doi :10.3969/j.issn.1671-1122.2019.02.004

基于纠删码的区块链系统区块文件存储模型的研究与应用

赵国锋[1]，张明聪[1]，周继华[2]，赵涛[2]

（1. 重庆邮电大学通信与信息工程学院，重庆 400065；2. 重庆金美通信有限公司，重庆 400030）

摘　要：文章提出了一种区块链系统区块文件存储模型，并将其应用于超级账本区块链系统。模型基于纠删码技术改进区块链系统的存储性能，对其存储的区块文件进行编码分片存储，能够以更小的数据冗余度获得不弱于原始系统的数据可靠性，大幅减少了节点对存储资源的需求。同时，模型提供了区块文件解码恢复方案及适配的区块同步方案，增强了模型的可用性。将模型在多个区块链节点中运行测试，结果表明该模型显著减少了节点的存储空间占用，有效改善了区块链系统的存储可扩展性。

关键词：区块链；纠删码；存储可扩展

中图分类号：TP309　**文献标识码：**A　**文章编号：**1671-1122（2019）02-0028-08

中文引用格式：赵国锋，张明聪，周继华，等．基于纠删码的区块链系统区块文件存储模型的研究与应用 [J]. 信息网络安全，2019，19（2）：28–35.

英文引用格式：ZHAO Guofeng, ZHANG Mingcong, ZHOU Jihua, et al. Research and Application of Block File Storage Model Based on Blockchain System of Erasure Code[J]. Netinfo Security, 2019, 19(2):28–35.

Research and Application of Block File Storage Model Based on Blockchain System of Erasure Code

ZHAO Guofeng[1], ZHANG Mingcong[1], ZHOU Jihua[2], ZHAO Tao[2]

(1. *School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*; 2. *Chongqing Jinmei Communication Co.Ltd., Chongqing* 400030, *China*)

Abstract: In this paper, a novel block file storage model is proposed by using erasure code technology, and is applied to the blockchain system. The storage system of blockchain is improved by using erasure code technology, and the stored block file is coded and stored in slices. With less data redundancy, the data reliability is not weaker than the original system, and the node's demand for storage is reduced. Moreover, the model provides a block file decoding recovery scheme, as well as an adapted block synchronization strategy, which enhances the usability of the model. Finally, this paper tests the model by running in multiple fabric nodes, the results shows that it can reduce the storage space of the nodes and effectively improve the storage scalability of the blockchain system.

Key words: blockchain; erasure code; storage expandable

收稿日期：2018-12-10

基金项目：重庆市重点研发项目 [cstc2018jszx-cyzdX0120]

作者简介：赵国锋（1972—），男，陕西，教授，博士，主要研究方向为互联网技术、网络测试与测量；张明聪（1996—），男，四川，硕士研究生，主要研究方向为未来网络、区块链；周继华（1979—），男，重庆，研究员级高级工程师，博士，主要研究方向为无线通信、移动网络及数字集群技术；赵涛（1983—），男，河南，高级工程师，硕士，主要研究方向为宽带无线通信、移动网络及无人智能集群等。

通信作者：张明聪 15280928920@163.com

0 引言

随着比特币 [1] 等加密数字货币受到广泛的关注，区块链技术也逐渐进入了公众的视野。区块链技术以其去中心化、多副本存储、数据共识难以篡改、数据可回溯等特点为改善当前众多中心化、可信度低、回溯取证困难的场景 [2,3] 提供了技术支持。

然而区块链技术也存在不足之处。由于区块链系统各节点采取副本全冗余存储，在使得区块链系统中每一完整节点能够对每条信息实现自主验证、得出结果、达成共识的同时，也带来了较为严重的存储负担。以区块链 2.0 的代表以太坊 [4] 为例，截至 2018 年 10 月 1 日，产生了超过 6500000 个区块，单节点仅同步完整的区块链需 1 TB 的存储空间。

超级账本 [5,6] 作为区块链 3.0 的代表，于 2015 年被发起，通过框架方法和专用模块，包括共识机制、存储方式、身份服务、访问控制和智能合约，致力于提供企业级区块链应用服务。然而，目前超级账本各节点同样采用副本全冗余存储，极大限制了区块链系统应用在大规模数据需要存储的场景。

本文结合纠删码（Erasure Code）[7-9] 技术，提出了一种基于纠删码的区块链系统区块文件存储模型，并将其应用到超级账本中。通过纠删码技术将超级账本的区块文件进行分块编码，每个节点根据算法仅保留部分编码块，节点通过收集一定数量的编码块进行解码即可恢复出原始数据块 [10,11]。利用纠删码技术的容错能力保障原始区块文件在超级账本网络中可用的同时，又因编码块分布式存储减少了节点因存储完整区块文件带来的存储开销，节约了存储资源，增加了超级账本区块链平台的存储扩展性。

1 相关工作

目前提出将纠删码应用于区块链领域的文献较少。文献 [12] 对将纠删码技术应用在区块链领域的可能性及潜在问题进行了分析。本文提出的区块链系统区块文件存储模型主要采用 RS 纠删码技术实现，并将其应用到超级账本区块链系统中。

1.1 区块链系统

区块链各节点之间具备开放互联的基本特点，基于 P2P 技术 [13] 使得每个节点平等获得数据并输出数据，无需任何中心化机构的审核。每个区块链节点有完整的副本数据，拥有独立的数据审计能力。所有处于同一区块链网络中的节点都遵循共识协议 [14,15]，计算并存储相同的数据，在整个区块链网络少量数据节点失效或者接收到错误数据的情况下，节点依然拥有数据校验及审计能力。

超级账本的区块文件存储系统与现有大部分区块链系统一致，如图 1 所示。

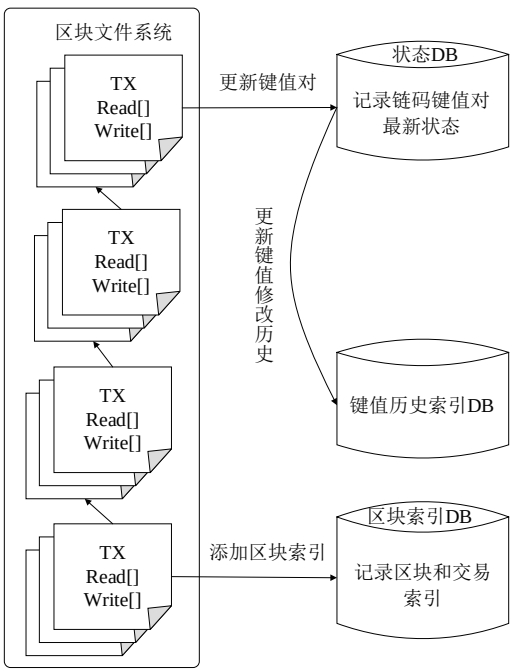


图 1 超级账本存储架构

区块文件系统负责对接收到的最新区块以文件的形式进行存储。状态数据库描述了某一时刻账本的状态，通过对新接收的区块内的交易进行模拟来更新键值。区块索引数据库用于回溯某个区块或者某笔交易的内容。事实上，在超级账本中对交易的验证及执行主要依赖于状态数据中记录的最新键值，仅有在回溯场景及节点间的区块同步场景下才需要使用到区块文件系统中记录的区块文件，然而这部分文件又占据了

节点大量的存储空间。

超级账本的交易或区块共识采用的是一种基于通信的共识方式，如图 2 所示，其区块的产出主要依赖排序服务集群。超级账本与其他区块链系统的不同之处在于准入授权机制，节点的加入需要通过注册机构进行会员注册，获得证书。通过注册的会员以组织的形式存在，即一个组织旗下可以动态地拥有多个记账节点，但背书节点的数量需要遵循超级账本指定的背书策略。

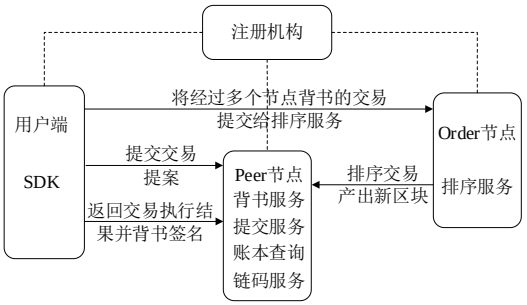


图 2 超级账本网络架构

1.2 纠删码

纠删码是一种数据保护方法，应用于网络传输中解决包丢失问题，存储系统利用它来提高存储可靠性。相比多副本全冗余存储而言，纠删码能够以更小的数据冗余度获得更高的数据可靠性。

本文使用的 RS 纠删码属于线性纠删码，如图 3 所示。线性纠删码将 k 份原始数据增加 m 份校验数据，共同形成 n 份编码数据，并能通过 n 份编码数据中的任意大于等于 k 份的编码数据还原出原始数据，如果有任意小于等于 m 份的数据失效，仍然能通过剩余的数据还原出原始数据。

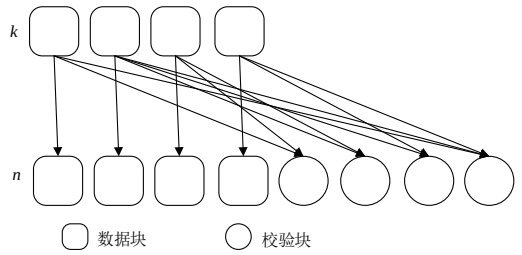
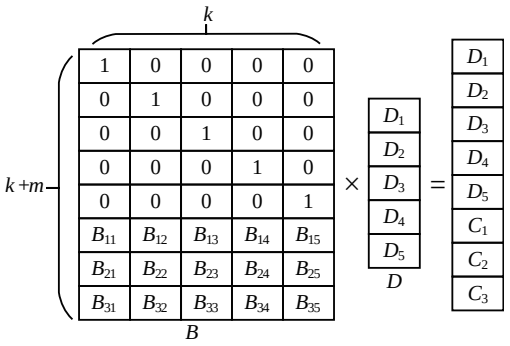
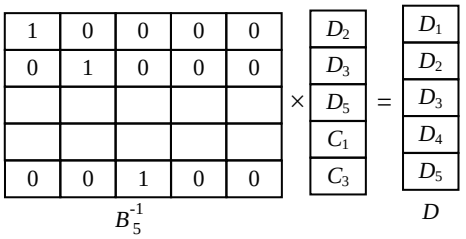


图 3 线性纠删码编码原理

RS 纠删码实际上就是利用生成矩阵与数据列向量乘积得到最后的信息列向量，当发现数据丢失，需要恢复原始数据时，利用残余可靠信息对应生成矩阵的逆矩阵与残余可靠信息进行乘积，即可得到完整的原始数据。具体编码过程及解码过程如图 4 所示。



a）编码过程



b）解码过程

图 4 RS 纠删码编码过程及解码过程

2 基于纠删码的超级账本区块文件存储模型

本文针对区块链网络中因账本膨胀带来的存储开销问题，结合纠删码技术，提出了基于纠删码的区块链系统区块文件存储模型，并将其应用在超级账本区块链平台。区块链节点将其保存的完整区块文件通过纠删码技术编码成多个编码块（如图 5 所示），每个节点仅保留部分编码块，全网节点拥有完整的编码块信息，使得全网中各个节点在尽可能减少存储空间占用的同时，又能保证原始区块文件不丢失，如图 6 所示。

2.1 区块文件编码存储

基于纠删码的区块文件存储模型中的节点在通过纠删码技术将其保存的区块文件分块编码成多个编码块时，既要使每个节点尽可能少地保留编码块，又要使全网拥有完整的编码块信息，其核心在于节点如何选择编码块进行保存，如何尽可能少地与其他节点重

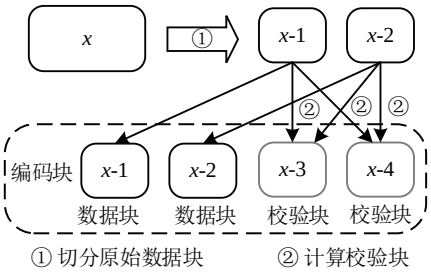


图 5 编码块获得流程

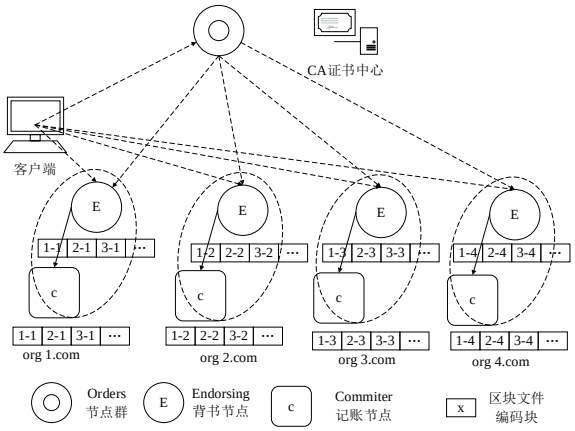


图 6 基于纠删码的存储可扩展模型

复保存。本文采用一种基于通信的方案，在超级账本区块链网络搭建完成初期产生第一个区块文件之前，排序服务集群会确认加入网络的组织的先后顺序，当有新的组织加入该网络或者退出该网络时，排序服务集群都会定期向其余组织的背书节点广播编码控制信息，如图 7 所示，包括用于指示切分区块文件的原始块数量、纠删码编码算法、编码容错率以及指示节点如何选择保留编码块的组织顺序。

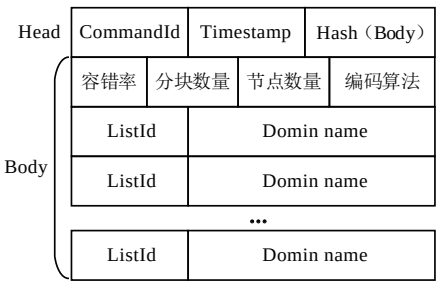


图 7 编码控制信息

第一个区块文件完成后，所有记账节点将根据最近一次接收到的编码控制信息对区块文件进行纠删码编码存储。设定纠删码容错率为 λ、原始文件切分为 k 个数据块、组织数量为 org 个的情况下，生成 $r = k\frac{l}{1-l}$ 个校验块，共 $s = r + k$ 个编码块。在完成对得到的编码块相关摘要信息的记录后，每个节点仅保存的编码块数量为

$$n_{now} = \frac{s}{org} = \frac{k}{(1-l)org} \quad (1)$$

具体保留哪一部分编码块由编码控制信息中的 $listId$ 决定。对于 $listId = i$ 的组织，其组织内的节点将保留由 $i \times n_{now} + 1$ 到 $(i+1)n_{now}$ 号的编码块。

如图 8 所示，假设此时区块链有 $node$ 个节点，每个节点都有 m 大小的区块文件满足纠删码编码条件，则通过本文区块文件存储模型，每个节点需要保存的编码块文件大小为

$$m_{node} = \frac{m}{k}s = \frac{m}{(1-\lambda)org} \quad (2)$$

全网共需要存储的编码块文件大小为

$$m_{total} = m_{node} \times node = \frac{m \times node}{(1-\lambda)org} \quad (3)$$

相较于普通超级账本区块链网络的单节点需要消耗 m 大小的存储空间，本文模型单节点存储编码块占用的存储空间随网络中组织数量的增多及容错率的降低而降低。并且在组织内仅有单节点的情况下，其全网存储编码块占用的存储空间仅受纠删码容错率影响。

2.2 区块文件解码恢复

客户端在使用基于纠删码的超级账本区块文件存储模型回溯某个区块或者查询区块内的某笔交易时，会向连接的背书节点发起查询请求。该背书节点会根据用户的查询请求访问存储该区块的区块文件，在查询得知该区块文件已被编码成块，并分布存储的情况下，节点会访问记录了该区块文件编码块的编码块索引数据库，同时向排序服务群发起区块编码块请求，期望尽可能多地收集原始数据块，如图 9 所示。

由于本文模型采用 RS 纠删码，其解码开销随解码过程中使用的校验块数量的增加而增加，因此在解码恢复时会尽可能利用原始数据块进行解码恢复。

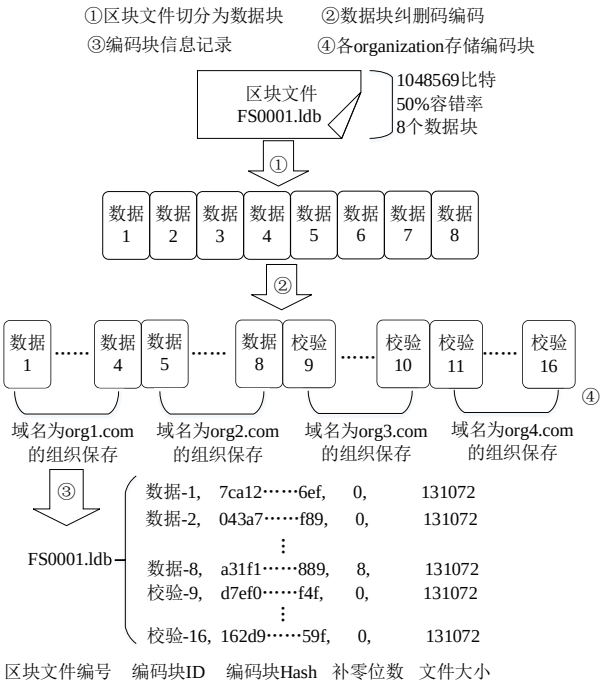


图 8 区块文件编码存储

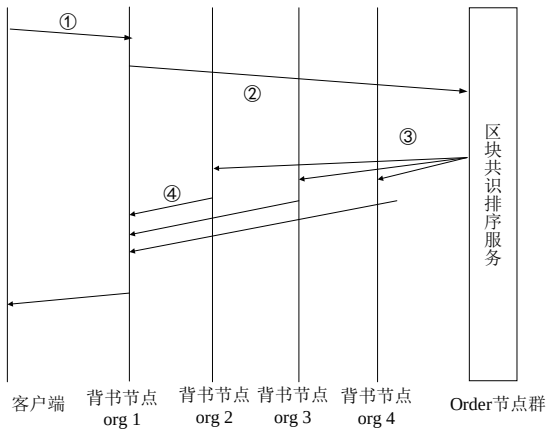


图 9 区块解码恢复

2.3 区块可扩展模型节点区块同步方案

在普通超级账本区块链网络中，节点之间同步区块采用 gossip[16] 最终一致性协议。每个节点随机向邻近节点定时发送本地账本当前快照，如最高块高度、最高块哈希值、状态树哈希值等一些账本的摘要信息，并接收来自其他节点发送的账本快照，节点间通过快照进行账本对比及区块请求，最终使得全网所有节点在较短时间内达到账本一致状态。图 10 为超级账本基于 gossip 协议实现区块同步的过程。

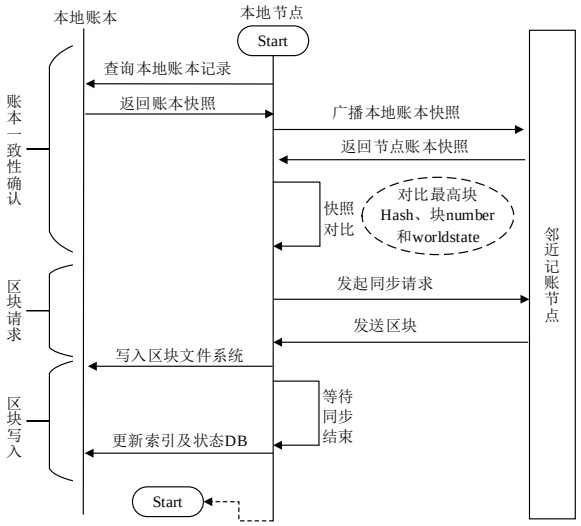


图 10 超级账本区块同步过程

然而本文模型存在超级账本节点长时间离线的情况，因为临近节点已对区块文件进行纠删码编码，仅保留了部分编码块，只能通过索引数据库及状态数据库提供节点账本快照，无法提供完整的原始区块文件供离线节点进行区块同步。在此种情况下，本文提出了适配的区块文件同步方案，如图 11 所示。

节点需要根据临近节点发送的账本快照，再结合本文提供的区块文件解码恢复方案请求恢复区块。邻近节点对待同步节点请求的区块进行判断，若未被纠删码编码分块存储，则直接传输区块信息；若请求区块所在的区块文件已被编码分块存储，则会返回消息告知待同步节点该区块所在文件已被编码分块存储，此时待同步节点根据本文提供的区块文件解码恢复方案请求恢复区块。待区块同步成功后，节点将根据本文所提供的区块文件编码存储方案对需要编码分块存储的区块文件进行纠删码编码，仅保留部分编码块。

3 实验及性能分析

3.1 实验环境

实验测试环境为两台 Intel Core i5-6400 2.70 GHz

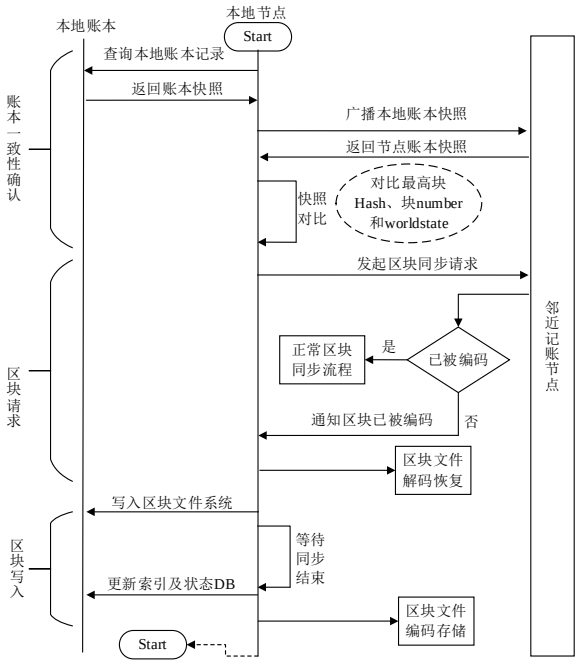


图 11 基于纠删码的超级账本区块文件存储模型区块同步过程

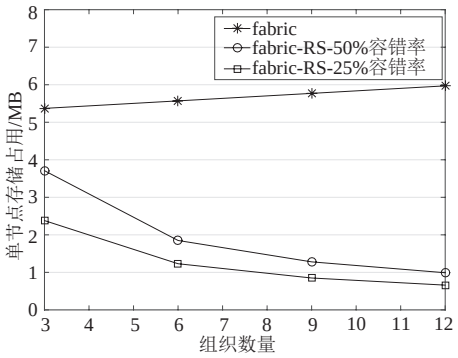
CPU、4 核、8 GB 内存、ubuntu 系统的 PC 机。每台 PC 机可以启动 12 个 docker 节点，借助 IBM 开发的开源 Hyperledger fabric v1.0 版本搭建超级账本区块链系统，并基于该版本代码进行修改，实现基于纠删码的区块链系统区块文件存储模型，搭建 fabric 存储可扩展区块链网络，简称 fabric-RS。未经修改的 Hyperlede fabric v1.0 版本简称 fabric。本文实验测试网络中的排序服务只由一个排序节点实现。

3.2 性能分析

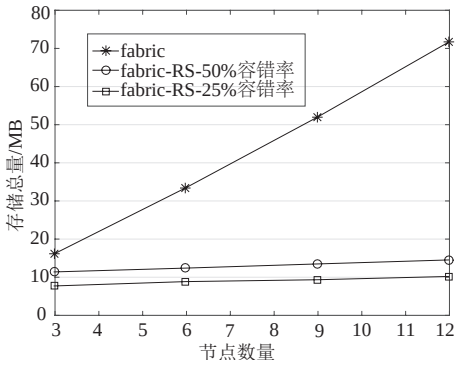
3.2.1 存储开销

在对不同组织数量、不同纠删码容错率的情况下，fabric-RS 对于存储资源的占用情况的测试中，实验固定了区块文件大小为 128 KB，原始数据块数量为 18，发起 1000 笔交易，RS 纠删码容错率为 50% 和 25%，组织数量分别为 3、6、9、12，每个组织内仅拥有一个背书节点。由于可扩展模型中的各个节点在完成区块文件纠删码编码后，需要对各个编码块的如哈希值、编码块 ID、补零位数等与解码恢复相关的信息进行存储，因此这些相对于 fabric 区块链额外的存储开销也被纳入测量范围。

图 12 a）与图 12 b）分别为在不同组织数量的情况下，fabric-RS 与 fabric 的单个节点占用的存储空间及所有节点占用的存储空间测试情况。



a）单个节点存储空间占用情况



b）总存储空间占用情况

图 12 存储空间占用测试

通过分析测试结果，可以得到以下结论。

1）fabric-RS 单节点的存储空间占用随着网络中组织数量增加而减小。当网络中组织数量为 12 时，fabric 的单节点存储占用为 fabric-RS 的 6 倍左右。这是因为组织数量增多则每个组织需要存储的编码块数量将减少。

2）fabric-RS 单节点的存储空间占用随着组织数量增加而减小的速率逐渐趋于平缓。这是因为在各组织存储编码块信息占用存储空间较小的情况下，影响节点存储空间占用的有另一重要因素，即记录编码块解码恢复相关信息占用的存储空间会增加。

3）在不考虑存储编码块解码相关信息额外占用存储空间的情况下，fabric-RS 全网总存储空间占用主要受编码容错率影响，但全网总存储占用保持恒定。例如，fabric-RS-50% 容错率在组织数为 3 时全网占用 11 MB，组织数为 12 时全网占用 14 MB，其全网总存储占用基本不受网络中组织数的影响。fabric 在组织数为 3 时全网占用 16 MB，组织数为 12 时全网占用 76 MB，其全网总存储占用受组织数的影响较大。

因此，基于纠删码的区块链系统区块文件存储模型能有效减少单节点及区块链全网对于存储空间的占用，并且在一定的组织数量范围内，其单节点的存储空间占用随组织数量增多而减少。

3.2.2 时间开销

区块文件编码分块存储所需要的时间开销主要来自于编码块开销。为了测试不同区块文件大小对编码计算造成的影响，固定需要编码的区块文件大小为 100 MB，纠删码容错率为 50%，测试的区块文件从单文件大小为 0.125 MB 至单文件大小为 100 MB 逐渐增加。图 13 为单区块文件大小不同的情况下使用 RS 纠删码编码成多个编码块，测试 100 次的平均编码开销。

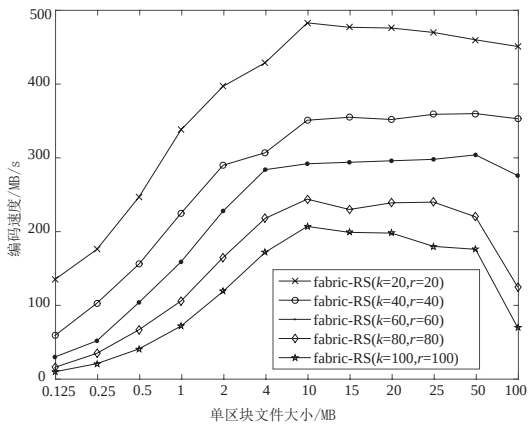


图 13 不同区块文件大小的编码开销

由图 13 可知：

1）不同区块文件大小影响着区块文件的编码速度。对大小为 100 MB 的区块文件进行编码，随着单区块文件大小的增加，其编码速度也逐渐增加。例如，单区块文件为 0.125 MB 左右时编码速度最低，这是因为在文件较小时，最终得到的编码分段文件数量较多，文件 IO 开销较大，严重影响编码速度；单区块文件为 4 MB 左右时，其编码速度基本达到峰值。

2）不同的编码组合也影响着区块文件的编码速度。在相同编码容错率的情况下，(k=20, r=20) 组合的编码速度要远高于 (k=100, r=100) 组合的编码速度。

对于 fabric 或者 fabric-RS 而言，无论是交易还是区块的共识，主要依赖状态数据库及索引数据库中保存的记录，而不依赖区块文件，因此区块文件编码所带来的时间开销并不影响区块链共识等服务。

区块文件解码恢复的时间开销主要集中于编码块收集过程中带来的通信上的时间开销以及解码计算上的时间开销。应用 RS 纠删码在恢复区块文件时连接的节点越少，下载速度越快，下载时延越短。而解码计算上的时间开销主要受解码过程中是否有校验块及校验块的数量的影响 [8]。

3.3 可靠性分析

该存储扩展模型的可靠性体现在对恶意节点的抵御能力上，即区块文件恢复时接收到错误的编码块的情况。客户端向节点请求恢复某个已被编码的区块，节点在接收到请求后发起解码恢复请求。通过对部分背书节点的程序进行修改，使其在接收到传递区块的编码块请求时始终传递错误的编码块信息。实验固定区块文件大小为 4 MB，原始数据块数量为 18，RS 纠删码容错率为 50%，组织数量为 9，每个组织内仅拥有一个背书节点。图 14 为在有组织提供错误编码块的情况下，发起区块恢复请求的节点对于区块文件的恢复情况。

通过实验可以发现，在有 4 个组织提供错误数据块的情况下，解码耗时 3 s 以上，但 fabric-RS 有能力恢复出原始区块；而在组织数量为 5 的情况下则无法恢复出原始区块，这是受到纠删码容错率的限制。在纠删码容错范围内，通过对区块链网络内的编码块信息的收集，fabric-RS 有能力恢复出原始区块数据，因此本文存储模型具有极高的可靠性保障。

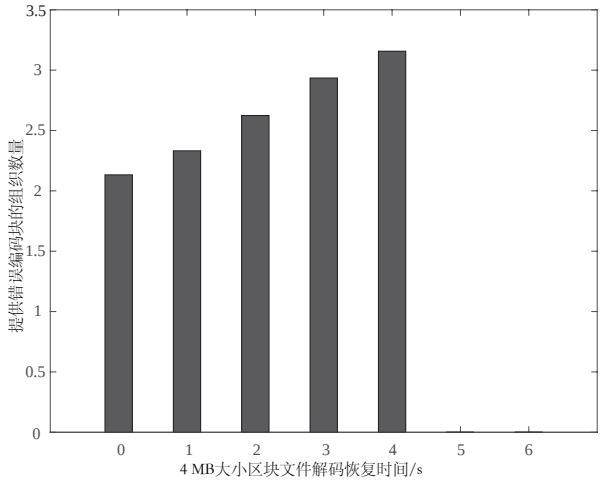


图 14 接收到错误编码块情况下的区块文件恢复

4 结束语

本文提出了基于纠删码的区块链系统区块文件存储模型，利用纠删码技术将区块文件分块编码成多个编码块，每个节点仅保留部分编码块，降低节点的存储负担；同时利用纠删码的容错能力，在部分节点失效的情况下，全网依然有能力恢复出原始区块。最后经实验测试该存储模型在多组织、多节点环境中能够安全稳定运行，并能够有效减小节点的存储负担。下一步工作将研究编码块分配策略及编码块收集策略，提出更为均衡、高效的分配及收集策略，充分利用区块链网络中的所有节点。同时，还将考虑利用该存储模型实现大文件在区块链网络上的存储，以实现对大文件信息的永久记录。（责编 潘海洋）

参考文献：

[1] NAKAMOTO S. Bitcoin: A peer-to-peer Electronic Cash System[EB/OL]. https://bitcoin.org/en/bitcoin-paper, 2018-11-10.

[2] WU Jian, GAO Li, ZHU Jingning. Digital Copyright Protection Based on Blockchain Technology[J]. Radio and Television Information, 2016(7): 60-62.
吴健，高力，朱静宁．基于区块链技术的数字版权保护 [J]. 广播电视信息，2016（7）：60-62.

[3] ZHAO Feng, ZHOU Wei. Analysis on the Protection of Digital Copyright Based on Blockchain Technology[J]. Technology and Law, 2017(1): 59-70.
赵丰，周围．基于区块链技术保护数字版权问题探析 [J]. 科技与法律，2017（1）：59-70.

[4] Ethereum Foundation. Introduction of sharding. https://github.com/ethereum/sharding/blob/develop/docs/doc.md, 2018-11-10.

[5] ANDROULAKI E, BARGER A, BORTNIKOV V, et al. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains[EB/OL]. https://arxiv.org/abs/1801.10228, 2018-11-10.

[6] Linux Foundation. Welcome to Hyperledger Fabric[EB/OL]. https://hyperledger-fabric.readthedocs.io/en/release-1.0, 2018-11-10.

[7] WICKER S B, BHARGAVA V K. Reed-Solomon Codes and Their Applications[M]. New York: Wiley-IEEE Press, 1994.

[8] LUO Xianghong, SHU Jiwu. Summary of Research for Erasure Code in Storage System[J]. Journal of Computer Research and Development, 2012, 49(1): 1-11
罗象宏，舒继武．存储系统中的纠删码研究综述 [J]. 计算机研究与发展，2012，49（1）：1-11.

[9] WANG Yijie, XU Fangliang, PEI Xiaoqiang. Research on Erasure Code-based Fault-tolerant Technology for Distributed Storage[J]. Chinese Journal of Computers, 2017, 40(1): 236-255.
王意洁，许方亮，裴晓强．分布式存储中的纠删码容错技术研究 [J]. 计算机学报，2017，40（1）：236-255.

[10] SATHIAMOORTHY M, ASTERIS M, PAPAILIOPOULOS D, et al. XORing elephants: novel erasure codes for big data[J]. VLDB Endowment, 2013, 6(5): 325-336.

[11] DENG Kai, TIAN Zhihong, MA Danyang. Research and Implementation of a Highly Reliable Distributed Storage Scheme Based on Wirehair Code[J]. Netinfo Security, 2018, 18(2): 20-26.
邓凯，田志宏，马丹阳．一种基于 wirehair 码的高可靠分布式存储方案的研究与实现 [J]. 信息网络安全，2018，18（2）：20-26.

[12] DAI Mingjun, ZHANG Shengli, WANG Hui, et al. A Low Storage Room Requirement Framework for Distributed Ledger in Blockchain[EB/OL]. https://www.researchgate.net/publication/323712562_A_Low_Storage_Room_Requirement_Framework_for_Distributed_Ledger_in_Blockchain, 2018-11-10.

[13] DONET J A D, PÉREZ-SOLÀ C, HERRERA-JOANCOMARTÍ J. The Bitcoin P2P Network[C]//Springer. 2014 International Conference on Financial Cryptography and Data Security, March 3-7, 2014, Christ Church, Barbados. Heidelberg: Springer, 2014: 87-102.

[14] GERVAIS A, KARAME G O, WÜST K, et al. On the Security and Performance of Proof of Work Blockchains[C]//ACM. 2016 ACM SIGSAC Conference on Computer and Communications Security, October 24-28, 2016, Vienna, Austria. New York: ACM, 2016: 3-16.

[15] CACHIN C, VUKOLIĆ M. Blockchain Consensus Protocols in the Wild[EB/OL]. https://arxiv.org/abs/1707.01873, 2018-11-10.

[16] GANESH A J, KERMARREC A M, MASSOULIÉ L. Peer-to-peer Membership Management for Gossip-based Protocols[J]. IEEE Transactions on Computers, 2003, 52(2): 139-149.