电大学学报(自然科学版),2019,39(6):62-72.
CHEN Siguang,YANG Yi,HUANG Liming,et al. Fog computing based secure and privacy-aware data aggregation in smart grid[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition),2019,39(6):62-72.

[21] KIM K,KIM J,KIM H. Improving thereliability of pumped-storage power plants in the operational phases using data mining algorithms[J]. KSCE Journal of Civil Engineering,2018,22(12):4771-4778.

[22] 屈志坚,袁慎高,范明明. 电能质量在线监测系统海量数据的双列族存储设计[J]. 电力系统保护与控制,2019,47(2):154-160.
QU Zhijian,YUAN Shengao,FAN Mingming. Double column family storage design for massive data of power quality online monitoring system[J]. Power System Protection and Control,2019,47(2):154-160.

[23] 曾飞,杨雄,杨景刚,等. 一种基于区块链的电力数据存储方法及电力数据共享方法:CN112948868A[P]. 2021-06-11.
ZENG Fei,YANG Xiong,YANG Jinggang,et al. Electric power data storage method and electric power data sharing method based on block chain:CN112948868A[P]. 2021-06-11.

作者简介:


曾飞

曾飞(1984),男,硕士,高级工程师,从事配电自动化、调度自动化、电网故障诊断等相关工作(E-mail:zeng_nj2021@126.com);

杨雄(1983),男,博士,高级工程师,从事配电自动化、配电物联网、交直流配电网、分布式新能源发电、储能技术等相关工作;

苏伟(1993),男,硕士,工程师,从事配网运行、配电自动化等相关工作。

Power data storage and sharing method based on blockchain and data lake

ZENG Fei, YANG Xiong, SU Wei, XIAO Xiaolong, YI Wenfei

(State Grid Jiangsu Electric Power Co.,Ltd. Research Institute,Nanjing 211103,China)

Abstract:Aiming at the requirements of data storage and sharing between side devices and among marketing distribution dispatching systems in cloud master station platform of power Internet of Things,a power data storage and sharing method based on blockchain and data lake is proposed. Firstly,the distributed power data storage architecture on edge layer is designed. Through ring signature and CryptoNote protocol,the data interaction between side device storage nodes is encrypted to realize the identity authentication of both parties,and the data sharing between nodes in power data security storage system is realized based on intelligent contract of blockchain. Then,the data sharing and access control model between the operation and dispatching platforms is constructed based on the data lake and the smart contract,in which the hash value of the data is stored in the blockchain,and the data is stored in the data lake,so as to realize the cross platform and cross domain secure sharing and access of data. Finally,an experimental platform is built to demonstrate the proposed method. The results show that the maximum storage delay time of the proposed method is short,and the throughput and security are also high,so it has a good application prospect.

Keywords:blockchain;power Internet of Things;distributed storage;ring signature;smart contract;data lake

(编辑 钱悦)