面向分布式同步网络归档的可组合账本

Thien-Nam Dinh
thidinh@sandia.gov 桑
迪亚国家实验室 阿尔伯克
基，新墨西哥州，美国

Nicholas Pattengale
ndpatte@sandia.gov 桑
迪亚国家实验室 阿尔伯克
基，新墨西哥州，美国

摘要
同步网络是一种高度可扩展的公证基础设施，可为历史网络数据提供防篡改的数据溯源。本文介绍了这一基础设施在网络归档中的适用性，并探讨了设想中的三个采用阶段。我们将实现其价值主张的核心机制形式化：一种在区块链支持的账本之间灵活拆分和合并加密信息的流程。最后，我们展示了初步性能结果，表明该方法能够适用于现代网络归档所需的规模。

CCS 概念
• **安全与隐私** → 分布式系统安全 .

关键词
区块链、溯源、网络存档

1 引言
为了将数字历史保存下来，留给后代，区块链技术展现出令人信服的价值：能够在离散时间维度上确保安全的数据溯源。研究人员已在时间戳标记 [4], 文档编辑 [5], 和学术投稿 [3], 等多种不同应用场景中展示了可行方案，但大规模采用仍遥不可及。我们认为，近期对同步网络 [2], 的编纂——这种区块链基础设施具有简单、可扩展且通用的特点——或许正是该领域实现临界规模所需的技术创新。一旦这种基础设施全面建成，世界各地的实体便能以密码学方式证明和验证与领域无关的数字内容的溯源，从而为所有符合规范的存档实体奠定一种可实现的可信度基础。我们的工作通过识别并编纂这一范式中的关键过程，推动了这一努力：分解和重组同步网络承诺[1]，以便在档案库之间安全地移动数据。我们将前者建模为本地账本上的拆分操作，将后者建模为包含加密元数据的本地账本上的合并操作。通过这项工作，我们确立了在动态网络归档生态系统中安全、流畅地移动数据的可能性。

2 用法
本节介绍拆分/合并过程在设想的三个采用阶段中的重要性。

2.1 数据管理
第一阶段考虑由单个网络档案组织采用。如果该组织采用分布式、大规模或其他复杂架构 [1], , 那么它就能受益于同步网络所提供的低成本、灵活的溯源机制。在这一阶段，拆分/合并过程对于在基础设施变更期间保留历史承诺是必要的。当组织变更其数据管理或数字身份架构时，该过程可使其在不损失安全性的情况下迁移原有承诺。

2.2 数据共享
第二阶段考虑多个网络档案组织的采用情况。由于不同的档案库保存着不同类型的内容，档案库集合将受益于同步网络所提供的、具有互操作性和领域无关性的溯源。在这一阶段，拆分/合并过程的作用是将先前收集的数据安全地从一个档案库传输到另一个档案库。例如，当旧档案库停止运营、新档案库出现，或两个档案库的收集兴趣有所重叠时，就可能需要进行此类数据传输。

2.3 数据溯源
第三阶段是由非档案组织采用。出于多种原因2，普通（非档案）网站可能希望通过同步网络承诺来保护其内容。在这一阶段，网站会将承诺拆分出来并展示在网站上，以便将其收集并合并到网络档案中。图1直观地展示了这一预期工作流程。

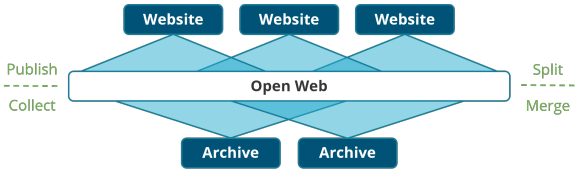

图1：溯源流程。网站将承诺拆分出来并展示在开放网络上，档案库则将其合并。

3 设计
本节将标准同步网络设置[3]中的拆分/合并过程编纂成文。在标准设置中，可归档内容由承诺 账本提供支持，这些账本由公证人保障安全，由日志维护，并由验证者检查。离散时间由区块链中区块的索引 定义。数字空间则

[1]承诺是一段用于说明某个内容溯源的元数据。

桑迪亚国家实验室是一家多任务实验室，由桑迪亚国家技术与工程解决方案有限责任公司管理和运营。该公司是霍尼韦尔国际公司的全资子公司，依据合同 DE-NA0003525 为美国能源部国家核安全管理局服务。

[2]白皮书第 4 节 [2] 归纳了若干原因。[3]白皮书第 3 节 [2]提供了定义和更多细节。

由可验证映射 构成的层级结构定义，每个映射都由一组键、值和证明组成。在这一设置下，首要任务是在保留承诺安全保障的同时，将内容从一个日志转移到另一个日志。针对这一任务，核心要求是提供一个过程，以（1）将大小为 n 的原始账本分解为一组承诺，并（2）将 m 个承诺的子集重新组合成一个新的派生账本。本节将介绍这样一个过程。

3.1 核心过程

算法1定义了用于拆分和合并默克尔树的函数，该树用于保护可验证映射[4]。对于一般的 *SplitTree* 操作，输入大小为 $O(n)$，输出传输大小为 $O(m \log n)$，耗时为 $O(m \log n)$。对于一般的 *MergeTree*操作，输入传输大小为 $O(m \log n)$，输出大小为 $O(^{m}\log_{logm}{}^{n})$，耗时为 $O(m \log n)$。拆分原始账本或合并原始承诺集时，m等于 n。

算法1 树的拆分/合并。 用于将承诺从一个可验证映射中拆分出来并合并到另一个映射中的函数。获取证明 定义于白皮书 [2] 的算法3中。

```
1: 函数 SplitTree(映射,树)
2:     返回 [[键，值，GetProof(tree，键)]
3:         对于每个键，值 ∈ map]
4: 函数MergeTree(entries)
5:     函数Recurse(items，depth)
6:         如果items。length = 1 ∧ items[0][2].length = depth 则
7:             返回 [哈希(items[0][0] + items[0][1]) [∅, ∅]]
8:         零，一 ← [ ]，[ ]
9:         对于每个 项目 ∈ 项目 执行
10:            如果 二进制(item[0])[深度] = 0 则
11:                零。追加(item)
12:            else
13:                一。追加(item)
14:        如果 零。长度 = 0则
15:            左侧 ← Recurse(零，深度 + 1)
16:        else
17:            左侧 ← [一[0][2][深度]，[∅, ∅]]
18:        如果一。长度 = 0则
19:            右侧 ← Recurse(一，深度 + 1)
20:        else
21:            右侧 ← [零[0][2][深度]，[∅, ∅]]
22:        返回 节点(
23:            哈希(left[0] + right[0]), [left,right])
24:    返回 Recurse(entries)
```

3.2 优化

我们描述了两种留待实现层面设计的优化。第一种是使用多重证明 [6] 在可验证映射之间转移批量承诺。这项优化可将传输大小降至$O(^{m}\log_{logm}{}^{n})$。第二种是扩展多重证明，以压缩部分相似的可验证映射中的相关承诺。例如，在压缩同一账本跨越 i 个区块的多个状态时，这项优化

[4]非密码学部分的拆分与合并被认为是微不足道的。

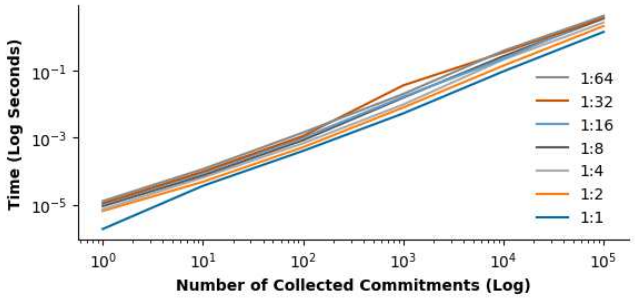

图 2：MergeTree 性能。合并 x 对承诺所耗费的时间（10 次试验的几何平均值）。每条曲线对应 m 与 n的不同比例。

可将传输大小降至$O(i\,^{m}\log_{logm}{}^{n})$，其中$m'$表示在任意两个连续步骤之间发生变化的承诺子集。

4 实现

我们将这一过程集成到现有的同步网络原型中，该原型目前包含一台公证服务器、一个期刊 Python SDK，以及一个验证器浏览器扩展。图 2 展示了一组在配备 Intel Core i9-1195OH 2.60GHz CPU 的个人笔记本电脑上进行的实验。初步结果表明，本文所述过程有望部署到现代网络存档所需的规模。[5]

5 后续方向

下一步是探索这一过程的具体应用。对于网络档案，仍需将同步网络承诺集成到高性能数据基础设施中。在开放网络方面，我们认为开发兼容的网络爬虫和浏览器扩展大有裨益。对于网站，我们预计新一代版本控制的同步网络服务器将会出现。归根结底，这些工具能否成功，将决定我们的努力能否催生安全网络存档的新局面。

参考文献

[1] Yasmin AlNoamany、Ahmed AlSum、Michele C Weigle 和 Michael L Nelson。2014年。谁链接到互联网档案馆，以及链接了什么。《国际数字图书馆期刊》14（2014），101ś115。[2] Thien-Nam Dinh、Nicholas Pattengale 和 Steven Elliott。2023年。同步网络。*arXiv* 预印本 *arXiv:2301.10733*（2023）。[3] Bela Gipp、Corinna Breitinger、Norman Meuschke 和 Joeran Beel。2017年。Crypt-Submit：介绍利用区块链对稿件提交和同行评审反馈进行安全时间戳标记。在*2017*年 *ACM/IEEE* 联合数字图书馆会议（*JCDL*）。IEEE，1ś4。[4] Bela Gipp、Norman Meuschke、Joeran Beel 和 Corinna Breitinger。2017年。利用加密货币区块链为数字文化遗产添加时间戳。*IEEE*数字图书馆技术委员会通讯（*TCDL*）13，1（2017）。[5] Muhammad Syafiq Mohd Pozi、Gopinath Muruti、Asmidar Abu Bakar、Adam Jatowt 和 Yukiko Kawai。2018年。利用区块链技术保存作者编辑历史。在第*18*届 *ACM/IEEE* 联合数字图书馆会议论文集。165ś168。[6] Lum Ramabaja 和 Arber Avdullahu。2020年。紧凑型 Merkle 多重证明。*arXiv* 预印本 *arXiv:2002.07648*（2020年）。

收稿日期：2023年2月12日

[5]互联网档案馆在2020年的一篇博客文章中称，该组织每天抓取 ≈ 750 百万个网页（http://blog.archive.org/2020/12/19/looking-back-on-2020/）。