Journal of Computer Applications
计算机应用, 2020, 40(4): 942 - 946
ISSN 1001-9081
CODEN JYIIDU
2020-04-10
http://www.joca.cn

文章编号:1001-9081(2020)04-0942-05
DOI:10.11772/j.issn.1001-9081.2019111917

区块链增强型轻量级节点模型

赵羽龙,牛保宁*,李　鹏,樊　星
(太原理工大学 信息与计算机学院 太原 030600)
(*通信作者电子邮箱 niubaoning@tyut.edu.cn)

摘　要: 区块链固有的链式结构意味着其数据量无休止地线性增长,随着时间的积累,对单个节点的存储造成很大的压力,对整个系统的存储空间造成极大的浪费。在比特币白皮书中提出的SPV(Simplified Payment Verification)节点模型,大大减少节点对存储空间的需求,但是,它使得全节点的个数减少、压力增大,减弱了整个系统的去中心化程度,存在拒绝服务攻击、女巫攻击等安全隐患。通过对比特币区块数据进行分析,提出一种功能完整的增强型轻量级节点模型ESPV(Enhanced SPV)。ESPV把区块分为新区块和旧区块,对它们采用不同的存储管理策略。新区块以全部副本(每个节点保存一份)的方式保存以用于交易验证,用较少的存储空间代价让ESPV具有交易验证(挖矿)功能;旧区块分片存储在网络的节点中,通过分级区块分区路由表访问,在保证数据可用性和可靠性的前提下减少系统对存储空间的浪费。ESPV节点具有完整的节点功能,从而保证区块链系统的去中心化特性,增强其安全性和稳定性。实验结果表明,ESPV节点具有80%以上的交易验证率,在数据量和增长量上这些节点仅为全节点的10%,ESPV的数据可用性和可靠性有保障,适用于系统的整个生命周期。

关键词: 区块链;数据分析;分片存储;轻量级;比特币
中图分类号: TP302.1　　**文献标志码:** A

Blockchain enhanced lightweight node model

ZHAO Yulong, NIU Baoning*, LI Peng, FAN Xing
(*College of Information and Computer, Taiyuan University of Technology, Taiyuan Shanxi* 030600, *China*)

Abstract: The inherent chain structure of blockchain means that its data volume grows linearly and endlessly. Over time, it causes a lot of pressure on the storage of the single node, which greatly wastes the storage space of the whole system. The Simplified Payment Verification (SPV) node model proposed in the Bitcoin white paper greatly reduces the node's need for storage space. However, it reduces the number of nodes and increases the pressure, which weakens the decentralization of the entire system and has security risks such as denial of service attacks and witch attacks. By analyzing the Bitcoin block data, a fully functional enhanced lightweight node model Enhanced SPV (ESPV) was proposed. The block was divided into new blocks and old blocks by ESPV, and different storage management strategies were adopted for them. The new block was saved in full copy (one copy per node) for transaction verification, allowing ESPV to has transaction verification (mining) function with less storage space cost. The old block was stored in the nodes of the network in slices, and was accessed through the hierarchical block partition routing table, thereby reducing the waste of the storage space of the system under the premise of ensuring data availability and reliability. The ESPV nodes have full node functionality, thus ensuring the decentralization of the blockchain system and enhancing the security and stability of the system. The experimental results show that the ESPV nodes have more than 80% transaction verification rate, and the data volume and growth amount of these nodes are only 10% of those of all nodes. The data availability and reliability of ESPV are guaranteed, and it is applicable to the whole life cycle of the system.

Key words: blockchain; data analysis; slice storage; lightweight; Bitcoin

0　引言

区块链技术由于其不可篡改、可追溯、去中心化等特性逐渐得到广泛的关注。它最早起源于中本聪的比特币白皮书[1],在数字加密货币[1]、供应链金融[2]、数据公证[3]、资源共享[4]等领域有许多的应用场景。区块链使用链式的结构和一致性协议保证区块数据不可篡改和可追溯,但无休止的数据追加对单个节点的存储造成很大压力。采用完全副本(每个节点保存一份)的数据存储方式,对系统存储空间也造成很大浪费。

以比特币为例,截至2019年3月27日,共产生569 001个区块、17 612 496个比特币,总交易量395 438 152笔,数据总容量196.15 GB,链上认证地址49 245 944,市值近5 000亿[5]。据BitNodes[6]统计,全网使用70001协议(>=Satoshi:0.8.x)同时在线的全节点有近10 000个。单个节点需要的磁盘空间约

收稿日期: 2019-11-05;**修回日期:** 2019-11-26;**录用日期:** 2019-11-26。
基金项目: 国家重大研发计划项目(2017YFB1401000);国家自然科学基金资助项目(61572345)。
作者简介: 赵羽龙(1995—),男,山西汾阳人,硕士研究生,主要研究方向:区块链数据管理;　牛保宁(1964—),男,山西太原人,教授,博士,CCF高级会员,主要研究方向:大数据管理与分析、数据库系统性能管理、空间数据管理、多媒体数据管理、区块链数据管理;　李鹏(1994—),男,山西离石人,硕士研究生,主要研究方向:区块链应用系统;　樊星(1992—),女,山西太原人,博士研究生,CCF会员,主要研究方向:区块链数据管理。

200 GB,每个节点保存一份,那么保守的估计整个系统需要 2 PB的存储容量来存储区块数据,而且每年还以线性的速度增长。

SPV(Simplified Payment Verification)节点模型[1]不存储区块链交易数据,只具有钱包功能,可以减轻存储压力;但减弱节点间的对等性,使得系统日趋中心化。全节点存储全部的区块数据,可以验证交易(挖矿),为其他节点提供数据,具有完备的节点功能。针对存储资源较少的设备,比特币白皮书中介绍了SPV节点,它在初始化同步过程中只下载区块头,然后根据自身需要从全节点请求数据。这种节点所需的存储大小只与区块高度成线性关系,与区块大小无关。每个区块头80 B,一年仅需4.2 MB的空间,极大地减轻了存储压力。但这种节点完全依赖于全节点,无法验证交易,容易遭受拒绝服务攻击[7]、女巫攻击[8]等安全性问题。随着数据量的增加,SPV节点增多,全节点个数减少,区块链系统的去中心化程度减弱,数据安全性、稳定性下降。SPV节点减少了数据存储,但是增大了网络带宽压力,在数据存储和数据共享方面对系统没有贡献。

本文提出一种增强型轻量级节点模型ESPV(Enhanced SPV):采用完全副本方式保存新区块(最近产生的n个区块),让轻量级节点具有交易验证功能;同时把旧区块(新区块之前的区块)进行分片存储,降低数据的冗余度;并且创建分级区块分区路由表,加快数据检索的速度,保证数据的可用性。

本文的主要工作为:

1)提出增强型轻量级节点模型ESPV,使轻量级节点可以验证交易,具有完整的节点功能,增强节点间的对等关系,保障区块链系统的去中心化、稳定性和安全性

2)在确保数据的可靠性和可用性的前提下对旧区块进行分片存储,减少存储空间浪费,增强系统的可扩展性;保存完整的区块头数据,保障系统中区块数据的真实性。

3)提出适合ESPV存储特征的路由表,既提高了数据查找效率,又达到负载均衡的作用,避免全节点压力过大的问题。

1　相关工作

比特币网络中的节点模型主要包括全节点和SPV节点,它们具有不同的功能和机制。全节点[1]是第一个也是最安全的节点模型,它通过下载和验证从创世块到最近发现的区块来确保区块链的有效性,可以独立地共享数据和验证交易。SPV节点只保存区块头数据,仅可以验证支付,不能验证交易(挖矿)。验证支付时,SPV节点需要依赖从全节点拉取符合布隆过滤器条件的MerkleBlock消息[9],通过其中的Merkle树的哈希认证路径,判断目标交易是否在该块中;同时,利用区块头检测该块是否在链中已被压入足够的深度,来确认交易是否成功。Delgado-Segura等[10]对比特币的UTXO(Unspent Transaction Output)数据进行了分析,发现大部分UTXO产生于最近的少部分区块中。

在简化单节点数据存储方面,目前主要有区块修剪、副本策略、纠删码技术、共识单元等解决方案。Bitcoin-core[11]提出一种区块修剪策略,下载的区块数据构建完UTXO集后就可以删除,极大地减少节点所需的存储空间。但随着修剪策略的流行,系统中区块数据的可靠性会下降。Jia等[12]提出一种存储可扩展性模型,将区块链中的区块数据保存在一定比例的网络节点中,增加了额外的两条链,在减少存储空间的同时也增加了系统复杂性。Dai等[13]提出一种低存储要求的区块链存储框架,使用纠删码技术[14]对区块数据进行分块存储,降低单节点的存储和带宽压力,增加了节点的计算量。Xu等[15]提出共识单元的方法,让节点形成社区,在社区内自治式分片存储区块数据,但仅针对私有链,对公有链存在查询开销太大的问题。本文仅在单节点中维护分片信息,建立路由表,采样处理数据,系统复杂度和计算量都比较小,且适用于公有链。

P2P(Peer-to-Peer)网络[16]中节点具有不同的类型和数据存储状况,为了加快数据的检索,通常会建立集中式或基于DHT(Distributed Hash Table)[17]的分散式路由表。集中式的路由表适合于网络节点数目较少的情况[18],网络中的节点在搜索数据时首先向路由中心请求数据所在位置,在得到目标节点的位置后它与目标节点单独进行联系。系统的响应时间短,数据可用性强,但中心服务器压力较大,存在单点故障问题。基于DHT技术的路由表查找算法性能为$\log N$[19],进行了负载均衡,适用于较大规模的节点网络,不存在单点故障问题。它是基于内容的查找方式,可以直接找到每个小数据分片的位置,同时也都需要维护一份索引数据,导致整个索引表很大,数据更新维护的成本高。

2　增强型轻量级节点模型

由前两章内容可知,完备的节点功能包括交易验证(挖矿)、数据存储和数据共享。为了让节点功能完整,增强型轻量级节点模型的设计也从这三部分出发。SPV节点为减少数据存储不再保存完整的区块数据,也不能进行挖矿,致使其对全节点产生严重的依赖,节点之间的关系不再对等。同时,比特币中每一笔交易的输入为历史交易的输出或者为coinbase交易,即挖矿奖励。在进行交易验证时需要追溯到交易输入中的每个交易,而且大部分的交易输入都集中在后面生成的区块中。于是我们构想是否可以通过保存最近生成的部分区块数据(新区块),让轻量级节点可以验证大部分交易,具有挖矿功能。而对于旧区块,主要用途是构建完整的UTXO库,请求量较少且不会发生更改,因此,不需要全部冗余保存,可以采用数据分片技术,降低数据的冗余度,减少存储空间浪费。旧区块分片保存后,为了保证数据可用性,可以通过设计适当的路由机制,加快数据检索速度。

如图1所示,本文对高度为567 301~568 201的区块进行固定间隔采样,查询向前追溯一定块数范围时可以验证当前区块中交易的比率,即通过向前遍历区块,查找其中是否包含该区块中交易的输入。可以发现仅需保存少量最新的区块就可以验证较高比例的交易。于是本文针对比特币的交易特性和现有节点模型存在的问题,提出增强型轻量级节点模型。模型的核心思想是区别对待新区块和旧区块,对新区块进行完全副本存储,对旧区块进行分片存储,采用不同的存储策略,为轻量级节点增加交易验证的功能,提高系统在存储方面的可扩展性,维护节点之间的对等性,保障系统的去中心化程度和稳定性。

2.1　新区块存储

网络中的节点对于新区块的请求量较大,为网络带宽型。网络中的全节点需要同步最新的区块数据,参与挖矿的节点

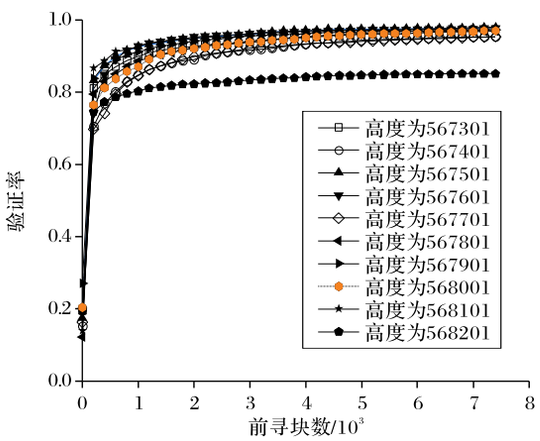


图1　交易验证率

Fig. 1　Transaction verification rate

也想尽早获取最新的区块数据，在它之上开始挖矿以更大概率地获取区块奖励。SPV 节点需要同步其区块头来完成支付验证。这是因为比特币采用链式的追加模式，每笔交易的成功需要确保其被打包在区块中，连接上链，且在该区块之上有不小于 6 块新区块的确认。

根据新区块的网络特性，ESPV 采用完全冗余的存储策略来保存新区块数据，即保存一定窗口大小的新区块数据。这样可以为其他节点提供数据共享服务，减少全节点带宽压力，增加系统的稳定性。

节点在初始化时首先查询最新的区块高度 h，然后从网络中下载区块高度为 $h-3000$ 到 h 的区块数据，同时构建 UTXO 库。在进行交易验证时如果未在 UTXO 库中发现交易的输入，那么将此交易转发给全节点。由图 1 可知这样可以验证大部分的交易，当已验证的交易数目足够多时，对应的区块大小会达到系统设定的最大阈值进而可以将它们打包成块开始挖矿计算。

目前区块交易数目正以线性的速率增长，为保证系统的可用性，ESPV 设计了相应的动态调整策略。从图 1 可知，交易的验证率随着新区块数的增长而增长，据此我们展开相应的设计。与挖矿难度值调整周期相同，ESPV 也采用每 2 周进行一次调整。比特币系统平均 10 min 生成一个区块，则每 2 016 块区块就需要进行一次调整。节点需要记录上次调整的区块高度，并以此为起点直到之后产生 2 016 块数据。在这 2 016 块中随机采样 40 块区块，如果这些区块在其前 3 000 块区块中可验证交易数低于 80% 的区块数大于 4，则为新区块存储区块数 n 加 32；如果可验证比率超过 95% 的区块数大于 32，则 n 减 32；否则 n 不发生改变。这里的 32 是由现有区块数和能保证较高比例交易验证率时所需的最新区块数按时间增长比率得出，较为合适。这些具体的参数都可以进行自定义配置，以适应不同节点的硬件环境和需求。

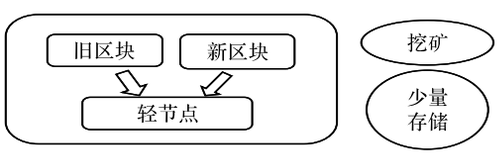


图 2 增强型轻量级节点模型

Fig. 2 Enhanced lightweight node model

2.2 旧区块存储

系统对旧区块的请求量较小，为非带宽型。只有当网络中有新的全节点加入或者重新构建全节点时，才需要拉取它们。这是因为其使用类似日志的形式，一旦生成就已经成为不会发生更改的历史数据，节点无须重复获取它们。在全节点进行验证交易时，需要从创世区块到最新产生的区块之中构建出完整的 UTXO 集，从中查找交易的输入，确保它是未花费的，余额大于等于支出花费，且验证签名确认资产所属权，所以旧区块的可靠性和可用性对于整个系统来说是十分重要的。

考虑到旧区块的访问特性，ESPV 对其采用分片存储的方式，即每个节点保存部分历史区块数据。这使得每个节点的存储压力变小，减少系统对存储空间的浪费，增加系统的可扩展性。

本文使用开源项目 Bitnodes 对比特币网络中的全节点数据进行了统计分析。截至 2019 年 4 月 10 日，近两年内比特币使用 70001 版本协议的全节点同时在线的个数中最大值为 12 770，最小值为 6 671，平均值为 9 931[6]；并且同时在线的节点个数具有一定的稳定性。由此本文对节点的可靠度进行了统计，可得图 3。在 P2P 系统中，可靠性关系[20]为：

$$a=\sum_{i=m}^{n} C_n^i p^i (1-p)^{n-i}$$

其中：a 为系统可靠度；p 为节点可靠度；r 为副本数。可解得：

$$r=\log(1-a)/\log(1-p)$$

根据 Borel 定律[21]定义低于 10^{-50} 的概率都是不可能的。故设

$$a=1-10^{-50}$$

根据图 3 可以保守估计节点可靠度 p=0.1，由此可以计算出 r 约为 1 000。

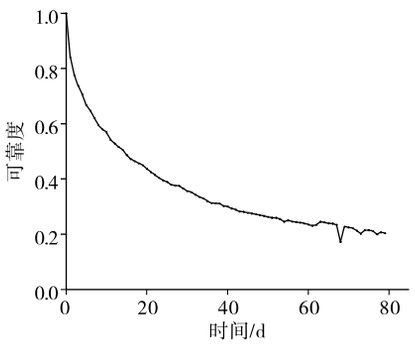


图 3 节点可靠性

Fig. 3 Node reliability

由于比特币系统中区块大小是分布不均匀的[5]，且节点通常需要获取连续的区块数据，所以进行分片时采用连续、固定存储空间大小的方式较为合适。根据现有数据存储情况进行分析，可设置一个分片的大小为 5 GB。为更好地适应节点间不同的存储空间差异，对旧区块制定不同的初始分片大小，即小、中、大 3 种，对应的分片个数为 4、8、12。节点在接收到区块时需要统计其大小，并记录系统区块分片的起始和终止区块高度，将这些高度值添加入到一个数组中，定义为分片锚定高度集，这些值将不会发生改变。每个节点在加入 P2P 网络或者进行数据的扩充和删减时都以这些固定的区块高度为界限。在节点加入网络时，节点产生一个从 0 到最新区块高度的随机数，以分片锚定高度集中离这个随机数最近的高度值作为其数据存储的起始高度，根据其可用空间的大小保存对应级别的分片数目。

假设整个系统中种子节点和矿池等性能良好、存储空间充足、稳定性强的节点运行全节点，其他硬件等条件受限的节点运行 ESPV 节点，这样系统中区块数据的可靠性和可用性具有基础的保障。目前完整的区块数据量在线性地增长，为保障系统的可用性每年需要将初始分片大小加 1，已经加入的节点向后延伸一个分片。

为了激励节点尽可能多地保存区块数据，ESPV 对存储量不同的节点请求设置不同的优先级，在节点内部通过获取节点分片属性进行请求队列排序，存储数据较多的节点优先得到请求回复。

2.3 分级区块分区路由表和链信息

结合旧区块的存储策略，ESPV 设计了新的路由机制：分级区块分区路由表。根据节点存储分片的大小，构建 4 个区块分区路由表，分别为小、中、大、全节点路由表。路由表以 Map 形式存储，key 为分片起始块号，value 是节点数组，每个节点为 ip:port。

```
Map{
    key : blocknumber
    value : [ip:port,…]
}
```

查找指定高度区块所在节点时，需要在分片锚定高度集中找到离它最近的分片起始块号 h。之后依次按小、中、大、全节点路由表查询 key 为 h 的节点列表。根据获取的节点列表长度 L，产生从 0~L 的随机数 N。把节点列表中第 N 个元素作为起始节点依次尝试连接节点：如果在小路由表中未找到连通的节点则继续向下，从中路由表中进行查找；如果找到目标节点则终止查询，向目标节点发起请求。为加快查询，将

小、中、大路由表尝试节点数的最大值分别设置为8、4、2,全节点路由表不作限制。采用直接定位的方式避免请求数据的洪泛广播,减少系统带宽压力,提高数据检索速度,保障数据可用性。根据节点存储数据量的大小,采用分级的方式从小到大进行节点访问,可以尽可能地利用不稳定节点,减少全节点和存储数据较多节点的带宽压力,进行负载均衡,避免局部热点产生。ESPV在节点加入路由表时设置了审核机制,其持续在线时长需要超过30 min,由此得到的节点相对稳定,防止数据频繁更新对系统网络带宽造成压力。节点在加入网络后可以从其他节点获取路由表信息,并且定期检测节点的连通性,更新路由表。

为验证数据的真实性,ESPV保存全部的区块头数据。节点每次从网络中获取到区块数据后将链上的区块哈希值与计算出的该区块哈希值进行比对就可以验证区块数据的真实性,维护系统数据的安全性。区块头的数据量很小,不会对节点的存储造成压力。

2.4　ESPV模块架构

ESPV使用不同的端口与网络中的节点进行数据共享。用端口1进行交易的接收和转发,端口2进行区块的请求和发送,端口3获取和共享路由信息,端口4发送和接收区块头。

各模块都有相应的功能。新旧区块的获取都需要由区块分区路由表取得目标节点地址,以快速从网络中拉取数据。通过得到的新区块构建UTXO库,用于交易的验证。为加快检索速度为UTXO构建缓存机制,将一部分最近产生的UTXO加载到内存中。旧区块为系统中的其他节点提供数据服务。区块头信息可以校验从其他节点得到的区块的真实性,保障系统安全。

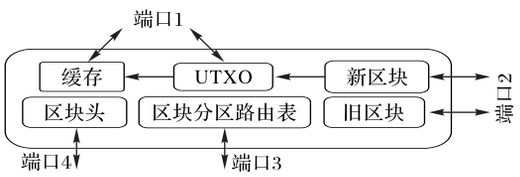


图4　ESPV模块架构

Fig. 4　ESPV module architecture

3　实验与结果分析

实验的开发环境为Intel Xeo E5-2609 v4 1.70 GHz CPU和16 GB内存的服务器。利用比特币现有区块数据进行模拟实验。

通过搭建真实节点,使用BitcoinETL[22]开源工具,对区块数据进行处理,只保留所需字段信息,获得了实验所需数据。实验过程中使用BloomFilter算法解决了超大数据量的关联查询问题。

首先,以10万块为一组,从中随机抽取100个区块,从目标区块开始向前追溯3 000块区块,查询这些区块中包含该区块交易的输入的比率,即可验证交易的比率,求均值可得图5。

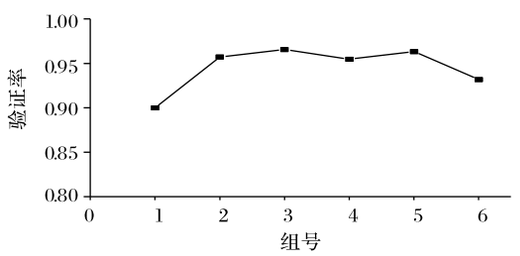


图5　总体交易验证率分布

Fig. 5　Overall transaction verification rate distribution

从图5可以看出,ESPV适用于整个比特币现有生命周期。比特币总体的交易特征是类似的,在得到数字货币后较大可能会在近期进行交易。

其次,使用最新的区块数据测试ESPV的交易验证功能。实验从高度为568 201块开始,每2 016块为一个周期,随机抽取其中10%的区块进行采样处理,计算交易验证的平均比率值,可得图6。

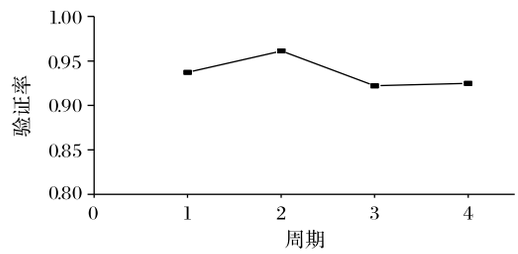


图6　交易验证率

Fig. 6　Transaction verification rate

由图6可知,ESPV对新区块采用的策略可以验证较高比率的交易,动态调整策略有效。

然后,ESPV与全节点在568 201块时的存储空间对比如图7所示。

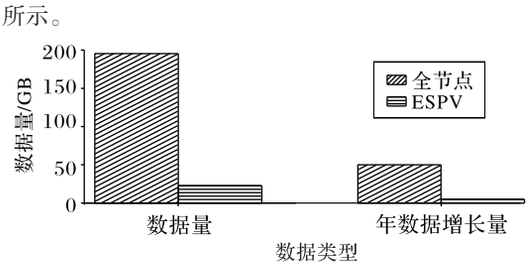


图7　存储空间对比

Fig. 7　Storage space comparison

通过分析图7,可以得出下面的结论:

1)ESPV节点比全节点所需的存储空间明显减少;

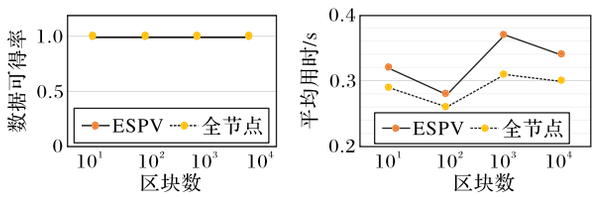
2)ESPV的数据量增长速度远小于全节点,可以符合普通PC的硬件条件,增加系统的可扩展性。

假设全网70%的节点使用ESPV节点,30%为全节点,同时在线节点约1万个,那么保守估计整个系统可以节省约1 PB数据存储空间。

最后,测试ESPV的可靠性和可用性。本文使用Java语言设计了节点对象,它具有节点类型、可靠度、IP、端口、区块段和网络带宽属性。通过创建1万个节点对象来模拟P2P节点,建立分级区块分区路由表;同时,创建1万个全节点对象,作对比实验。分别从两组节点中随机获取长度为10、100、1 000和10 000的区块数据。

```
Class Node {
    String type;
    Double reliability;
    String ipPort;
    Int[ ] blockSegment;
    Double bandwidth;
}
```

根据现有网络节点的分布和可靠性情况设置节点属性。200个种子节点(全节点)的可靠度为0.9,网络带宽为10 MB/s;剩余30%为全节点,70%为ESPV节点,可靠度都为0.1,网络带宽为0.2 MB/s~5 MB/s的随机数。在创建节点对象时给区块段属性随机赋予由真实比特币数据得到的分片锚定高度集中的值,区块大小采用现有比特币中区块的大小数据。每试错一个节点需要延时10 ms。遍历节点,为节点建立路由表。用于对比的1万个模拟全节点的带宽和可靠度与非种子节点相同,建立路由表。由实验结果可得图8。



(a) 可靠性 (b) 可用性

图8 数据可靠性和可用性

Fig. 8 Data reliability and availability

从实验结果可以看出,ESPV 节点模型和全节点的数据可得率相同都为100%,可靠性有保障。在获取区块数据的平均用时上,ESPV 节点模型略高于全节点,这是因为系统尽可能地利用不稳定节点造成延时;但P2P网络本身不稳定,较低的延时差对于单个节点的影响很小,还可以满足应用需求,能够应用于实际生产环境中。

4 结语

区块链的数据量呈线性增长模式,对单个节点的存储造成很大的压力。SPV 节点模型解决了存储问题,但它们完全依赖于全节点,使得全节点压力增大,系统的去中心化程度减弱。本文提出一个功能完备的增强型轻量级节点模型。通过对区块数据进行分析,发现保存最近的少量数据就可以验证一定量的交易,让节点具有挖矿功能。通过对网络中全节点数据进行统计分析,在保障数据可靠性和可用性的前提下,对旧区块进行随机分片存储,降低单个节点的存储压力,增加系统存储可扩展性、安全性和稳定性。为加快数据查找,设计出分级区块分区路由表,防止数据请求进行洪泛广播对网络带宽造成压力。使用链信息保证数据的真实性。为适应区块数据线性增长的模式,提出动态调整策略,保证增强型轻量级节点模型的可用性。

增强型轻量级节点模型在数据存储空间和交易验证之间进行了折中,对早期旧区块产生的UTXO支持性较差,还需要全节点进行验证和打包,未来还需要进一步优化。

参考文献(References)

[1] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system [EB/OL]. [2019-04-01]. https://www.bitcoincash.org/bitcoin.pdf.

[2] KORPELA K, HALLIKAS J, DAHLBERG T. Digital supply chain transformation toward blockchain integration [EB/OL]. [2019-04-01]. https://scholarspace.manoa.hawaii.edu/bitstream/10125/41666/1/paper0517.pdf.

[3] TURKANOVIĆ M, HÖLBL M, KOŠIČ K, et al. EduCTX: a blockchain-based higher education credit platform [J]. IEEE Access, 2018, 6: 5112-5127.

[4] CHOWDHURY M J M, COLMAN A, KABIR M A, et al. Blockchain as a notarization service for data sharing with personal data store [C]// Proceedings of the 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering. Piscataway: IEEE, 2018: 1330-1335.

[5] CoinMarketCap. Bitcoin[EB/OL]. [2019-03-27]. https://coinmarketcap.com/zh/currencies/bitcoin/.

[6] Bitnodes. 730 days nodes [EB/OL]. [2019-04-01]. https://bitnodes.earn.com/dashboard/?days=730(#nodes.)

[7] LAU F, RUBIN S H, SMITH M H, et al. Distributed denial of service attacks [C]// Proceedings of the 2000 IEEE International Conference on Systems, Man and Cybernetics. Piscataway: IEEE, 2000: 2275-2280.

[8] WANG L H. Research on Security of P2P Technology [J]. Applied Mechanics and Materials, 2014, 644-650: 2826-2829.

[9] Github.com. Bitcoin-spv[EB/OL]. [2019-04-01]. https://github.com/bitcoin-s/bitcoin-s-spv-node/blob/master/src/main/scala/org/bitcoins/spvnode/networking/PaymentActor.scala.

[10] DELGADO-SEGURA S, PÉREZ-SOLÀ C, NAVARRO-ARRIBAS G, et al. Analysis of the bitcoin UTXO set [C]// Proceedings of the 2018 International Conference on Financial Cryptography and Data Security, LNCS 10958. Berlin: Springer, 2018, 2018: 78-91.

[11] Github.com. Bitcoin [EB/OL]. [2019-04-01]. https://github.com/bitcoin/bitcoin/blob/v0.11.0/doc/release-notes.md.

[12] JIA D, XIN J, WANG Z, et al. ElasticChain: support very large blockchain by reducing data redundancy [C]// Proceedings of the 2018 Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data, LNCS 10988. Cham: Springer, 2018: 440-454.

[13] DAI M, ZHANG S, WANG H, et al. A low storage room requirement framework for distributed ledger in blockchain [J]. IEEE Access, 2018, 6: 22970-22975.

[14] LIN W K, CHIU D M, LEE Y B. Erasure code replication revisited [C]// Proceedings of the 4th International Conference on Peer-to-Peer Computing. Piscataway: IEEE, 2004: 90-97.

[15] XU Z, HAN S, CHEN L. CUB, a consensus unit-based storage scheme for blockchain system [C]// Proceedings of the IEEE 34th International Conference on Data Engineering. Piscataway: IEEE, 2018: 173-184.

[16] AMALARETHINAM D I G, BALAKRISHNAN C, CHARLES A. An improved methodology for fragment re-allocation in peer-to-peer distributed databases [C]// Proceedings of the 4th International Conference on Advances in Recent Technologies in Communication and Computing. Piscataway: IEEE, 2012: 78-81.

[17] HASSANZADEH-NAZARABADI Y, KÜPÇÜ A, ÖZKASAP Ö. Decentralized and locality aware replication method for DHT-based P2P storage systems [J]. Future Generation Computer Systems, 2018, 84: 32-46.

[18] 李志永. 高可用性P2P文件共享系统关键技术研究[D]. 武汉: 华中科技大学, 2007: 4-5. (LI Z Y. Research on key technologies of high availability P2P file sharing system [D]. Wuhan: Huazhong University of Science and Technology, 2007: 4-5.)

[19] 刘建超. 面向DHT的P2P分布式存储认证系统[D]. 沈阳:东北大学, 2010: 17. (LIU J C. DHT-oriented P2P distributed storage authentication system [D]. Shenyang: Northeastern University, 2010: 17.)

[20] 许劲斌. P2P网络存储系统的数据可靠性研究[D]. 哈尔滨:哈尔滨工程大学, 2011: 19. (XU J B. Research on data reliability in peer-to-peer network storage system [D]. Harbin: Harbin Engineering University, 2011: 19.)

[21] BOREL E. Probabilities and Life [M]. Mineola, NY: Dover Publications, 1962: 23-87.

[22] Github.com. Bitcoin-etl[EB/OL]. [2019-04-01]. https://github.com/blockchain-etl/bitcoin-etl.

This work is partially supported by the National Key Research and Development Program of China (2017YFB1401000), the National Natural Science Foundation of China (61572345).

ZHAO Yulong, born in 1995, M. S. candidate. His research interests include blockchain data management.

NIU Baoning, born in 1964, Ph. D., professor. His research interests include big data management and analysis, database system performance management, spatial data management, multimedia data management, blockchain data management.

LI Peng, born in 1994, M. S. candidate. His research interests include blockchain application system.

FAN Xing, born in 1992, Ph. D. candidate. Her research interests include blockchain data management.