以 Tagion、Filecoin 和 Celestia 为例 探索数据完整性证明

作者：Stanford Blockchain Club 来源：W-SOURCE 翻译：善欧巴，本站

## 引言

自其诞生以来，区块链和加密货币一直致力于通过提供更广泛的访问权限和去除中介来改变金融格局。随着时间的推移，web3的发展扩大了其应用场景，强调了区块链技术在创建一个创作者繁荣、用户掌控数据的互联网中的潜力。

在确保去中心化的同时，赋能终端用户的基础设施的关键部分是确保数据存储在一个有韧性、抗审查的数据库中。尽管集中式数据库方便且熟悉，但它们未能提供必要的安全保障，并且需要数据库所有者的许可，限制了全球范围的应用。

分布式数据存储系统通过构建一个存储、管理和共享数据的节点网络来满足对容错性和高韧性存储的需求。去除中央权威的需要，并以P2P的方式分发数据，增强了安全性和透明度。分布式存储系统通常基于区块链或类似技术，倾向于通过冗余和可用性来复制数据。

尽管分布式存储系统依靠未使用的存储容量，提供了安全性、数据韧性以及潜在的成本效益，但它们在法规和互操作性方面面临挑战。

对于真正开放和可访问的网络，分布式存储是必不可少的。任何存储系统的一个关键方面是如何证明数据被存储和维护的方式。对用户来说，更重要的问题是，数据如何被证明是存储和维护的？这是通过数据证明来解决的。

总体而言，我们区分了两种类型的证明：

确定性证明：允许数据持有者为特定数据创建证明，以证明其创建并匹配数据的哈希值。这种类型的证明仅公开使用数据作为输入生成的哈希。

概率证明：依赖概率来展示基础数据高度可能可用。这是一种理性地展示某一特定假设的确定性程度的证明，适用于已发布并在必要时可检索的数据。

本文其余部分将讨论三种不同系统中证明数据存储和完整性的设计选择。首先是专注于高容量和可扩展数据的Tagion，然后讨论去中心化存储网络Filecoin如何确保数据的规模化存储。最后，本文将探讨专注于存储和提供区块链数据的Celestia。

## Tagion

### 架构

Tagion是一个去中心化网络，致力于高容量交易，目标是建立一个基于技术和民主治理的独特货币系统。该项目依赖于创新的数据库架构和加密技术，以实现大规模。它不是区块链，而是一个使用DART数据库优化存储的分布式账本。Tagion的证明机制是确定性证明的一个例子。

DART数据库的核心功能是作为一个分布式哈希表，根据哈希键对数据进行存储。随着信息存储的增加，该结构自然会生成更多分支，每个分支支持多达256个组合档案和子分支。



除了类似于分布式哈希表，Tagion的基础设施也可以理解为稀疏默克尔树（SMT）。SMT是基于键值对的认证数据结构，支持标准数据库操作，如查找、插入、更新和删除。每个键值对代表一个叶子，通过递归地哈希子节点到默克尔根来派生父分支的哈希。

SMT通过使证明者能够在无需访问不相关数据元素或下载特定数据片段的情况下验证元素的存在，显著提高了效率。此外，树中值的独立性允许按任意顺序更新，而不改变树的最终结构。



Tagion的系统利用包含所有子分支哈希的根哈希，以最小的计算快速验证数据状态。为进一步增强处理能力，系统可以为特定生态系统创建子DART，类似于分片区块链。这些指定节点管理数据子集，增加吞吐量，并使网络能够根据不同应用定制，类似于应用链。

使用DART创建一个无状态系统，无需维护系统转换的完整历史记录。这意味着可以删除数据，总体上降低存储需求，并通过轻量化潜在地增加系统的去中心化。

Tagion采用HiBON（哈希不变二进制表示对象）进一步促进存储过程，确保数据在进入时保持哈希不变，简化了基于关联哈希的数据检索。哈希不变性指的是数据在处理顺序不同的情况下总是会生成相同的哈希。这是一种被验证的技术，用于加速数据库中的数据检索和写入。

通过这些机制，Tagion不仅安全地存储数据，还高效地验证其在网络中的包含和完整性。

## 数据完整性

Tagion的所有子系统都会进行所谓的随机游走，检查数据是否按需存储和提供。未通过保留验证挑战的节点将被排除在网络之外。

所有档案都包含时间戳，需要支付费用以延长存储时间。在游走过程中，系统会检查是否收到付款，如果没有，则删除数据，释放空间。

## Filecoin

Filecoin是一个去中心化存储网络，通过其原生代币Filecoin激励矿工提供存储容量。为了赚取这些奖励，矿工必须生成证明，以验证他们的存储能力。

Filecoin的基本存储单位称为扇区，具有标准大小和提供者可延长的生命周期。该设计在安全性和可用性之间进行了精心平衡。Filecoin上存储的所有用户数据都是加密的，多个副本分布在网络上，确保矿工无法访问文件内容。

矿工在Filecoin网络中的影响力与其提供的存储量成正比，这也使他们可以参与网络的共识机制。Filecoin虚拟机负责执行智能合约，并促进市场操作，如将存储提供者与用户配对。

Filecoin的架构是模块化的，允许节点根据需要操作系统的特定部分。例如，一个节点可以只作为存储节点，而不参与市场操作。

为了确保数据完整性和可用性，Filecoin依赖于两种算法：存储证明和复制证明。

## 存储证明

Filecoin中的矿工生成证明，以验证他们在任何给定时间持有数据副本。通过挑战来实现这一证明：系统向矿工提出问题，只有在拥有数据的情况下才能正确回答。

为了确保矿工不会只在提出挑战时复制数据，挑战设计为在不可预测的时间间隔内随机针对数据的不同部分。随机性和时间间隔的不确定性相结合，使得矿工在提出挑战时才获取数据变得不可能、不经济且不理性。

## 时空证明（PoSt）

Filecoin引入了时空证明（PoSt），以确保持续的存储和数据可用性。时空证明通过向矿工提出加密挑战，验证在时间间隔内的存储。只有在文件存储在指定时间框架内时，矿工才能通过挑战。

PoSt包括两种类型的挑战：

赢得时空证明（Winning PoSt）：矿工验证他们在特定时间存储了数据副本，通常是在算法选择矿工挖掘下一个区块时。短期截止时间确保他们拥有数据。

窗口时空证明（WindowPoSt）：一种重复性挑战，矿工提交证明，表明他们按要求维护了数据。只在提交数据时封存数据对矿工来说会更昂贵。

封存是复制证明算法的一部分，计算密集型，使理性矿工希望尽可能减少封存需求。

## 复制证明

封存是复制证明算法的一部分，计算密集型，鼓励矿工减少封存频率。复制证明确保用户，矿工在其物理硬件上创建并存储了唯一的副本。此证明包括：

数据本身

封存数据的矿工

封存的时间和日期

数据封存时的区块高度



要求矿工生成两种证明为用户提供了文件安全存储的保障，且只有提供实际存储的矿工才能获得奖励。由于证明过大无法上链，矿工生成零知识简洁非交互性知识论证（zkSNARK），然后提交到链上，使Filecoin成为最大的zkSNARK用户，每天生成600万到700万个证明。

总体而言，Filecoin结合了确定性证明（PoRep）和概率证明（PoSt）的方法，采用混合方法。

## Celestia

本文的第三个例子是Celestia，一个所谓的数据可用性区块链，为模块化区块链提供执行和数据存储，允许它们将核心功能外包出去。

随着以太坊Rollup的兴起，像Celestia这样的数据可用性解决方案因其提供比以太坊归档节点更便宜的Rollup交易数据存储替代方案而受欢迎。

## 证明数据可用性

与Filecoin不同，Celestia不为终端用户提供存储解决方案，而是专注于解决数据可用性问题。数据可用性确保区块链数据已正确发布。通常，区块链节点必须下载整个区块以验证可用性，这是一个资源密集型过程，可能会妨碍验证。

为了简化这一过程，Celestia采用数据可用性抽样（DAS）。此方法涉及轻节点仅下载一小部分数据，直到达到预定的置信水平。如果样本中的数据都可用，则认为数据已发布，作为数据可用性的概率证明。

其工作原理如下：

提案者创建一个数据区块。

将区块数据拆分为k×k的块，形成一个矩阵。

通过添加奇偶校验数据扩展该矩阵，使用Reed-Solomon编码创建一个2k×2k的矩阵。此类编码允许从数据的子集恢复整个数据集。

计算并组合扩展矩阵的每行和每列的独立默克尔根。

最后，将所有这些组合根的默克尔根添加到区块头中的区块数据承诺中，确认数据的可用性。



为了验证可用性，轻节点随机抽取扩展矩阵中的唯一坐标，然后向完整节点查询对应于这些坐标的默克尔证明的数据块。如果响应正确，则表示整个区块的数据有很高的概率是可用的。

节点随后将接收到的带有正确默克尔证明的数据块广播到网络的其余部分。只要抽样足够，节点可以重建整个区块，使Celestia可以更多地依赖资源有限的节点进行验证，从而有助于去中心化。

在撰写本文时，Celestia仍然很新。然而，数据可用性抽样是一项可能在Celestia之外被采用的技术，以太坊核心开发者正在讨论将其添加到协议中以帮助扩展。

## 结论

总之，分布式网络中用于存储和验证数据可用性的各种方法正在运作并被积极使用。

Tagion使用DART数据库，通过分片来提高吞吐量并支持由稀疏默克尔树保护的专门子生态系统的发展。

Filecoin的架构利用两种不同的算法，时空证明和复制证明，使矿工能够验证并证明他们可靠地存储了数据。这些证明随后以零知识证明的形式记录在链上。

Celestia作为数据可用性层，利用Reed-Solomon编码将数据块扩展为矩阵。这种结构允许轻客户端进行随机抽样以确认数据可用性，绕过下载整个数据集的需要。

随着分布式存储系统格局的不断演变，Tagion、Filecoin和Celestia各自提出了确保数据完整性、可用性和可访问性的独特策略。这些平台共同为构建支持去中心化网络的韧性数据发布和存储系统做出了重要贡献。