

国网江苏省电力有限公司科技项目

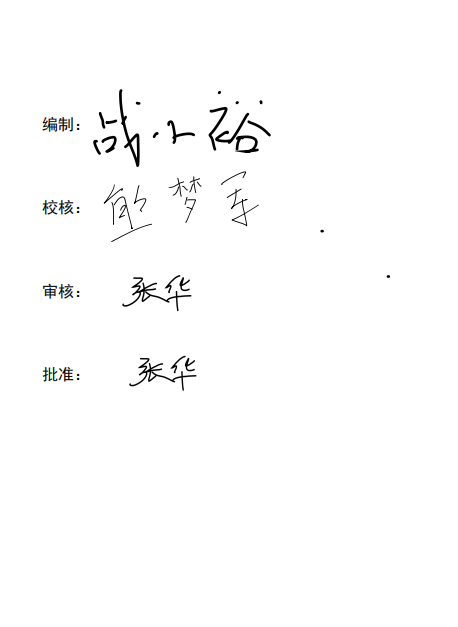
电力数据安全分发与溯源关键技术研究

课题 3 技术报告

基于区块链的电力数据安全分发与溯源技术研究报告

国网江苏营销服务中心

2022年4月



目 录

[1 概述 1](#_Toc132726206)

[1.1 研究背景 1](#_Toc132726207)

[1.2 研究目的和意义 1](#_Toc132726208)

[1.3 研究内容 2](#_Toc132726209)

[1.4 报告组织结构 3](#_Toc132726210)

[2 国内外研究水平综述 4](#_Toc132726211)

[2.1 区块链技术 4](#_Toc132726212)

[2.2 数据安全分发管控技术 5](#_Toc132726213)

[2.3 数据溯源技术 6](#_Toc132726214)

[2.4 本章小结 7](#_Toc132726215)

[3 典型电力数据分发应用场景与数据泄露风险 8](#_Toc132726216)

[3.1一般企业数据分发方式与风险 8](#_Toc132726217)

[3.1.1分发方式 8](#_Toc132726218)

[3.1.2订阅分发架构 9](#_Toc132726219)

[3.1.3主动分发架构 13](#_Toc132726220)

[3.1.4数据风险 15](#_Toc132726221)

[3.2电力数据特点与风险分析 19](#_Toc132726222)

[3.2.1电力数据特点 19](#_Toc132726223)

[3.2.2电力数据风险与数据安全法 19](#_Toc132726224)

[3.2.3 电力系统数据使用管理办法 20](#_Toc132726225)

[3.3电力数据分发场景与数据应用特征解析 26](#_Toc132726226)

[3.3.1典型分发场景 26](#_Toc132726227)

[3.3.2数据应用类型 28](#_Toc132726228)

[3.4 电力数据外泄形式与安全防控需求 30](#_Toc132726229)

[3.4.1电力数据外泄形式 30](#_Toc132726230)

[3.4.3数据可用性和防外泄之间的约束关系 52](#_Toc132726231)

[3.4.4云存储 52](#_Toc132726232)

[3.4.5云存储安全 56](#_Toc132726233)

[3.5本章小结 60](#_Toc132726234)

[4 基于区块链的电力数据共享库构建技术 61](#_Toc132726235)

[4.1一般共享数据库描述 61](#_Toc132726236)

[4.1.1 数据库结构 61](#_Toc132726237)

[4.1.2共享权限结构 62](#_Toc132726238)

[4.1.3数据联邦技术 63](#_Toc132726239)

[4.1.4信息共享平台中共享数据库的使用优势 63](#_Toc132726240)

[4.2区块链技术 64](#_Toc132726241)

[4.3.1区块链基础架构 64](#_Toc132726242)

[4.3.2共识机制 67](#_Toc132726243)

[4.3基于区块链的电力数据分发记录模式 71](#_Toc132726244)

[4.3.1数据溯源 71](#_Toc132726245)

[4.3.2数据模型 73](#_Toc132726246)

[4.3.3模型构建及上链过程 75](#_Toc132726247)

[4.3.4模型实例 77](#_Toc132726248)

[4.4基于区块链的共享方案 80](#_Toc132726249)

[4.4.1共享方案主体思想 80](#_Toc132726250)

[4.4.2共享方案溯源流程 80](#_Toc132726251)

[4.4.3文档追踪合约 82](#_Toc132726252)

[4.4.4投票合约 83](#_Toc132726253)

[4.4.5方案安全与隐私分析 85](#_Toc132726254)

[4.5本章小结 86](#_Toc132726255)

[5 基于区块链的电力数据安全分发技术研究 86](#_Toc132726256)

[5.1典型数据分发架构 86](#_Toc132726257)

[5.1.1中心化数据分发架构 86](#_Toc132726258)

[5.1.2分布式数据分发架构 88](#_Toc132726259)

[5.1.3P2P数据分发架构 89](#_Toc132726260)

[5.1.4区块链数据分发架构 91](#_Toc132726261)

[5.2电力数据分发风险分析及防护点 92](#_Toc132726262)

[5.2.1电力数据分发安全需求分析和风险描述 92](#_Toc132726263)

[5.2.2 电力数据安全分发管控不足与防护点 93](#_Toc132726264)

[5.3基于区块链的数据分发过程 95](#_Toc132726265)

[5.3.1 传统数据分发服务与区块链分发的对比 95](#_Toc132726266)

[5.3.2 电力区块链架构的选取 96](#_Toc132726267)

[5.3.3 基于区块链分发的主体 99](#_Toc132726268)

[5.3.4 分发中用到的智能合约 100](#_Toc132726269)

[5.4基于区块链的数据流转过程 101](#_Toc132726270)

[5.4.1数据流转过程中的数据信息 101](#_Toc132726271)

[5.4.2 数据分发的一个典型过程 102](#_Toc132726272)

[5.4.3 转发过程中链上记录的信息 102](#_Toc132726273)

[5.5本章小结 103](#_Toc132726274)

[6 基于加密的数据安全分发技术研究 104](#_Toc132726275)

[6.1常见加密方式 104](#_Toc132726276)

[6.1.1加密的概念流程 104](#_Toc132726277)

[6.1.2对称加密体制算法 105](#_Toc132726278)

[6.1.3非对称加密体制算法 109](#_Toc132726279)

[6.1.4不同加密算法的比较 112](#_Toc132726280)

[6.2常见的数字签名算法 115](#_Toc132726281)

[6.2.1数字签名的相关概念 115](#_Toc132726282)

[6.2.2数字签名的过程及应用 116](#_Toc132726283)

[6.2.3数字签名的常见哈希算法 118](#_Toc132726284)

[6.3基于加密的数据流转图 122](#_Toc132726285)

[6.3.1分发过程介绍 122](#_Toc132726286)

[6.3.2数据流转过程中的加密应用 122](#_Toc132726287)

[6.3.3数据流转过程中的签名应用 123](#_Toc132726288)

[6.4基于明文和密文的两种数据安全分发方案 124](#_Toc132726289)

[6.4.1两种方案流程 124](#_Toc132726290)

[6.4.2两种方案的不同点 126](#_Toc132726291)

[6.5基于区块链的数据溯源方案 126](#_Toc132726292)

[6.5.1常见溯源方案 126](#_Toc132726293)

[6.5.2溯源场景及方法 127](#_Toc132726294)

[6.6 本章小结 131](#_Toc132726295)

[7 结论 133](#_Toc132726296)

概述

研究背景

2020年国家将“数据”定位为新型生产要素，这就对数据分发交互中的安全性提出了更高的要求，需要保证数据分发过程的安全性、数据的可用性和发生数据泄露问题的可溯源。电网作为国家关键基础设施，服务范围涉及各行各业，其安全稳定运营关系国家安全，是国家关注的重点。随着我国工业和居民生活用电负荷持续快速增长，如何高质量满足供电需求对电力系统提出了新的要求。与此相应，电力数据在系统各环节流转面临的风险也与日俱增。数据共享度的提高，相同数据分发给多个接收者的场景日益多见，现有方法对多个数据接收者基于所接收的数据，存在数据泄露和无法溯源的风险。

区块链技术是利用块链式数据结构来验证与存储数据、利用分布式节点共识算法来生成和更新数据、利用密码学的方式保证数据传输和访问的安全、利用由自动化脚本代码组成的智能合约来编程和操作数据的一种全新的分布式基础架构与计算方式。技术实现上，区块链具备透明、可追溯、不可篡改和信息可信共享等特性，可以解决各个领域内数据可信传递与共享的问题。同时，区块链具备进行自主编写智能合约实现业务逻辑的功能，因此区块链技术将会给未来整个互联网甚至各行各业带来巨大的发展。

研究目的和意义

近年来，数据的重要性已得到国家和社会的高度关注，各行业积极推进数字化转型，发展数字经济，伴随各行各业信息化水平的不断提升，数据分发交互变得日趋频繁， 2020年国家将“数据”定位为新型生产要素，这就对数据分发交互中的安全性提出了更高的要求，需要保证数据分发过程的安全性、数据的可用性和发生数据泄露问题的可溯源。

电力数据包含反映社会生活和生产的居民和企业真实用电数据，具有广泛的应用场景和深层的使用价值。然而，随着电力系统信息化和电力大数据应用的不断深入，电力数据分发日趋频繁，分发场景多样，不同分发对象和分发场景下电力数据泄露形式复杂，泄露风险激增，电力数据的安全分发和发生数据泄露后的溯源问责，已成为提升行业数据安全管理能力，发展数字经济，助力公司建设中国特色国际领先的能源互联网企业的重要环节。针对上述问题，本课题开展电力数据分发应用场景与数据泄露风险形成机理研究；研究基于区块链的数据分发与溯源技术；结合典型应用场景，实现电力数据安全分发管理，支撑电力数据泄露后的溯源问责。

本项目研究成果的推广和应用具有以下几方面的重要意义：

一是通过开展电力数据分发应用场景与数据泄露风险形成机理研究，梳理电力数据应用的主要类型、数据分发场景，分发中数据泄露形式和风险成因，形成数据分发管控需求。通过解析各类电力数据分发应用场景和数据非授权泄露形式，以及对数据可用性的考量，制定电力数据安全分发与溯源约束，作为行业内电力数据安全分发与溯源规范参考。

二是开展基于区块链的数据分发与溯源技术研究，解决电力数据分发中数据泄露盲点，提升数据分发的安全性和溯源能力。针对各类典型电力数据分发应用场景，研究基于区块链等技术的电力安全分发和数据溯源方法，构建电力数据安全分发与溯源架构，兼顾数据可用性的同时支撑典型电力数据分发应用。

研究内容

本报告的主要研究内容主要包括以下3个方面：

（1）基于区块链的电力数据共享库构建技术

构建电力数据的区块链数据共享系统，以区块链的分布式特性为基础，确定可用于溯源的数据内容类型，设计电力数据在分发时的记录模式，作为后续安全分发的基本单元，保证数据分发的不可篡改和可溯源性，实现数据泄露的溯源问责。

（2）基于区块链的电力数据安全分发技术

基于区块链的不可篡改特性，记录数据分发的完整流程与凭据，并保存于区块链存储结构中，作为数据泄露时的溯源依据，实现对数据发送方和数据接收方的有效监管，作为基本流程结合加密技术，保证电力数据分发的安全性以及数据分发过程可溯源。

（3）基于加密的数据安全分发技术

基于所构建区块链数据共享库作为记录单元，结合数据安全分发流程，利用加密、签名等数据保护技术，在保证数据可用性的前提下，防止数据在传输过程中遭到泄露以及篡改，为电力数据安全分发提供技术保障。

## 报告组织结构

本报告共分为六个章节：第一章节主要介绍课题的研究背景、研究目的和意义、研究内容等基础内容。第二章节对现阶段区块链技术，数据安全分发技术，数据溯源等做了详细介绍。第三章节研究了典型电力数据分发应用场景和数据泄露风险分析。第四章节围绕电力数据共享库的概念展开讨论。第五章节研究了电力数据安全分发技术，其中包括了数据分发的流程与链上数据记录等信息。第六章研究了基于加密的数据分发场景分析，重点介绍了国密算法的应用。第七章对全文所做工作进行一个概括与归纳。

国内外研究水平综述

区块链技术

区块链技术是利用块链式数据结构来验证与存储数据、利用分布式节点共识算法来生成和更新数据、利用密码学的方式保证数据传输和访问的安全、利用由自动化脚本代码组成的智能合约来编程和操作数据的一种全新的分布式基础架构与计算方式。

技术实现上，区块链具备安全，隐私，透明，可追溯，不可篡改和信息可信共享等特性，所以恰好可以解决各个领域内数据可信传递与共享的问题。同时，目前区块链可以进行自主编写智能合约实现业务逻辑的功能，因此区块链技术将会给未来整个互联网甚至各行各业带来巨大的发展。

区块链技术具有去中心化、信任机制、公开透明、时序不可篡改和可追溯性等五个典型特征。

（1）去中心化

区块链的去中心化指的是存储在其上的数据是分布在不同的节点，因为区块链本一个分布式的数据库。这不同于以往的存储方式，只有一个或几个主要节点存储数据。这样参与的各方都保有数据而不是一个统一的机构去管理，使得中心被消除掉。去中心化解决了以下几个问题：

提高容错性。去中心化系统的容错能力更强，因为它不是只依靠一个或几个节点工作的，当一个节点被攻击失去工作能力后，其他节点仍可以正常工作保证系统的正常运行。

抗攻击性。如果去攻击一个去中心化系统可能需要很高的成本，因为攻击一个中心化的系统只需将其中心破坏即可，而去中心化的系统需要将其部分系统的节点都攻击，这是需要消耗大量资源的，最终可能会得不偿失。

抗勾结性。在去中心化系统中各个参与者都是平等的，要想勾结在一起很难实现，除非超过半数的人勾结在一起。而传统的中心化系统，上级可以为了自己的利益勾结在一起已侵害下级的权益。

（2）信任机制

区块链的信任机制的实现是依靠一些密码学原理与算法的，通过这些技术是系统实现参与者在不需要第三方中介的情况下，达成共识，建立信任的关系。

（3）公开透明

区块链的公开透明体现在只要通过验证，任何节点都可以加入区块链网络，在其上进行账本的查询审计，任何人都可以知道具体发生了什么。

（4）不可篡改

区块链采用密码学算法和共识机制保障区块链的不可篡改。如果有人想恶意破坏账本修改数据，那他必须耗费足够的资源修改一半以上的节点中的账本数据才行，显然这是很难做到的。

（5）可追溯性

区块链采用带有时间戳的链式区块结构存储数据，具有极强的可追溯性和可验证性。后面的区块必须识别前面区块的哈希值才能将自己加到链后面，因此在区块链上的任意一条数据都可以通过其链式结构查询其根源，直到创世区块。

数据安全分发管控技术

在政策层面，全球各国政府近年来对数据共享和数据利用都高度重视，不断推出各种政策法规，以推进国内和区域间数据管理和数据共享的发展。2002年欧盟发布《布加勒斯特宣言——迈向信息社会：原则、战略和优先行动》，提出了公共数据共享的指导思想，为欧盟的科学数据共享提供了依据。在2005年，联合国通过了《全球科学信息共有倡议》（The Global Commons For Science Initiative），倡议人们促进科学数据的合理传播和合作利用，并由联合国经济与社会事务署推出服务于全球的数据管理系统（UNDESA）。在 2007年由澳大利亚政府制定的《澳大利亚科学数据管理文件》中，涉及科学数据共享的内容包括科研责任和科研数据管理两部分，为学术界提供了普遍可接受的全面框架，对数据共享提出了明确的要求。

在技术方面，1979年，Shamir提出第一个秘密共享方案，称为(k,n)门限方案，其核心思想是将要共享的秘密数据D分为n块，任意k块或者k块以上组合就可以重构该秘密数据，但是任意k−1块或者k−1块以下组合均得不到关于数据D的任何信息。2010年，Yu 等人在半诚实的服务器模型下实现了细粒度的数据共享，但是它不能实现灵活的数据共享，而且随着用户数量的增加，创建文件、用户授权和撤销的复杂性会线性增大，这就使得加密和用户密钥生成不适用于一些特定的应用场景，比如移动云环境。2010年，Li等人提出一种新的访问控制框架来管理云环境下的私人健康记录，方案使用ABE技术对私人健康记录进行加密，将系统分为多个安全域来减少密钥分发的复杂性，同时还支持灵活、按需的用户权限撤销。2015年，Shao等人基于在线–离线基于属性的代理重加密技术和转化密钥技术提出一个移动云环境下的数据共享方案，它允许数据共享者外包一部分解密权限，但是移动用户的离线计算开销很大。现有的大多数数据共享方案都是利用基于属性的加密技术(ABE)构造，但是它们中的大部分并没有考虑或实现可追踪性。因此，Ahuja等人于2016年提出一个可追踪的基于属性的签密方案(ABS)来实现数据共享。2017年，Sookhak等人基于云环境提出一个车载自组网中(VANETs)不同车辆间安全地共享私密数据的方案，方案基于双线性对构造，该方案的提出为许多移动云环境下的数据共享方案提供了设计思路。以上所有方案均表明：现有的大多数数据共享方案在安全与高效方面存在不足，特别是在隐私安全越来越受到业界重视的条件下，安全成为一大隐患，因此，区块链技术开始广泛与数据可信共享结合。

数据溯源技术

对数据进行溯源追踪，是用户对委托到云存储的数据安全保护的重要需求。在数据外包到不完全可信的云服务提供商（CSP）后，用户对数据脱离自己的控制的担忧，自然引发了对云上数据进行溯源追踪的需求。

首先提出的方法就是基于日志记录的溯源追踪。利用 CSP 提供的系统日志服务，从日志中提取数据操作事件，实现对数据的追踪。这一方法本质上是服务器的系统日志分析，目标是提取数据操作记录，且被 Amazon 等当前大多数主流云服务提供商普遍支持。但这一方法存在两大问题：一是系统日志记录条目繁多、数据量巨大，而其中的数据操作记录较少，会造成事件提取操作的效率低下；二是系统日志的高度结构化，易被篡改伪造，作为溯源追踪的证据来说，存在可靠性的问题；如果生成系统日志的 CSP 不完全可信，那么日志甚至存在真实性的问题。

为了使用户能够像使用本地设备一样使用云存储，发展云存储的公开溯源至关重要，这使得用户可以求助于第三方溯源者（TPA）对数据进行验证而无需担心，但同时溯源过程也不应该为用户的数据隐私带来漏洞，现有的研究基于此提出了一个支持隐私保护的公开溯源安全云存储模型，并且对 TPA 进行了扩展使其能够同时高效的为多个用户执行溯源。利用环签名计算溯源所需的验证信息，共享数据中每个数据块上签名者的身份信息作为隐私对第三方溯源者不可见，第三方溯源能在不检索整个文件的情况下验证数据的完整性。针对云存储中多个数据中心的多台服务器上保存的副本可能不一致的情况，研究者们提出了一种有效的多副本可证明数据持有（PDP）方案，该方案中利用完全同态加密（FH）生成数据块副本，数据拥有者会先对数据分块加密后再利用同态加密生成块标签，并且该方案支持数据拥有者对数据块的动态更新操作和第三方溯源。

在大多数研究中，为了使溯源结果公平公正，使用第三方溯源是一种可行的选择。但同时也带来了问题，如何找到一个用户和云存储提供方都信任的第三方机构是一个挑战，因为在现实中完全可信的第三方几乎不存在。但是随着 2009 年比特币首次亮相，其底层的区块链技术引起了学术界的广泛关注。区块链技术去中心化的分布式架构和去信任化的运行机制使得建立一个不依赖于可信第三方的去中心化溯源架构成为可能。

综上，已有的数据进行溯源追踪方法的研究主要在于对数据完整性的溯源以及获得不可抵赖的溯源证据，如真实的不可篡改的操作记录。但现有的方法为了实现溯源证据的不可抵赖性，都不可避免的引入了 CSP 的协助产生操作记录。用户对 CSP信任度的担忧，是上述方法均无法完全避免的问题。另外，目前对数据的溯源追踪，出现了溯源证据海量，溯源效率不足的隐患。而区块链技术提供了不可篡改与去中心化两大特点，对溯源证据的不可抵赖性与溯源效率提高提供了新的研究思路。

本章小结

本章节针对研究相关内容，对国内外已有的研究成果进行了分析总结，包括区块链技术、数据安全分发管控技术、数据溯源技术等的国内外研究现状，通过对国内外相关技术的研究现状进行研究和了解，为项目工作的开展奠定良好的基础。

典型电力数据分发应用场景与数据泄露风险

3.1一般企业数据分发方式与风险

3.1.1 分发方式

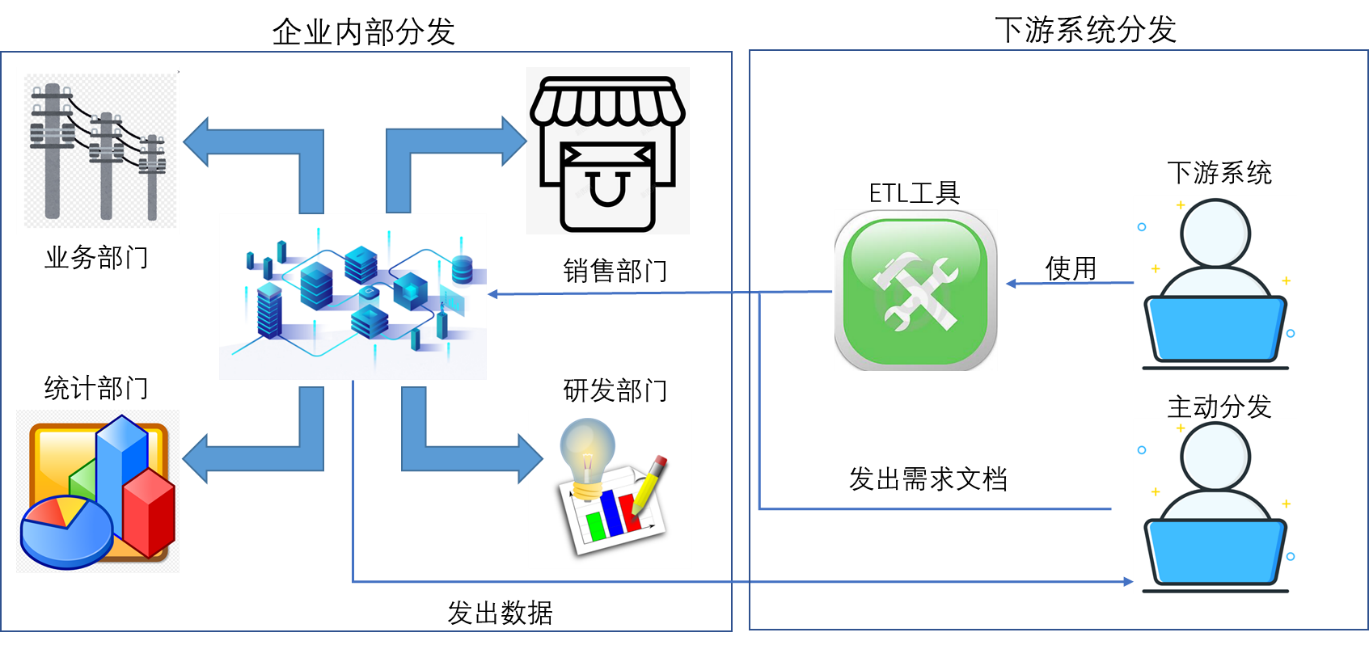
（1） 数据中心方式：各部门通过数据中心直接或间接地获取数据。

（2） 被动方式：数据应用系统主动通过ETL（数据抽取、数据转换和数据加载）工具读取数据表或数据文件。适用于数据需求时效高、比较紧急的应用。由于ETL工具直连数据库，直接访问数据库数据的方式会影响操作型数据仓储的性能，使用此种方式需要被严格限制。

1） ETL数据集成技术：该技术能共享数据信息，实现业务数据库和整个平台之间的信息流通，并通过访问权限实现信息共享。针对数据集成特征上的差异可以采用不同集成技术。ETL的集成三要素分别为抽取（extract）、转换（transform）、加载（load）。数据抽取：将用户所需的数据在业务数据库中抽取。数据转换：按照共享数据格式要求对抽取的数据进行转换，对不一致、错误的数据整理加工，以便在共享数据库中存储。数据加载：将转换完成的数据加载到共享数据库中实现共享。

2） 基于中间件的数据集成技术：部分业务数据库由于安全防护因素不能直接连接，在此条件下可以采用中间件数据集成技术。业务源系统将共享数据转化为文件形式，通过中间件将数据信息传递给共享平台。集成系统可以划分为中间件、接口适配器、数据转化组件等，根据业务系统的实际需求可添加文件适配器、消息适配器等不同数据组件。中间件通过接口适配器与数据源实现交互。转换组件在共享平台上根据数据管理标准，将业务源数据进行转化处理后再入库，避免出现数据信息丢失。

（3） 主动方式：汇总各下游系统的需求清单，根据下游应用的需求下档数据文件，并同时生成校验文件，供下游应用检查和校验。

图 3-1 分发方式

3.1.2 订阅分发架构

数据订阅分发技术实现数据的交换和集成，以便将数据通过一定方式发布到目的数据库。它能提供跨数据库类型、跨数据库版本、跨操作系统平台的订阅分发功能，通过将源数据库的某些主题信息发布到一个公共的目录服务器，使有需求的订阅端可从目录服务器获得主题，并订阅主题。数据订阅分发技术可以确保分布在不同地点的数据库自动同步更新并且保持数据的一致性。订阅分发体系结构如图3-2所示：

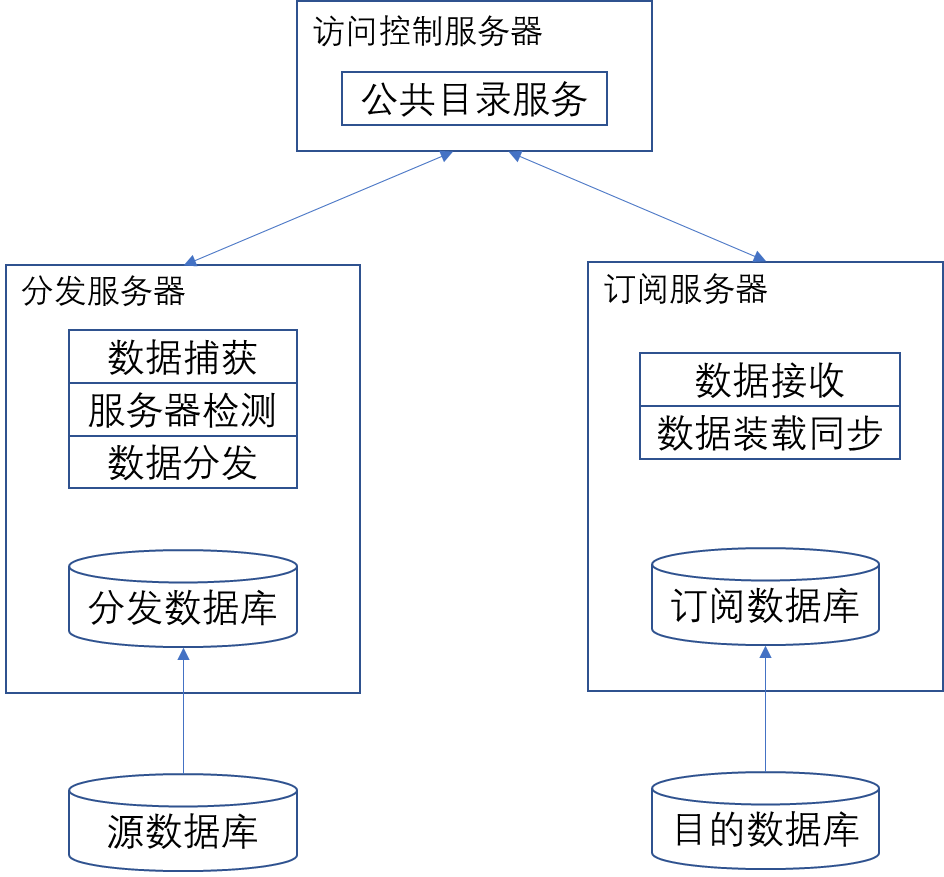


图 3-2 订阅分发体系

数据订阅分发技术的工作原理是：访问控制服务器提供集中式服务，将描述数据源的元数据（共享数据模型）发布到公共服务目录，供各个订阅服务器来订阅。分发服务器通常直接连接源数据库服务器，接收源数据的所有改变并保存这些改变，再把这些改变分发给订阅者。订阅服务器通常是数据的接收者，负责接收并同步分发服务器推送过来的变更数据。在数据订阅分发技术中，最关键的是变化数据捕获技术CDC。

CDC能捕获发生变化的数据到其他数据库或数据源。CDC能够识别自上次提取之后发生变化的数据，在对源表进行增、删、改操作的同时就可以进行提取。CDC将变化的数据保存在变更表中，然后以一种可控的方式利用视图提供给目标数据库。当前CDC有２种捕获变化数据方式，即同步CDC和异步CDC。异步CDC又分为３种模式：异步HotLog模式、异步分布式HotLog模式和异步AutoLog模式。

同步CDC模式是通过在源数据库上建立触发器的方式来获取增量数据，因此可以做到实时抽取（对源系统性能影响较大，而且一般同步CDC只适用于同一数据库中）。异步模式需要通过解析日志文件实现变化数据的捕获和提取，会有一段时间延迟（对源系统影响较小），但实施的成本比较高且技术的实现难度较大。根据企业数据共享和交换的实际需求，采用如图３-3所示的数据共享和交换模式。

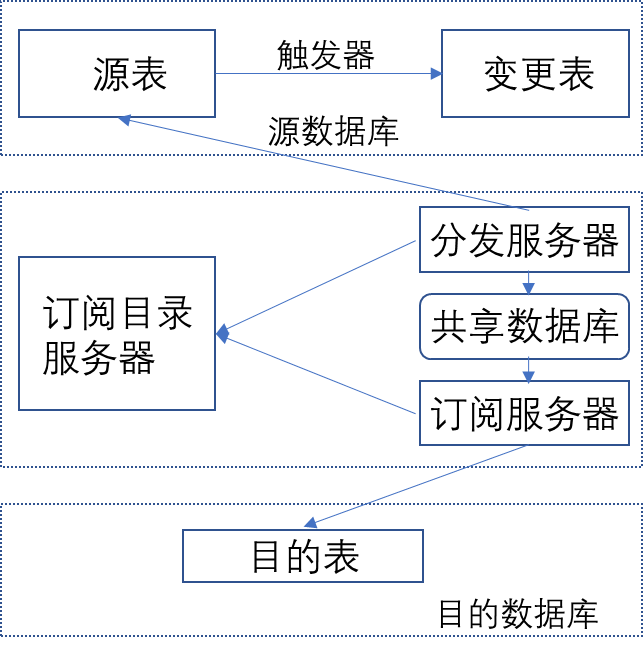


图 3-3 数据共享和交换模式

数据共享和交换的步骤如下：

（1） 源表在发生增、删、改操作时，唤醒触发器，写入变更表。

（2） 分发服务器读取分发目录表（见表3-1），然后到相应的源数据库的变更表和源表中读取数据，之后再写入共享数据库。共享数据库的数据表除了记录业务信息外，还增加了“更新时间”和“更新标记”两个字段（见表3-2），每次分发服务器写入共享库时，都会将更新时间置为当前时间，并将变更表的更新标记写入共享数据库。在分发服务器完成一次分发后，将本次分发涉及的变更表中的记录删除。

表 3-1 分发目录表实例

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 源数据库ip | 源数据库登录信息 | 源表 | 更新表 | 共享数据库ip | 共享数据库登录信息 | 共享数据库表 |
| 10.0.5.1 | XXX | staff | Staff\_change | 10.0.7.1 | XXX | gy\_staff |
| 10.0.5.2 | XXX | client | client\_change | 10.0.7.1 | XXX | gy\_client |
| … | … | … | … | … | … | … |

表 3-2 共享数据库示例

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 职工号 | 姓名 | … | 更新标记 | 更新日期和时间 |
| 000001 | 张三 | … | U | 2021-7-9T15：30：01 |
| 000002 | 李四 | … | I | 2021-7-9T15：34：01 |
| 000003 | 王五 | … | D | 2021-7-9T15：37：01 |

（3） 订阅服务器定期读取订阅目录表（见表3-3），根据订阅目录表上记录的相应共享数据表的同步时间戳，读取共享数据表中变更时间大于等于此时间戳的记录，然后根据更新标志对目标数据库的相应记录进行操作。

表 3-3 订阅目录表示例

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 共享数据库IP | 共享数据库登录信息 | 共享数据库 | 目的数据库IP | 目的数据库登录信息 | 目的数据库表 | 同步日期和时间 |
| 10.0.7.1 | XXX | gy\_staff | 10.0.8.1 | XXX | ps\_staff | 2021-7-9T15：30：01 |
| 10.0.7.2 | XXX | gy\_client | 10.0.8.1 | XXX | ps\_client | 2021-7-9T15：35：01 |
| … | … | … | … | … | … | … |

（4） 将本次读取变更记录中的最大变更时间回写到订阅目录服务上相应记录的更新时间戳。

3.1.3 主动分发架构

主要应用场景包括：不适合使用ETL直接访问数据或数据文件的下游管理系统、对时效要求不高的下游数据集市等。首先汇总各下游应用系统的需求清单，根据下游应用的需求下档数据文件，并同时生成校验文件，供下游应用检查和校验，主要可以实现以下几种分发方式：

（1） 定制字段分发方式。该方式支持下游应用系统对数据的特别需求，根据下游需求定制所要下发的数据，定制维度为字段级，可以灵活定制某一表中的某些字段。

（2） 定制数据分发方式。该方式也支持下游对数据的特别需求，根据下游需求定制下发数据，定制维度为数据字段级，可根据下游需求筛选下游所需要的数据。

（3） 单表数据分发方式。该方式支持对单表的全量或增量下载方式，根据各个下游系统需求，共享下档文件以最大化减少下档文件对系统资源的占用，可减轻系统压力。

分发类型分为直接转发和收齐转发。直接转发方式生成下档文件即转发，不用等到该系统的下档文件全部下档后再下发。收齐转发方式是等该系统的下档文件全部下档完成之后下发。

以上所述方法，可实现同一系统对不同表的不同定制需求。根据下游需求，对同一下游系统的数据需求可同时实现定制字段分发方式、定制数据分发方式、单表数据分发方式三种分发方式的支持，支持系统间并发执行及系统内并发。

如图3-4所示是一种并行数据下发方式。首先采用三级并发的方式将数据文件装入数据库中，初始化各个状态表；其次获取各个系统的转发方式，转发方式为直接转发和收齐转发；然后根据各下游系统对应的转发方式，下发相应的数据文件，循环读取状态表，直至所有数据下发完成；其中对于有相同需求的系统的数据文件存储在同一个目录下，这样可以有效的节省存储空间。数据文件下档及下发时采用系统间及不同数据文件间的双并发模式，可以大大提高数据文件的下发速度，有效节省系统运行时间。

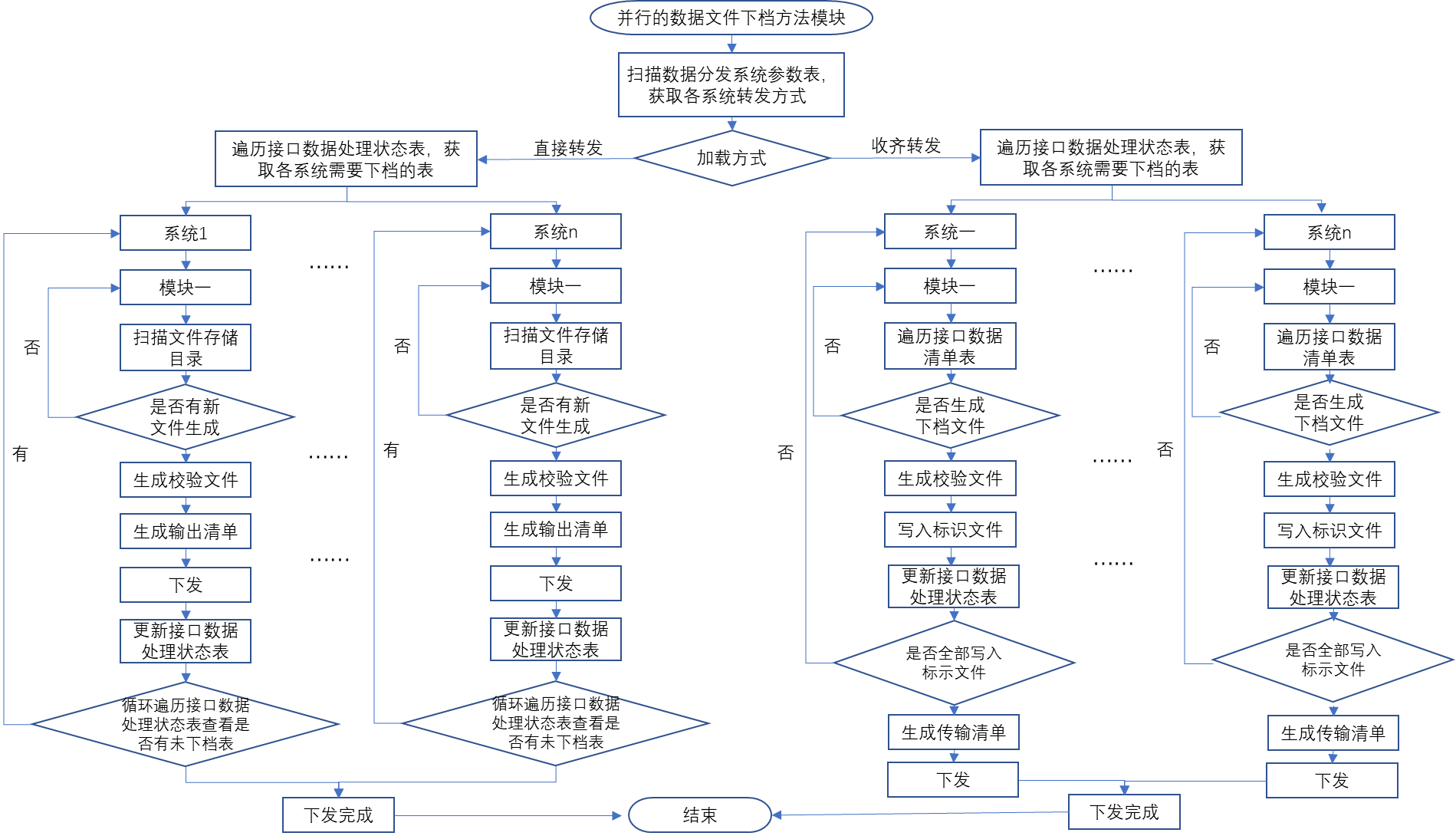


图 3-4 订阅分发架构

下表为各种分发方式的对比结果：

表 3-4 分发方式对比

|  |  |  |  |
| --- | --- | --- | --- |
| 分发方式 | 选用架构 | 优点 | 缺点 |
| 数据中心 | 中心分发架构 | 有效降低了数据交互的复杂度，通过统一的安全管理策略来避免因各系统间松散且无标准的点对点交互所产生的数据安全隐患问题 | 同步方式单一、源头管理困难、对外数据安全风险高 |
| 被动分发 | 订阅-分发架构 | 它能提供多种方式，跨数据库类型、跨数据库版本、跨操作系统平台的订阅分发功能 | 直连数据数据安全风险高，适用于数据需求时效高、比较紧急的应用 |
| 主动分发 | 主动分发架构 | 支持超大数据量、多并发、高性能的数据分发 | 数据溯源、追责难 |

3.1.4 数据风险

（1） 中心化风险

1） 数据汇聚到数据中心之后，基础部门的数据权属变得不清晰，同时也不利于出现问题时责任的正确匹配。

2） 中心化结构特有的数据安全、信息泄露等问题。

3） 数据具有可复制和易传播的特性，会导致数据脱离数据中心的控制后，存在数据泄露、越权使用的风险，还可能有数据被篡改、破坏的可能，也无法保证隐私数据安全性。

（2） 数据价值及相关风险

数据本身由于其价值特性也会有相应的风险。本小节参考标准对数据本身价值特性进行了划分，并列出了数据特性不同程度的定义及发生风险时可能后果。

1） 完整性（integrity）：保证信息及信息系统不会被非授权更改或破坏的特性。包括数据完整性和系统完整性。

表 3-5 完整性

|  |  |  |
| --- | --- | --- |
| 赋值 | 标识 | 定义 |
| 5 | 很高 | 完整性价值非常关键，未经授权的修改或破坏会对组织造成重大的或无法接受的影响，对业务冲击重大并可能造成严重的业务中断，难以弥补 |
| 4 | 高 | 完整性价值较高，未经授权的修改或破坏会对组织造成重大影响，对业务冲击严重，较难弥补 |
| 3 | 中等 | 完整性价值中等，未经授权的修改或破坏会对组织造成影响，对业务冲击明显，但可以弥补 |
| 2 | 低 | 完整性价值较低，未经授权的修改或破坏会对组织造成轻微影响，对业务冲击轻微，容易弥补 |
| 1 | 很低 | 完整性价值非常低，未经授权的修改或破坏对组织造成的影响可以忽略，对业务冲击可以忽略 |

2） 脆弱性（vulnerability）：可能被威胁所利用的数据的薄弱环节。

表 3-6 脆弱性

|  |  |  |
| --- | --- | --- |
| 赋值 | 标识 | 定义 |
| 5 | 很高 | 如果被威胁利用，将对资产造成完全损害 |
| 4 | 高 | 如果被威胁利用，将对资产造成重大损害 |
| 3 | 中等 | 如果被威胁利用，将对资产造成一般损害 |
| 2 | 低 | 如果被威胁利用，将对资产造成较小损害 |
| 1 | 很低 | 如果被威胁利用，造成损害可以忽略 |

3） 可用性（availablity）：数据或资源的特性，被授权实体按要求能访问和使用数据或资源。

表 3-7 可用性

|  |  |  |
| --- | --- | --- |
| 赋值 | 标识 | 定义 |
| 5 | 很高 | 可用性价值非常高，合法使用者对信息及信息系统的可用度达到年度99.9%以上，或系统不允许中断 |
| 4 | 高 | 可用性价值较高，合法使用者对信息及信息系统的可用度达到每天90%以上，或系统允许中断时间小于10min |
| 3 | 中等 | 可用性价值中等，合法使用者对信息及信息系统的可用度在正常工作时间达到70%以上，或系统允许中断时间小于30min |
| 2 | 低 | 可用性价值较低，合法使用者对信息及信息系统的可用度在正常工作时间达到25%以上，或系统允许中断时间小于60min |
| 1 | 很低 | 可用性价值可以忽略，合法使用者对信息及信息系统的可用度在正常工作时间低于25% |

4） 保密性（Security）

表 3-8 保密性

|  |  |  |
| --- | --- | --- |
| 赋值 | 标识 | 定义 |
| 5 | 很高 | 包含组织最重要的秘密，关系未来发展的前途命运，对组织根本利益有着决定性的影响，如果泄露会造成灾难性的损害 |
| 4 | 高 | 包含组织的重要秘密，其泄露会使组织的安全和利益遭受严重损害 |
| 3 | 中等 | 组织的一般性秘密，其泄露会使组织的安全和利益受到损害 |
| 2 | 低 | 仅能在组织内部或在组织某一部门内部公开的信息，向外扩散有可能对组织的利益造成轻微损害 |
| 1 | 很低 | 可对社会公开的信息，公用的信息处理设备和系统资源等 |

(3) 威胁来源

保存在信息媒介上的各种数据资料，包括源代码、数据库数据、系统文档、运行管理规程、计划、报告、用户手册、各类纸质的文档等。

威胁可以通过威胁主体、资源、动机、途径等多种属性来描述。造成威胁的因素可分为人为因素和环境因素。根据威胁的动机，人为因素又可分为恶意和非恶意两种。环境因素包括自然界不可抗的因素和其他物理因素。威胁作用形式可以是对信息系统直接或间接的攻击，在保密性、完整性和可用性等方面造成损害；也可能是偶发的或蓄意的事件。

1） 环境因素

断电、静电、灰尘、潮湿、温度、鼠蚁虫害、电磁干扰、洪灾、火灾、地震、意外事故等环境危害或自然灾害，以及软件、硬件、数据、通讯线路等方面的故障。

2） 人为因素

恶意人员：不满的或有预谋的内部人员对信息系统进行恶意破坏;采用自主或内外勾结的方式盗窃机密信息或进行篡改；外部人员利用信息系统的脆弱性，对网络或系统的保密性、完整性和可用性进行破坏，以获取利益或炫耀能力。

非恶意人员内部人员：由于缺乏责任心，或者由于不关心或不专注，或者没有遵循规章制度和操作流程而导致故障或信息损坏;内部人员由于缺乏培训、专业技能不足、不具备岗位技能要求而导致信息系统故障或被攻击。

(4) 威胁子类

软硬件故障：对业务实施或系统运行产生影响的设备硬件故障、通讯链路中断、系统本身或软件缺陷等问题。如：设备硬件故障、传输设备故障、存储媒体故障、系统软件故障、应用软件故障、数据库软件故障、开发环境故障等。

恶意代码：故意在计算机系统上执行恶意任务的程序代码。如：病毒、特洛伊木马、蠕虫、陷门、间谍软件、窃听软件等。

越权或滥用：通过采用一些措施，超越自己的权限访问了本来无权访问的资源，或者滥用自已的权限，做出破坏信息系统的行为。如：非授权访问网络资源、非授权访问系统资源、滥用权限非正常修改系统配置或数据、滥用权限泄露秘密信息等。

网络攻击：利用工具和技术通过网络对信息系统进行攻击和人侵。如：网络探测和信息采集、漏洞探测、嗅探(帐号、口令、权限等)、用户身份伪造和欺骗、用户或业务数据的窃取和破坏、系统运行的控制和破坏等。

泄密信息泄露：给不应了解的他人信息导致。如内部信息泄露、外部信息泄露等。

篡改非法修改信息：破坏信息的完整性使系统的安全性降低或信息不可用。如：篡改网络配置信息、篡改系统配置信息、篡改安全配置信息、篡改用户身份信息或业务数据信息等。

抵赖：不承认收到的信息和所作的操作和交易。如：原发抵赖、接收抵赖、第三方抵赖等。

3.2 电力数据特点与风险分析

3.2.1 电力数据特点

电力数据覆盖面广，实时性好，准确性高，对其他领域具有极大参考价值：电力数据对工业生产部门，尤其是大型工厂等单位的意义主要是指导这类机构优化用电策略，开展绿色生产，降低企业运营成本等。电力数据对售电方最直接的作用就是通过分析数据提升企业效益。用电数据可以帮助政府对工业生产状况、城区住房空置率、电价补贴政策影响等涉及经济宏观走势和人民日常生活的方方面面进行分析。

3.2.2 电力数据风险与数据安全法

(1) 数据稳定性问题。电力服务是国民经济的命脉，电力数据一旦遭到泄露和篡改将导致极其严重的后果。而以数据中心为支撑的中心化数据管理方案必须依赖一个保存所有数据的中心服务器，因而稳定性受到单台计算机稳定性的制约。

(2) 数据隐私问题。以数据中心为核心的共享方案的另一隐患是隐私问题。由于电力数据中存在大量不宜公开的隐私信息，如居民的地址、电话号码等，因此隐私性问题值得关注。

随着信息技术和人类生产生活交汇融合，各类数据迅猛增长、海量聚集，对经济发展、人民生活产生了重大而深刻的影响。数据安全已成为事关国家安全与经济社会发展的重大问题。党中央高度重视，就加强数据安全工作和促进数字化发展作出一系列重要部署。按照党中央决策部署，贯彻总体国家安全观的要求，全国人大常委会积极推动数据安全立法工作。经过三次审议，2021年6月10日，《数据安全法》经十三届全国人大常委会第二十九次会议通过并正式发布，将于2021年9月1日起施行。

《数据安全法》形成了数据安全的顶层设计。作为我国数据安全领域的基础性法律，《数据安全法》主要有以下三个特点：一是坚持安全与发展并重。设专章对支持促进数据安全与发展的措施作了规定，保护个人、组织与数据有关的权益，提升数据安全治理和数据开发利用水平，促进以数据为关键生产要素的数字经济发展。二是加强具体制度与整体治理框架的衔接。从基础定义、数据安全管理、数据分类分级、重要数据出境等方面，进一步加强与《网络安全法》等法律的衔接，完善我国数据治理法律制度建设。三是回应社会关切。加大数据违法行为处罚力度，建设重要数据管理、行业自律管理、数据交易管理等制度，回应实践问题及社会关切。《数据安全法》完善了国家数据安全工作体制机制，规定中央国家安全领导机构负责国家数据安全工作的决策和议事协调等职责，并提出建立国家数据安全工作协调机制。在网络数据安全工作方面，专门明确国家网信部门依照本法和有关法律、行政法规的规定，负责统筹协调网络数据安全和相关监管工作。

《数据安全法》重点确立了数据安全保护的各项基本制度，完善了数据分类分级、重要数据保护、跨境数据流动和数据交易管理等多项重要制度，形成了我国数据安全的顶层设计。

3.2.3 电力系统数据使用管理办法

（1）工作目标

科学界定数据开放需求，制订差异化开放策略，规范数据开放流程，构建协同高效，安全可控，依法合规的数据开放体系。

（2）工作原则

1)共享融合，互利共赢：树立数据开放共享理念，积极响应数据对外开放需求，加强与外部机构的沟通互动，探索共享融合、互利共赢的合作模式。通过内外数据的开放共享和融合应用，提升公司数据对外服务能力，发挥电力数据价值，真正将资源优势转化为竞争优势。

2)统筹规划，分类施策：加强数据对外开放的顶层设计，统筹推进公司数据开放工作，建立规范的数据开放标准。结合不同需求主体和需求内容，制订差异化的数据对外开放策略，充分发挥各部门的专业优势，建立分级统筹、专业把关的工作机制。

3)依法合规，确保安全：严格遵守国家法律、法规和公司规定，严格保护国家秘密、公司商业秘密，个人信息安全、严格执行国家跨境数据共享有关要求，在不损害公司、客户等相关方利益的前提下，推动数据有序开放。加强前沿技术创新应用，提升对共享数据的感知、监测、追溯和控制能力，保障数据安全。

4)健全机制，规范有序：健全数据开放制度和流程，持续规范和完善数据对外开放管理，推进对外提供数据工作的制度化和规范化。总结经验、迭代完善，逐步健全职责清晰，协同配合的数据对外开放体系。

（3）职责分工

表 3-9 职责分工

|  |  |
| --- | --- |
| 国网互联网部 | 对外开放工作归口管理，建立数据对外开放制度、流程和机制；制订数据脱敏策略、开展数据对外开放需求，落实技术支撑，并对数据开放安全进行指导，检查，评价 |
| 国网保密办 | 数据对外开放的保密工作归口管理和保密工作的指导和检查 |
| 国网法律部 | 负责数据对外开放涉及的决策、方案、合同等合规审核 |
| 总部各部门 | 负责本专业数据对外开放管理，受理本专业数据对外开放需求，管控本专业数据质量，制订本专业数据脱敏策略 |
| 国网大数据中心 | 负责公司数据对外开放的专业支撑，在总部互联网部和业务部门的委托和指导下，做好数据脱敏、数据处理、数据应用开发等技术支撑工作 |
| 各分部，公司各单位 | 负责本单位数据对外开放工作，参照总部职责分工，细化本单位落地措施。 |

（4）数据开放策略及流程

1）数据开放需求分类

表 3-10 开放需求

|  |  |
| --- | --- |
| 省公司层面对内需求 | 省公司内部各业务部门、业务支撑及实施机构、地市公司跨层级等的数据共享应用需求。 |
| 地市公司层面对内需求 | 地市公司内部本单位各部门、业务支撑及实施机构的数据共享应用需求。 |
| 政府监管类需求 | 面向国家监管机构报送发用电情况，供电质量等监管信息，依据法律法规提供相关数据 |
| 公益服务类需求 | 面向政府机构和非盈利性组织，提供公益性数据服务 |
| 商务增值类需求 | 面向公司外部各类机构，以新业务拓展和商务增值为目标，结合对方需要打造精准营销、企业征信等数据产品，对外提供数据产品服务 |
| 公共开放类需求 | 面向公司外部各类机构、电力用户和社会公众等，按照国家和政府要求开放收费标准、用电容量、电费情况等数据 |

2） 数据开放策略

表 3-11 开放策略

|  |  |
| --- | --- |
| 省公司层面对内开放策略 | 负面清单之外且明细数据不出中台的数据使用需求，基于中台自助按需获取和使用；列入负面清单数据、明细数据出库等的数据需求，经审批通过方可使用。 |
| 地市公司层面对内开放策略 | 涉及负面清单数据提交至业务部门（数据提供方）进行审批，涉及明细数据出库的需求提交互联网部审批。涉及跨单位数据使用的需求由地市科技互联网部提交至国网江苏省信通公司。 |
| 政府监管类开放策略 | 由相关业务部门承接，以政府部门和监管机构的公文或公函内容为依据，涉及提供保密数据的，履行流程并签署保密协议，保证数据使用安全合规。 |
| 公益服务类开放策略 | 由开展数据应用的业务部门承接，原则上不提供明细数据，以提供数据服务或打造数据产品方式响应需求，涉及保密数据的，需签订保密协议 |
| 商务增值类开放策略 | 由开展数据应用的业务部门承接，原则上不提供明细数据，以提供数据服务或打造数据产品方式响应需求，签订数据服务合同，明确数据产品服务价格、使用要求、权限范围 |
| 公共服务类开放策略 | 由相关业务部门负责，以信息发布和查询服务方式主动响应数据开放需求 |

3） 数据开放流程

表 3-12 开放流程

|  |  |
| --- | --- |
| 需求受理 | 由需求承接方对接需求，明确数据开放目的、需求内容、开放方式、申请理由等，判别需求类型，填写数据对外开放申请表 |
| 需求分析和初审 | 数据产生部门对数据需求进行初审，分析其数据开放目的、开放内容、开放方式等，提出审核意见 |
| 保密合规审核 | 数据产生部门负责本专业数据保密合规检查，保密办负责公司商业秘密数据的保密审核，法律部负责合规性审核 |
| 数据归口审核 | 互联网部负责数据归口管理审核，分析数据提供方式、数据内容、数据安全、技术支撑等 |
| 签订合同（协议） | 根据提供数据涉密情况，由需求承接方根据审核意见，与需求提出方签订数据保密协议和服务合同，明确需求提出方应遵循的保密要求 |
| 提供服务 | 与需求提出方对接，确定数据提交方式和保障措施等，按要求依法合规向需求提出方提供数据服务 |

4）数据共享策略

表 3-13 共享策略

|  |  |
| --- | --- |
| 做好数据识别和分类分级 | 通过数据盘点理清数据状况，建立数据目录，以数据目录为基础梳理形成数据共享负面清单，结合负面清单明细进行数据安全升级，识别数据保护对象。同时，应结合数据在挖掘、计算、分析、处理后安全属性动态变化的特点识别、调整数据分级和保护要求。 |
| 公司内部共享数据 | 数据共享需求清单在公司范围内统一发布、定期更新。按照数据共享负面清单开展，原则上非列入负面清单的数据可以在公司内部直接共享，商业秘密、工作秘密、个人信息等纳入数据负面清单的数据，由数据使用方提出申请，经数据产生的总部业务部门和数据归口管理部门审批后方可对内跨部门、跨单位给数据使用方使用。对涉及负面清单中个人信息等敏感数据的共享应用，应采取数据脱敏、数据水印、访问授权、安全审计等措施，保证数据应用安全。数据使用方对负面清单内的数据使用完毕后，应及时销毁相关数据。 |
| 公司内部中台 | 基于业务部门数据共享需求，大数据中心形成数据共享需求清单。原则上明细数据不出中台，应基于全业务统一数据中心和数据中台提供的内部受控环境，开展数据的在线查询和在线应用， 不得自行将数据拷贝和提供他人。依托全业务统一数据中心和数据中台，将分散在不同业务信息系统的数据进行整合汇聚，形成统一的数据资源中心。按照“数据一个源”的要求，实现公司数据的一次采集、多处使用。大数据中心根据公司要求，推进外部数据统一纳管。 |
| 公司对外提供数据 | 需要执行差异化的数据开放策略，涉密数据按照公司保密规章制度执行，严格履行相关保密合规审核流程。应通过与外部合作单位和供应商签订合同（含保密条款）、保密协议、保密承诺书等方式进行数据安全管控，合同、协议、承诺书由业务部门会同法律合规部门根据实际情况制订，内容需明确数据使用范围、途径等。数据对外提供应遵守国家有关政策法规和公司管理规定要求，除国家机关依法调取数据外，原则上公司明细业务数据不对外提供。禁止外部合作单位和供应商在对互联网提供服务的网络和信息系统中存储或运行公司商业秘密数据和重要数据。 |
| 开展数据产品研发，发布活动 | 应确保遵从国家法律法规、数据主体授权范围、公司相关规章制度等数据安全保护要求。对外发布特定数据产品和服务，需按照国家、行业相关法律法规要求，获得相应的备案、资质或牌照，并报业务部门和数据管理工作归口部门备案。数据产品和服务提供方与使用方应签署相关使用协议，明确约束使用规则等双方权利义务。 |
| 通过互联网电子渠道交互数据（详细数据对不对外开放） | 对利用互联网电子渠道交互或发布用户的业务数据，应采用符合公司安全防护方案的业务数据交互方式，并经安全专家委员会审查和安全检测机构测评。防止非授权发布或交互企业重要数据以及个人信息，扰乱互联网秩序等违法违规活动。 |
| 境内数据与跨境数据 | 我国境内产生和收集的个人信息和重要数据应在境内存储，由境外产生并跨境传输至境内的数据，应禁止非法、负面信息由境外流入境内；因业务需要，确需向境外提供的数据，应经公司保密办与业务部门审批，根据国家相关规定做好安全评估等工作，视情况向国家有关部门报备。 |

3.3 电力数据分发场景与数据应用特征解析

3.3.1 典型分发场景

根据国家电网有限公司文件，电力系统数据分发可以归纳为两大类型，即内网分发和外网分发：

（1）内网分发：在公司、机构、组织等构建的可信任的局域网内进行数据分发，一般采用数据中心方式，内网各部门直连数据中心获取数据。

内网分发流程图：

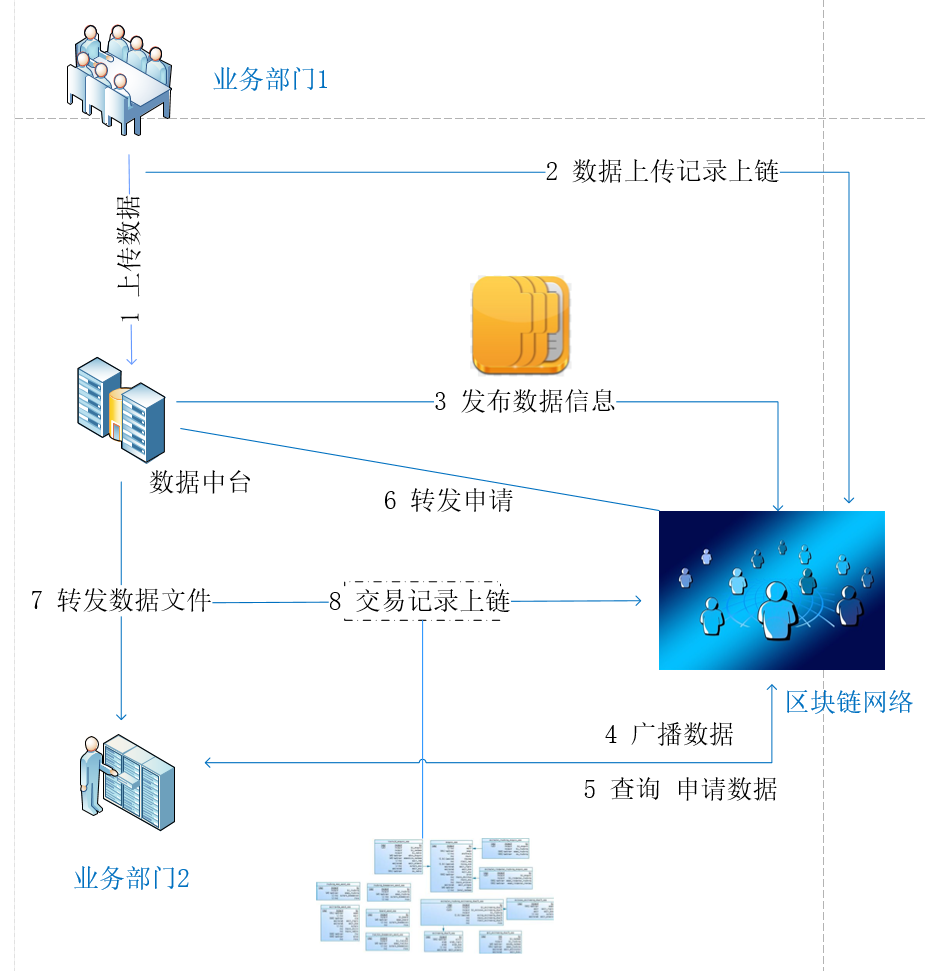


图 3-5 内网分发流程

1. 1~2.省公司业务部门采用加密技术将数据打包上传到数据中台，并将此次数据传输记录在区块链上。
2. 3~4.数据中台将内部数据加密存储在本地或是云服务器中，并向区块链网络中广播数据摘要信息，对内部单位提供要选择的数据信息。
3. 5~6.  需求单位向公司发起数据申请，对于涉及负面清单数据的需求由业务部门（提供方）、互联网部审批通过后方可共享，明细数据出库需求由互联网部审批通过后方可提供。地市公司的跨单位需求由总公司相应部门审批。
4. 7~8.  数据中台利用哈希和加密算法将数据传输给提出申请的单位，并将此次交易记录的具体信息上链。

 一旦发生数据的非法转发，可以根据区块链的不可篡改性找到传输数据的发送方，对于不经过区块链进行的数据传输，可以利用水印技术进行一个责任方的问责工作，确保数据对内开放的溯源追踪。

（2）外网分发：外网分发是指面向政府、企事业单位、用电客户、社会公众等提供电力数据或产品服务的行为。主要形式包括：直接提供业务明细数据；提供基于原始数据加工形成的数据产品和数据服务。

外网分发流程图：

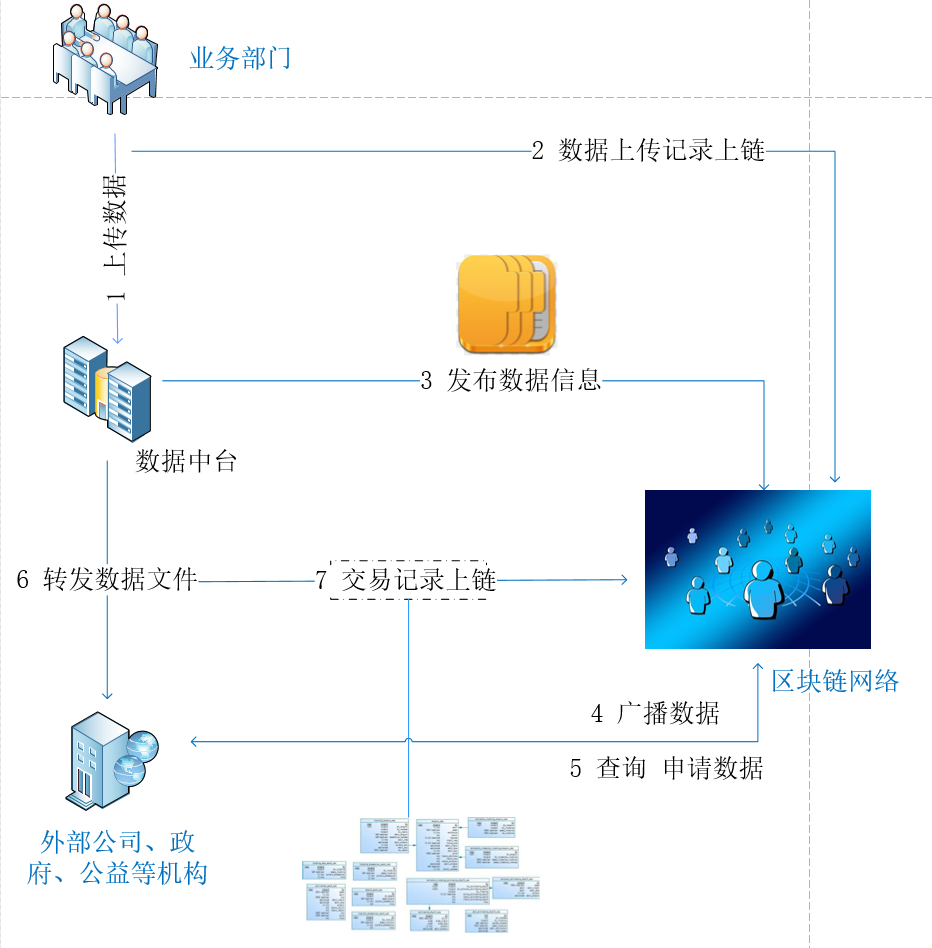


图3-6 外网分发流程

1. 公司业务部门采用加密技术将本部门产生的业务数据打包上传到数据中台，并将此次数据传输记录（包括发送方业务部门名称，数据类型，数据摘要信息，发送时间，接收方数据中台等）在区块链上，用于追溯业务部门是否进行了非法转发
2. 数据中台里有国网公司全部的数据信息，将其加密存储在本地或是云服务器中，并向区块链网络中广播数据摘要信息，用于为外部公司、政府等机构提供要选择的数据信息。
3. 外部单位向国网公司发起数据申请，根据要申请的数据信息种类，分析是否要经保密部门和法律部进行合规性审核，如果审核不通过，则对外数据流程终止。审核通过，组织技术部门通过数据中台向外部单位提供数据。
4. 数据中台利用哈希和加密算法将数据明细或数据产品服务经不安全信道传输给提出申请的外部机构，并将此次交易记录（发送方数据中台，数据类型，数据摘要信息，发送时间，接收方单位等）的具体信息上链。
5. 一旦发生数据的非法转发，可以根据区块链的不可篡改性找到传输数据的发送方（业务部门或者数据中台），对于不经过区块链进行的数据传输，可以利用水印技术进行一个责任方的问责工作，确保数据对外开放的溯源追踪。

3.3.2 数据应用类型

(1) 发电调控

通过电力数据帮助发电企业日常运营，作为发电机组的监测指标，对设备故障风险评估，实现精益生产；支撑发电企业精益化检修和发电调控。既提升发电企业经营效益，又促进源网协调。

(2) 经营异常监测

以电力数据为基础，融合金融市场数据和第三方数据，对企业实际经营情况进行挖掘分析，识别和监测有经营异常波动和虚假信息披露的公司，辅助金融监管部门加强市场风险识别和监管，提高监管效率和水平。

(3) 行业动能发展指数

用电消费、用电需求是行业投资中的一项关键数据，行业动能发展指数可对行业发展进行客观展现。反映到季度、月度、周甚至日的行业用电消费变化状况，可充分发挥电力数据价值。

(4) 用电信息征信体系服务

基于电力客户基本信息、长期的用电记录、缴费情况、缴费能力等数据，对各类数据进行统计分析，建立用户信用评级指标和标准，进行用户信用评价，并分析客户信用变化趋势和潜在风险。同时，利用相似的方法，基于电力客户基本信息、用电情况、利润贡献、设备装备水平等数据，建立用户价值评级指标和评分标准，综合考虑企业信用等级及企业经营情况，实现对客户价值等级的评估。

(5) 用户能耗分析及用电优化

基于用电信息、用户负荷等数据，研究用户的负荷特性及用电行为习惯，研究用户用电行为分析、用电负荷特性分析、用电影响模型、用户能耗分析、用户用电建议等。制订节能方案，提高能源利用率，降低电能损失，保障客户经济利益，促进节能减排。

(6) 产业结构用电分析

按区域进行全行业、高耗能行业、高附加值的用电情况分析挖掘。

(7) 地区用电分析

按行业进行经济区域、行政区域的用电量分析挖掘。

(8) 电力经济预测

按单位增加值变化情况对用电量进行分析与预测。

(9) 居民生活用电特性分析

按城乡进行居民用电、居民户均用电、零用电户情况、阶梯电价情况进行关联关系分析挖掘。

表 3-14 应用类型

|  |  |
| --- | --- |
| **应用类型** | **分发场景** |
| 发电调控 | 内网分发 |
| 经营异常检测 | 外网分发 |
| 行业动能发展指数 | 外网分发 |
| 用电信息征信体系服务 | 内网分发 |
| 用户能耗分析及用电优化 | 内网分发 |
| 产业结构用电分析 | 外网分发 |
| 地区用电分析 | 内网分发 |
| 电力经济预测 | 内网分发 |
| 居民生活用电特性分析 | 内网分发 |

3.4 电力数据外泄形式与安全防控需求

3.4.1 电力数据外泄形式

（1） 一般企业数据外泄形式

一般企业的数据外泄包括使用外泄、存储外泄和传输外泄：

一是使用泄漏。如操作失误、打印、外发文件、拍摄屏幕等方式泄漏内部数据。

二是存储泄漏。包括数据中心、服务器和数据库的数据被入侵造成泄漏；离职人员通过移动存储设备随意拷走机密资料，或者离职人员对在职时期的涉密邮件导出；移动终端被盗、丢失或维修造成数据泄漏；黑客攻击等。

三是传输泄漏。通过网络监听、拦截等方式对传输数据进行篡改、伪造和窃取。

（2） 电力数据威胁

数据传输风险：在数据传输过程中，数据被窃取或篡改，破坏了数据的机密性和完整性。

数据非法转发风险：数据以不合规的方式被转发，使得非授权单位获得数据。

数据泄露风险：数据被恶意人员以非转发方式窃取，如打印、拍照等。

数据存储风险：数据存储在本地或云服务器中出现的数据泄露情况。

3.4.2 电力系统安全防控需求

针对使用泄露：提升工作人员素质，减少操作失误，以免泄露内部数据。

针对存储泄露：数据实行加密存储，中心管理平台对防护产品进行统一管理和配置，定义部署安全策略，将策略下发到监测、阻断、发现设备，接收敏感数据信息事件，对事件进行审计、分类、告警、追踪取证。以集中策略为基础对企业敏感数据进行事前、事中、事后全方位管理。

针对传输泄露，对传输数据进行严格加密处理，防止数据明文在传输过程中被窃取。

对于国网数据安全要求：

加强数据全生命周期安全防护，数据采集环节应遵循合法、正当、必要、最小化、准确性、可问责性的原则，公开收集、使用等处理信息规则，明示收集、使用信息的目的、方式和范围；数据传输环节要采用公司统一密码基础设施签发的密钥或者证书加密传输；数据存储环节在存储重要数据时应加密存储，落实重要数据备份、访问控制、安全审计等技术应用；数据使用环节要结合数据业务场景采用脱敏、水印、审计等手段实现差异化防护，遵循最小授权原则访问和处理个人信息和企业重要数据，加强数据安全监测、预警、审计和处置能力建设，提高对各类数据安全事件发现、响应和溯源能力；数据销毁环节要根据公司要求合理选择数据恢复、擦除与销毁措施进行处理。

（1） 数据开放流程：数据开放流程主要包括需求受理、需求分析和初审、保密合规审核、数据归口审核、签订合同（协议）、提供服务六个主要环节。

需求受理：由需求承接方对接需求，明确数据开放目的、需求内容、开放方式、申请理由等，判别需求类型，填写数据对外开放申请表。涉及提供数据服务或数据产品的，需求承接方应对服务和产品进行保密合规审查，并提报审核结果。

需求分析和初审：数据产生部门对数据需求进行初审，分析其数据开放目的、开放内容、开放方式等，提出审核意见。判断相关数据是否泄密，存在泄密数据的，同步履行涉密审查审核手续。

保密合规审核：数据产生部门负责本专业数据保密合规审查，保密办负责公司商业秘密数据的保密审核；法律部负责合规性审核。审查审核存在问题时，相关部门应给出建设性意见建议，共同研究提出解决方案；保密合规审核不通过的，终止数据开放流程。

数据归口审核：互联网部负责数据归口管理审核，从数据归口管理角度，分析数据提供方式、数据内容、数据安全、技术支撑等，提出审核意见。

签订合同（协议）：根据提供数据涉密情况，由需求承接方根据审核意见，与需求提出方签订数据保密协议，明确需求提出方应遵循的保密要求。商务增值类需求还应按要求签署数据服务合同，合同须规定双方的权利和义务、数据使用范围、数据提供方式、授权许可类型和数据交付期限、协议期限、费用及支付方式、保密要求、免责条款等。合同和协议按要求开展依法合规审查。

提供服务：与需求提出方对接，确定数据提交方式和保障措施等，按要求依法合规向需求提出方提供数据。发现需求提出方存在违反国家法律、法规，或者与公司签署的保密协议、合同相关情况时，应立即要求其停止相关行为，同时终止数据开放，并按规定追究相关责任。合同（协议）到期未续签的，应终止数据开放，并督促需求提出方对接收数据进行擦除和销毁等工作。

（2） 数据对内开放

1） 负面清单

定义：“负面清单”是指不能直接共享使用，需经数据提供部门对数据用途、提供方式、脱敏要求等进行审核后方可共享的数据清单；

策略：基于负面清单的数据共享流程适用于公司内部数据共享，涉及数据对外提供、对外增值服务等数据共享需求，按相关规定和要求执行。列入负面清单的数据，经相关业务部门审批后共享，其余数据均可在公司内部共享使用。依托数据中台，实现数据“在线审批、在线获取、在线应用”，确保数据在线管理、规范使用。

职责：数据共享遵循数据分级授权策略，互联网部统一归口管理，明确职责、流程，各部门、各单位按照权限范围开展数据共享工作；地市公司使用本单位数据的需求，不再提交省公司审批，由地市公司自行审批后共享。

安全需求：遵循“谁主管谁负责，谁运行谁负责，谁使用谁负责”的原则，严格执行公司保密和网络安全相关规定， 落实数据安全责任；原则上明细数据不出中台，敏感数据须脱敏后方可共享。

2) 省公司层面

定义：省公司内部各业务部门、业务支撑及实施机构、地市公司跨层级等的数据共享应用需求。

策略：负面清单之外且明细数据不出中台的数据使用需求，基于中台自助按需获取和使用；列入负面清单数据、明细数据出库等的数据需求，经审批通过方可使用。涉及跨专业数据计算汇总的报表、指标等分析应用需求，原则上基于全业务统一数据中心和数据中台实现。数据增值变现等业务涉及的数据对外使用，应严格按照有关数据安全、隐私保护和公司保密要求，履行对外提供数据的审批，且在采取重要数据保护措施下对 外提供，确保公司数据对外使用安全。

流程：省公司各业务部门、业务支撑及实施机构等数据使用授权流程涉及需求受理、需求分析、需求审核及授权使用4个阶段9个流程环节。各业务部门等提出需求后由国网江苏省信通公司统一受理。涉及负面清单数据提交至业务部门（数据提供方）进行审批，涉及明细数据出库的需求提交互联网部审批。国网江苏省信通公司在中台完成授权配置。

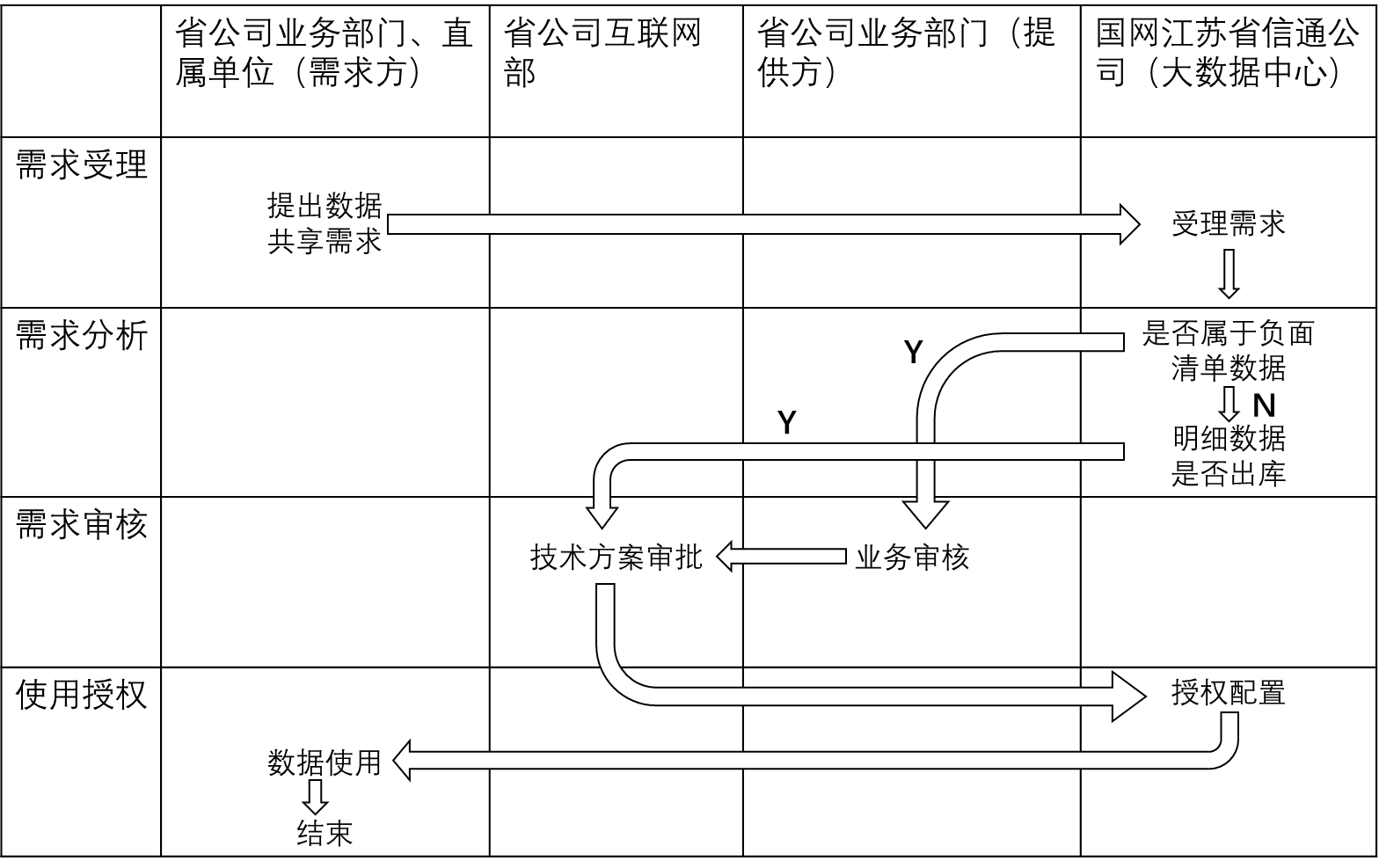


图 3-7 省公司流程

职责：省公司业务部门、业务支撑及实施机构（需求方）根据自身业务需求，提出数据使用申请，由国网江苏省信通公司统一受理。国网江苏省信通公司受理需求后，组织开展需求分析，判断需求数据是否为负面清单数据、明细数据是否出库等，明确数据提供方案。涉及负面清单数据的需求由业务部门（提供方）、互联网部审批通过后方可共享；明细数据出库需求由互联网部审批通过后方可提供。基于业务部门数据共享需求，国网江苏省信通公司形成数据共享需求清单。依托全业务统一数据中心和数据中台，将分散在不同业务信息系统的数据进行整合汇聚，形成统一的数据资源中心。按照“数据一个源”的要求，列入负面清单数据、明细数据出库等的数据需求，经审批通过，由国网江苏省信通公司在中台完成授权配置后共享使用。

安全需求：建立公司数据共享负面清单制度。业务部门提出本专业数据共享负面清单，对列入负面清单的数据应详细说明原因。对纳入数据共享负面清单的数据，互联网部组织制定数据使用审批程序，数据使用方提出申请，经数据责任部门审批授权后方可使用。应基于全业务统一数据中心和数据中台提供的内部受控环境，开展数据的在线查询和在线应用，不得自行将数据拷贝和提供他人。

表 3-15 风险描述

|  |  |  |
| --- | --- | --- |
| 风险类型 | 风险描述 | 解决方案 |
| 数据传输风险 | 公司业务部门在向数据中台提供或接收数据遭到窃取 | 选用安全的加密算法对明文信息加密，保障其机密性 |
| 公司业务部门在向数据中台提供或接收数据遭到篡改 | 使用哈希算法对消息进行验证，保证其完整性。 |
| 数据非法转发风险 | 业务部门以不合规方式转发数据到其他部门 | 根据转发情况，将转发过程中发送方，接收方，数据文件信息记录到区块链上。当发现数据非法转发时，由区块链信息来保证其可溯源性 |
| 中台转发数据到非授权部门 | 记录中台数据转发流程，将流程信息、文件信息记录到区块链上，通过区块链实现对中台非法转发的溯源 |
| 数据泄露风险 | 数据文件由非法途径如打印、拍照等方式，直接分发到非授权单位或人员中，避开了区块链的分发记录过程 | 对于重要的数据文件，在离开数据中台或业务部门时进行水印处理。当出现无法由区块链进行追责的溯源文件时，根据水印来对文件进行责任判断，保证其可溯源性 |
| 数据存储风险 | 数据存储在本地或云服务器中出现的数据泄露情况 | 对重要数据进行加密存储，保证其机密性。为保证存储安全，加强数据备份、访问控制等技术的应用 |
| 流程文件泄露 | 数据需求清单、审批文件、审核文件等遭到泄露，暴露数据文件信息 | 将需求清单、审批文件、审核文件分发过程中信息记录到区块链上，保证其可溯源性 |

3) 地市公司层面

定义：地市公司内部本单位各部门、业务支撑及实施机构的数据共享应用需求。

策略：负面清单之外且明细数据不出中台的数据使用需求，基于中台自助按需获取和使用；列入负面清单数据、明细数据出库等的数据需求，经审批通过方可使用。涉及跨专业数据计算汇总的报表、指标等分析应用需求，原则上基于全业务统一数据中心和数据中台实现。数据增值变现等业务涉及的数据对外使用，应严格按照有关数据安全、隐私保护和公司保密要求，履行对外提供数据的审批，且在采取重要数据保护措施下对 外提供，确保公司数据对外使用安全。

流程：地市公司数据使用授权流程涉及需求受理、需求分析、需求审核及使用授权 4个阶段15个流程环节。各业务部门等提出需求后由地市公司科技互联网部统一受理。涉及负面清单数据提交至业务部门（数据提供方）进行审批，涉及明细数据出库的需求提交互联网部审批。涉及跨单位数据使用的需求由地市科技互联网部提交至国网江苏省信通公司。

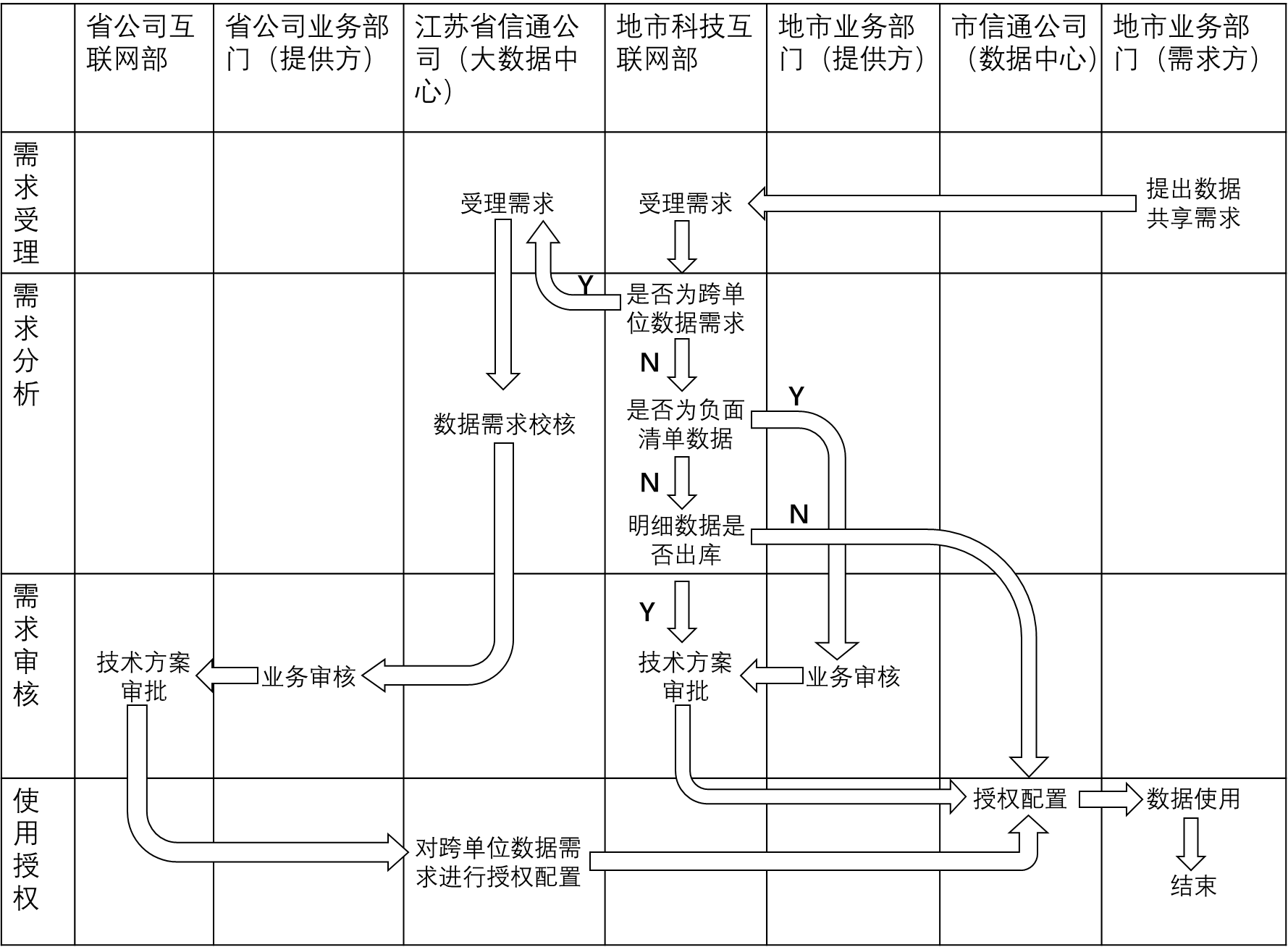


图 3-8 地市公司流程

职责：各业务部门、业务支撑及实施机构（需求方）根据自身业务需求，提出数据使用申请，由地市公司科技互联网部统一受理。科技互联网部受理需求后，组织开展需求分析，判断需求数据是否为负面清单数据、明细数据是否出库等，明确数据提供方案。涉及负面清单数据的需求由本单位业务部门（提供方）审批通过后方可共享；涉及明细数据出库需求由本单位科技互联网部审批通过后方可提供。跨单位数据的使用需求，由总公司业务部门、互联网部审批通过，分别由国网江苏省信通公司、地市信通公司在中台完成授权配置后共享使用。

安全需求：对纳入数据共享负面清单的数据，科技互联网部组织制定数据使用审批程序，数据使用方提出申请，经数据责任部门审批授权后方可使用。应基于全业务统一数据中心和数据中台提供的内部受控环境，开展数据的在线查询和在线应用，不得自行将数据拷贝和提供他人。

表 3-16 风险描述

|  |  |  |
| --- | --- | --- |
| 风险类型 | 风险描述 | 解决方案 |
| 数据传输风险 | 地市业务部在向数据中台提供或接收数据遭到窃取  地市业务部在向省公司中台提供或接收数据使用不安全信道遭到窃取 | 选用安全的加密算法对明文信息加密，保障其机密性 |
| 地市业务部在向数据中台提供或接收数据遭到篡改  地市业务部在向省公司中台提供或接收数据使用不安全信道遭到篡改 | 使用哈希算法对消息进行验证，保证其完整性。 |
| 数据非法转发风险 | 业务部门以不合规方式转发数据到内部其他部门或地市部门 | 严格控制内外网关，只允许中台对外的数据传输。  记录转发过程中流程、文件信息到区块链上。由区块链信息来保证其可溯源性 |
| 中台转发数据到非授权部门或未审批地市公司 | 记录中台数据转发流程，将流程信息、文件信息记录到区块链上，通过区块链实现对中台非法转发的溯源 |
| 数据泄露风险 | 数据文件由非法途径如打印、拍照等方式，直接分发到非授权单位或人员中，避开了区块链的分发记录过程 | 对于重要的数据文件，在离开数据中台或业务部门时进行水印处理。当出现无法由区块链进行追责的溯源文件时，根据水印来对文件进行责任判断，保证其可溯源性 |
| 数据存储风险 | 数据存储在本地或云服务器中出现的数据泄露情况 | 对重要数据进行加密存储，保证其机密性。为保证存储安全，加强数据备份、访问控制等技术的应用 |
| 流程文件泄露 | 数据需求清单、审批文件、审核文件等遭到泄露，暴露数据文件信息 | 将需求清单、审批文件、审核文件分发过程中信息记录到区块链上，保证其可溯源性 |

（3） 数据对外开放需求：

1） 政府监管类

定义：面向国家监管机构，报送的发用电情况、供电质量等监管信息，以及按照国家法律法规、行政法规、政策要求需配合提供的相关数据。该类需求可为政府科学管理和决策提供数据支撑。

对外策略：政府监管类需求由相关业务部门承接，以政府部门和监管机构的公文或公函内容为依据，涉及提供保密数据的，履行相关审核流程并签署保密协议，保证数据使用安全合规。流程主要包括需求受理、需求分析与初审、保密合规审核、签订合同（协议）、提供服务五个环节。对于政府监管类存量的已常态化提供的数据，可按照既有模式继续提供；新增的外部开放需求，由相关业务部门对接监管机构，组织开展需求分析，沟通数据提供方式，涉及涉密数据的，履行相关审核流程，审核通过后签署数据保密协议，按要求提供数据。

对外模式：

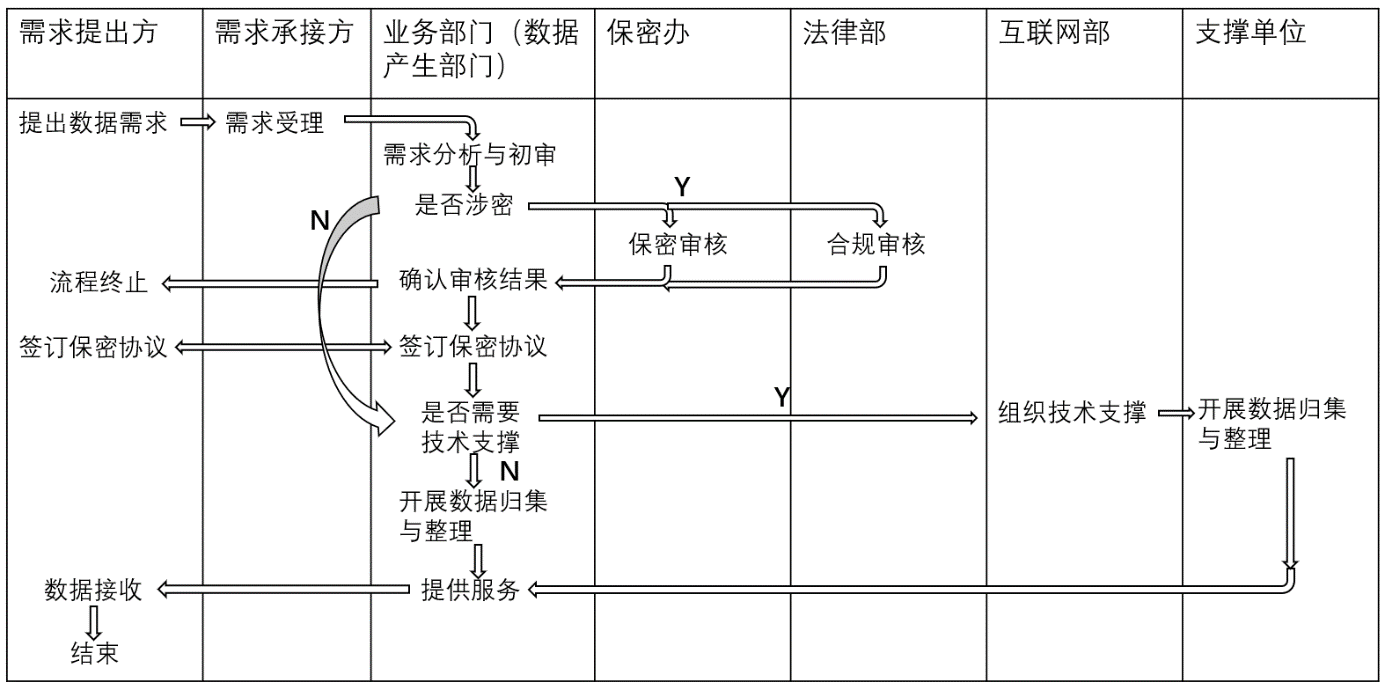


图 3-9 政府监管类

职责划分：由需求承接方对接需求，明确数据开放目的、需求内容、开放方式、申请理由等，判别需求类型，填写数据对外开放申请表。业务部门进行需求分析和初审，分析其数据开放目的、开放内容、开放方式等，提出审核意见。判断相关数据是否泄密，存在泄密数据的，同步履行涉密审查审核手续。如果要对外开放的数据涉密，则由保密办进行保密审核和法律部进行合规性审核，当审查审核不通过时，相关部门应给出建设性意见，共同提出解决方案。保密合规审核不通过的，终止数据对外开放流程。审核通过则由需求提出方与承接方签订合同协议，如果对外开放数据需要技术支撑，则由互联网部组织技术支撑，技术支撑单位开展数据归集与整理工作，否则由业务部门直接开展数据归集与整理为数据需求方提供数据服务。

安全需求：涉密数据按照公司保密规章制度执行，严格履行相关保密合规审核流程。应通过与外部合作单位和供应商签订合同（含保密条款）、保密协议、保密承诺书等方式进行数据安全管控，合同、协议、承诺书由业务部门会同法律合规部门根据实际情况制订，内容需明确数据使用范围、途径等。数据对外提供应遵守国家有关政策法规和公司管理规定要求，除国家机关依法调取数据外，原则上公司明细业务数据不对外提供。禁止外部合作单位和供应商在对互联网提供服务的网络和信息系统中存储或运行公司商业秘密数据和重要数据。在数据传输环节中，采用公司统一密码基础设施签发的密钥或证书，实现企业重要数据的加密传输。在数据使用环节中，应结合数据业务场景采用脱敏、水印、审计等技术手段实现差异化防护，遵循最小授权原则访问和处理个人信息和企业重要数据。加强数据安全监测、预警、审计和处置能力建设，提高对各类数据安全事件发现、响应和溯源能力。

表 3-17 风险描述

|  |  |  |
| --- | --- | --- |
| 风险类型 | 风险描述 | 解决方案 |
| 数据传输风险 | 业务部门产生的数据经由不安全信道上传到数据中台供统一调度或数据中台提供数据经不安全信道传输给政府监管部门，数据被篡改 | 利用加密技术，采用公司统一密码基础设施签发的密钥或者证书加密传输，保障其安全性 |
| 业务部门产生的数据经由不安全信道上传到数据中台供统一调度或数据中台提供数据经不安全信道传输给政府监管部门，数据发生泄露 | 利用哈希算法，采用公司统一密码基础设施签发的密钥或者证书加密传输，保障其真实有效 |
| 数据非法转发风险 | 数据文件存在由业务部门直接转发给非授权人员的情况 | 将业务部门转发数据的过程记录在区块链上，发生问题时查找记录，定位责任方 |
| 数据中台直接将数据转发给非授权人员的情况 | 数据中台向外分发的数据信息同样记录上链，发生问题时查找记录，定位责任方 |
| 数据泄露风险 | 数据文件不经过区块链直接从业务部门非法转发 | 对要分发的数据文件打上水印，根据泄露的数据文件的水印信息确定业务部门或者数据中台是数据泄露方 |
| 数据信息由数据中台链下分发 | 对要分发的数据文件打上水印，根据泄露的数据文件的水印信息确定业务部门或者数据中台是数据泄露方 |
| 数据存储风险 | 业务部门或者数据中台将数据文件存储在本地或者云服务器中 | 存储重要数据时应加密存储，落实重要数据备份、访问控制、安全审计等技术应用 |

2） 公益服务类

定义：面向政府机构或非营利性组织等，从服务中央决策部署落地实施、社会治理现代化等方面，提供的公益性数据服务。该类需求有助于体现公司的社会价值。

对外策略：公益服务类需求由开展数据应用的业务部门承接，原则上不提供业务明细数据，以提供数据服务或打造数据产品方式响应需求，涉及提供保密数据的，需签订数据保密协议，明确数据服务的使用要求、权限及范围等，保证数据使用安全合规。提供数据产品的社会公益类流程包括需求受理、数据归口审核、签订合同（协议）、提供服务四个环节，由需求承接方按要求开展需求对接，沟通数据产品服务方式。履行数据归口审核流程，开展合同和协议的依法合规审查审核，组织开展数据产品设计开发并提供服务。直接提供明细数据的社会公益类流程包括需求受理、需求分析和初审、保密合规审核、数据归口审核、签订合同（协议）、提供服务六个环节。由需求承接方、保密办、法律部进行保密合规审查审核；互联网部进行数据归口审核。审核通过后，需求承接方与需求提出方签订数据使用合同或协议，组织提供数据。

对外模式（提供数据产品和服务）：

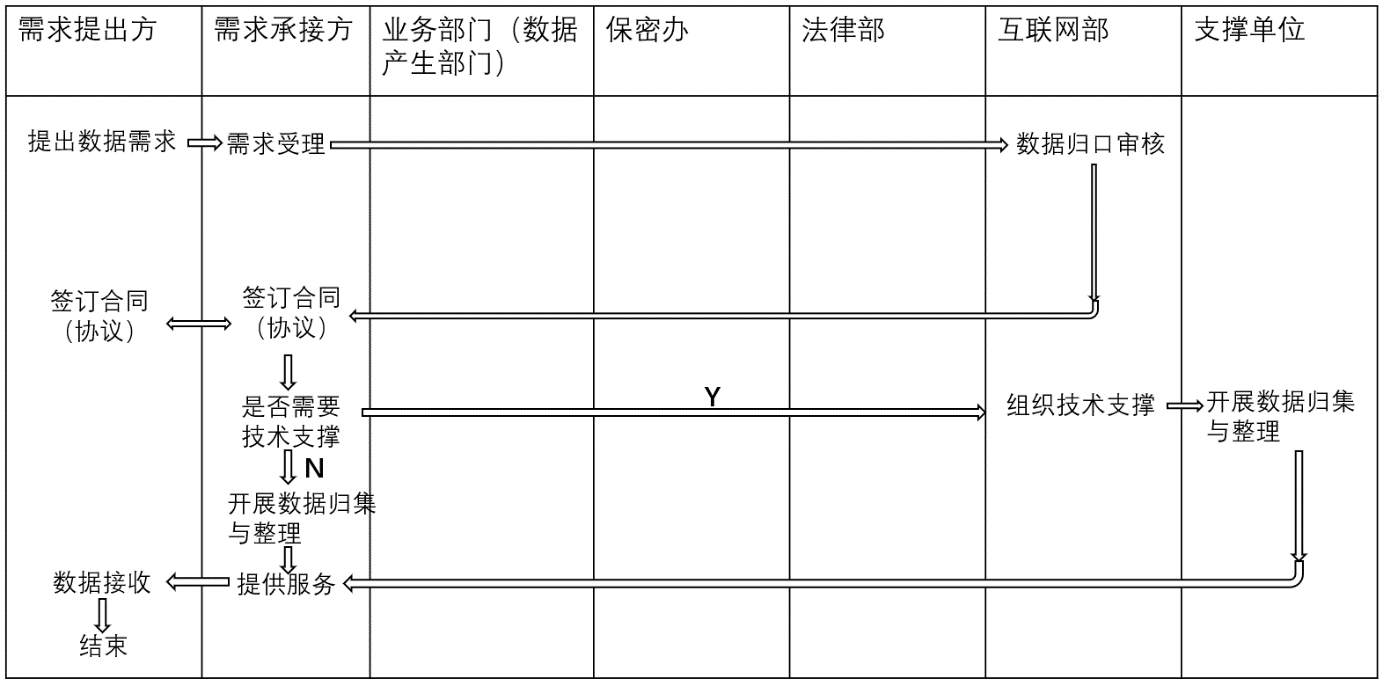


图 3-10 公益服务类

职责划分：由需求承接方对接需求，明确数据开放目的、需求内容、开放方式、申请理由等，判别需求类型，填写数据对外开放申请表。然后交由互联网部开展数据归口审核，从数据归口管理角度，分析数据提供方式、数据内容、数据安全、技术支撑等，提出审核意见。需求提出方与承接方签订合同协议，如果对外开放数据需要技术支撑，则由互联网部组织技术支撑，技术支撑单位开展数据归集与整理工作，否则由业务部门直接开展数据归集与整理为数据需求方提供数据服务。

对外模式（提供明细数据）：



图 3-11 公益服务类对外模式

职责划分：由需求承接方对接需求，明确数据开放目的、需求内容、开放方式、申请理由等，判别需求类型，填写数据对外开放申请表。业务部门进行需求分析和初审，分析其数据开放目的、开放内容、开放方式等，提出审核意见。判断相关数据是否泄密，存在泄密数据的，同步履行涉密审查审核手续。如果要对外开放的数据涉密，则由保密办进行保密审核和法律部进行合规性审核，当审查审核不通过时，相关部门应给出建设性意见，共同提出解决方案。保密合规审核不通过的，终止数据对外开放流程。审核通过互联网部开展数据归口审核，从数据归口管理角度，分析数据提供方式、数据内容、数据安全、技术支撑等，提出审核意见。需求提出方与承接方签订合同协议，如果对外开放数据需要技术支撑，则由互联网部组织技术支撑，技术支撑单位开展数据归集与整理工作，否则由承接方直接开展数据归集与整理为数据需求方提供数据服务。

安全需求：应确保遵从国家法律法规、数据主体授权范围、公司相关规章制度等数据安全保护要求。对外发布特定数据产品和服务，需按照国家、行业相关法律法规要求，获得相应的备案、资质或牌照，并报业务部门和数据管理工作归口部门备案。数据产品和服务提供方与使用方应签署相关使用协议，明确约束使用规则等双方权利义务。在数据传输环节中，采用公司统一密码基础设施签发的密钥或证书，实现企业重要数据的加密传输。在数据使用环节中，应结合数据业务场景采用脱敏、水印、审计等技术手段实现差异化防护，遵循最小授权原则访问和处理个人信息和企业重要数据。加强数据安全监测、预警、审计和处置能力建设，提高对各类数据安全事件发现、响应和溯源能力。

表 3-18 风险描述

|  |  |  |
| --- | --- | --- |
| 风险类型 | 风险描述 | 解决方案 |
| 数据传输风险 | 业务部门产生的数据经由不安全信道上传到数据中台供统一调度或数据中台提供数据经不安全信道传输给政府机构或者非营利性组织，数据被篡改 | 利用加密技术，采用公司统一密码基础设施签发的密钥或者证书加密传输，保障其安全性 |
| 业务部门产生的数据经由不安全信道上传到数据中台供统一调度或数据中台提供数据经不安全信道传输给政府监管部门，数据发生泄露 | 利用哈希算法，采用公司统一密码基础设施签发的密钥或者证书加密传输，保障其真实有效 |
| 数据非法转发风险 | 数据文件存在由业务部门直接转发给非授权人员的情况 | 将业务部门转发数据的过程记录在区块链上，发生问题时查找记录，定位责任方 |
| 数据中台直接将数据转发给非授权人员的情况 | 数据中台向外分发的数据信息同样记录上链，发生问题时查找记录，定位责任方 |
| 数据泄露风险 | 数据文件不经过区块链直接从业务部门非法转发 | 对要分发的数据文件打上水印，根据泄露的数据文件的水印信息确定业务部门或者数据中台是数据泄露方 |
| 数据信息由数据中台链下分发 | 对要分发的数据文件打上水印，根据泄露的数据文件的水印信息确定业务部门或者数据中台是数据泄露方 |
| 数据存储风险 | 业务部门或者数据中台将数据文件存储在本地或者云服务器中 | 存储重要数据时应加密存储，落实重要数据备份、访问控制、安全审计等技术应用 |
| 数据使用风险 | 数据在接收方的不合理使用 | 与需求方签订数据使用合同，规范数据使用策略，结合数据业务场景采用脱敏、水印、审计等手段实现差异化防护 |

3） 商务增值类

定义：面向公司外部各类机构，以新业务拓展和商务增值为目标，结合对方需要打造精准营销、企业征信等数据产品，对外提供数据产品服务。该类需求有利于公司拓展生态圈，实现数据增值变现。

对外策略：商务增值类需求由开展数据应用的业务部门承接，原则上不提供业务明细数据，以提供数据服务或打造数据产品方式响应需求，按照互利共赢的原则，签订数据服务合同，明确数据产品服务价格、使用要求、权限范围等，根据需要签订保密协议，保证数据使用安全合规。提供数据产品服务的商务增值类流程包括需求受理、数据归口审核、签订合同（协议）、提供服务四个环节，由需求承接方按要求开展需求对接，沟通数据产品服务方式，履行数据归口审核流程，签订数据使用合同或协议并开展合同和协议的依法合规审查审核，组织开展数据产品设计开发并提供服务。直接提供明细数据的商务增值类流程包括需求受理、需求分析和初审、保密合规审核、数据归口审核、签订合同（协议）、提供服务六个环节。由需求承接方受理需求，数据产生部门开展专业审核；包括涉密数据的，需求承接方、保密办、法律部进行保密合规审查审核；互联网部开展数据归口审核。审核通过后，需求承接方与需求提出方签订数据使用合同或协议，组织提供数据。

对外模式（提供数据产品和服务）：

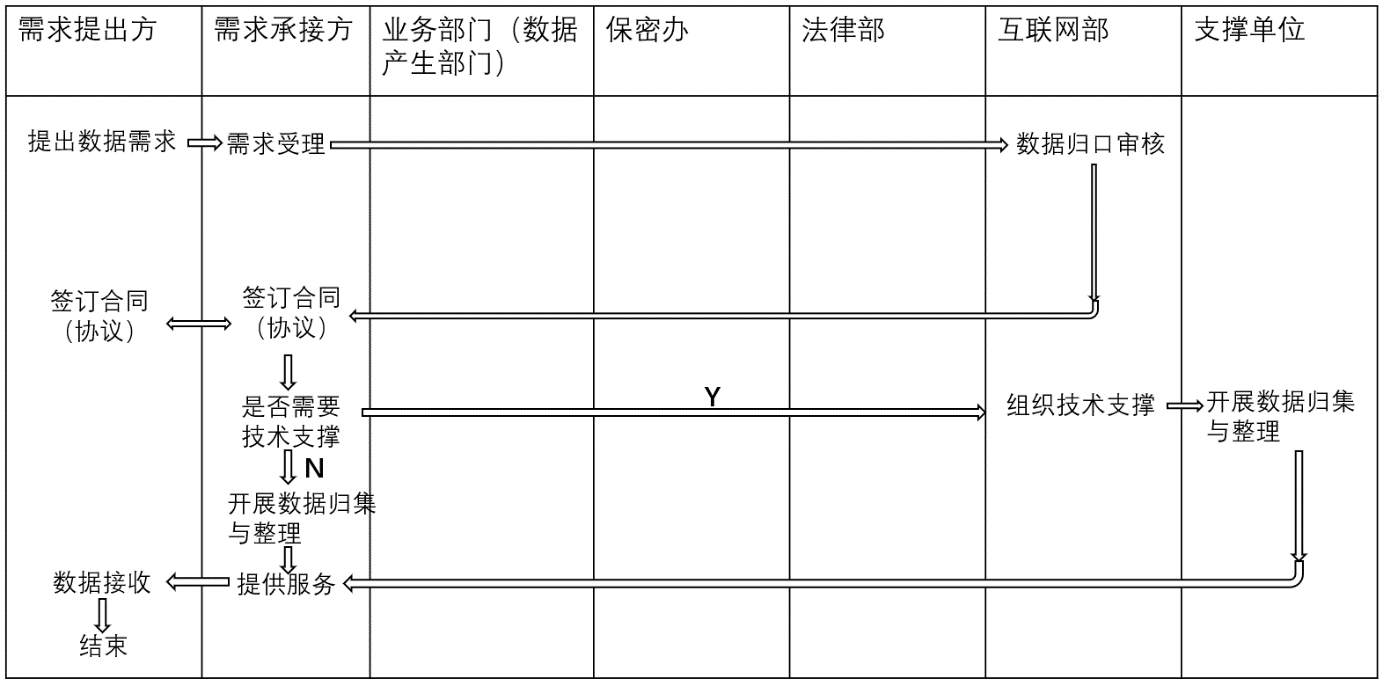


图 3-12 商务增值类

职责划分：由需求承接方对接需求，明确数据开放目的、需求内容、开放方式、申请理由等，判别需求类型，填写数据对外开放申请表。然后交由互联网部开展数据归口审核，从数据归口管理角度，分析数据提供方式、数据内容、数据安全、技术支撑等，提出审核意见。需求提出方与承接方签订合同协议，如果对外开放数据需要技术支撑，则由互联网部组织技术支撑，技术支撑单位开展数据归集与整理工作，否则由业务部门直接开展数据归集与整理为数据需求方提供数据服务。

对外模式（提供明细数据）：



图 3-13 商务增值类对外模式

职责划分：由需求承接方对接需求，明确数据开放目的、需求内容、开放方式、申请理由等，判别需求类型，填写数据对外开放申请表。业务部门进行需求分析和初审，分析其数据开放目的、开放内容、开放方式等，提出审核意见。判断相关数据是否泄密，存在泄密数据的，同步履行涉密审查审核手续。如果要对外开放的数据涉密，则由保密办进行保密审核和法律部进行合规性审核，当审查审核不通过时，相关部门应给出建设性意见，共同提出解决方案。保密合规审核不通过的，终止数据对外开放流程。审核通过互联网部开展数据归口审核，从数据归口管理角度，分析数据提供方式、数据内容、数据安全、技术支撑等，提出审核意见。需求提出方与承接方签订合同协议，如果对外开放数据需要技术支撑，则由互联网部组织技术支撑，技术支撑单位开展数据归集与整理工作，否则由承接方直接开展数据归集与整理为数据需求方提供数据服务。

安全需求：应确保遵从国家法律法规、数据主体授权范围、公司相关规章制度等数据安全保护要求。对外发布特定数据产品和服务，需按照国家、行业相关法律法规要求，获得相应的备案、资质或牌照，并报业务部门和数据管理工作归口部门备案。数据产品和服务提供方与使用方应签署相关使用协议，明确约束使用规则等双方权利义务。在数据传输环节中，采用公司统一密码基础设施签发的密钥或证书，实现企业重要数据的加密传输。在数据使用环节中，应结合数据业务场景采用脱敏、水印、审计等技术手段实现差异化防护，遵循最小授权原则访问和处理个人信息和企业重要数据。加强数据安全监测、预警、审计和处置能力建设，提高对各类数据安全事件发现、响应和溯源能力。最后应根据公司要求合理选择数据恢复、擦除与销毁措施进行处理。

表 3-18 风险描述

|  |  |  |
| --- | --- | --- |
| 风险类型 | 风险描述 | 解决方案 |
| 数据传输风险 | 业务部门产生的数据经由不安全信道上传到数据中台供统一调度或数据中台提供数据经不安全信道传输给外部企业，数据被篡改 | 利用加密技术，采用公司统一密码基础设施签发的密钥或者证书加密传输，保障其安全性 |
| 业务部门产生的数据经由不安全信道上传到数据中台供统一调度或数据中台提供数据经不安全信道传输给政府监管部门，数据发生泄露 | 利用哈希算法，采用公司统一密码基础设施签发的密钥或者证书加密传输，保障其真实有效 |
| 数据非法转发风险 | 数据文件存在由业务部门直接转发给非授权人员的情况 | 将业务部门转发数据的过程记录在区块链上，发生问题时查找记录，定位责任方 |
| 数据中台直接将数据转发给非授权人员的情况 | 数据中台向外分发的数据信息同样记录上链，发生问题时查找记录，定位责任方 |
| 数据泄露风险 | 数据文件不经过区块链直接从业务部门非法转发 | 对要分发的数据文件打上水印，根据泄露的数据文件的水印信息确定业务部门或者数据中台是数据泄露方 |
| 数据信息由数据中台链下分发 | 对要分发的数据文件打上水印，根据泄露的数据文件的水印信息确定业务部门或者数据中台是数据泄露方 |
| 数据存储风险 | 业务部门或者数据中台将数据文件存储在本地或者云服务器中 | 存储重要数据时应加密存储，落实重要数据备份、访问控制、安全审计等技术应用 |
| 数据使用风险 | 数据在接收方的不合理使用 | 与需求方签订数据使用合同，规范数据使用策略，结合数据业务场景采用脱敏、水印、审计等手段实现差异化防护 |

4） 公共开放类

定义：面向公司外部各类机构、电力用户和社会公众等，按照国家和政府要求开放收费标准、用电容量、电费情况等数据，提供相关数据查询服务。该类需求有利于打造开放透明的企业形象。

对外策略：公共开放类需求由相关业务部门负责，按照信息公开相关制度要求，以信息发布和查询服务方式主动响应数据开放需求。通过国家电网公司外网门户、“网上国网”APP等指定渠道发布，涉及单位商业秘密及个人信息的，应限定查询范围，保证数据使用安全合规。公共开放类流程主要包括需求分析与初审、提供服务两个环节。由相关业务部门主动响应政府开放需求，开展需求分析及方案制订，并组织实施。必要时可征询保密办、法律部意见。

对外模式：

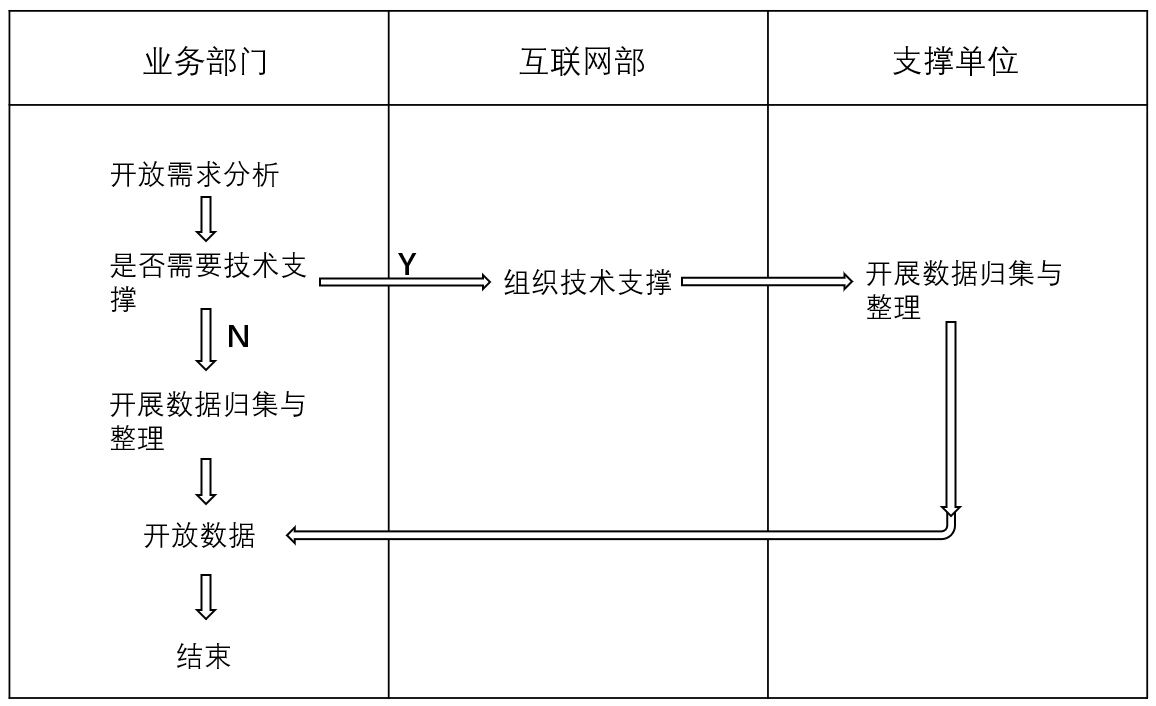


图 3-12 公共开放类

职责划分：业务部门进行需求分析和初审，分析其数据开放目的、开放内容、开放方式等，提出审核意见。如果对外开放数据需要技术支撑，则由互联网部组织技术支撑，技术支撑单位开展数据归集与整理工作，否则由承接方直接开展数据归集与整理为数据需求方提供数据服务。

安全需求：对利用互联网电子渠道交互或发布用户的业务数据，应采用符合公司安全防护方案的业务数据交互方式，并经安全专家委员会审查和安全检测机构测评。防止非授权发布或交互企业重要数据以及个人信息，扰乱互联网秩序等违法违规活动。在数据传输环节中，采用公司统一密码基础设施签发的密钥或证书，实现企业重要数据的加密传输。在数据使用环节中，应结合数据业务场景采用脱敏、水印、审计等技术手段实现差异化防护，遵循最小授权原则访问和处理个人信息和企业重要数据。加强数据安全监测、预警、审计和处置能力建设，提高对各类数据安全事件发现、响应和溯源能力。

表 3-19 风险描述

|  |  |  |
| --- | --- | --- |
| 风险类型 | 风险描述 | 解决方案 |
| 数据传输风险 | 业务部门产生的数据经由不安全信道上传到数据中台供统一调度或数据中台提供数据经不安全信道传输给电力用户和社会公众等，数据被篡改 | 利用加密技术，采用公司统一密码基础设施签发的密钥或者证书加密传输，保障其安全性 |
| 业务部门产生的数据经由不安全信道上传到数据中台供统一调度或数据中台提供数据经不安全信道传输给政府监管部门，数据发生泄露 | 利用哈希算法，采用公司统一密码基础设施签发的密钥或者证书加密传输，保障其真实有效 |
| 数据非法转发风险 | 数据文件存在由业务部门直接转发给非授权人员的情况 | 将业务部门转发数据的过程记录在区块链上，发生问题时查找记录，定位责任方 |
| 数据中台直接将数据转发给非授权人员的情况 | 数据中台向外分发的数据信息同样记录上链，发生问题时查找记录，定位责任方 |
| 数据泄露风险 | 数据文件不经过区块链直接从业务部门非法转发 | 对要分发的数据文件打上水印，根据泄露的数据文件的水印信息确定业务部门或者数据中台是数据泄露方 |
| 数据信息由数据中台链下分发 | 对要分发的数据文件打上水印，根据泄露的数据文件的水印信息确定业务部门或者数据中台是数据泄露方 |
| 数据存储风险 | 业务部门或者数据中台将数据文件存储在本地或者云服务器中 | 存储重要数据时应加密存储，落实重要数据备份、访问控制、安全审计等技术应用 |

3.4.3 数据可用性和防外泄之间的约束关系

数据可用性通常指的是数据以某种形式存储时的可用性，存储形式一般指通过网络或外部存储介质远程存储数据，可用性主要是说数据对于试图访问它的人来说的可靠性。

防止数据外泄，可以通过使用加密、追踪检查等手段来实现，但同时也会影响到数据的可用性。过于复杂的防外泄手段，可能导致数据难以使用，也就偏离了数据收集的意义。

我们一般通过访问控制，来保障一定防外泄能力前提下，使数据能较好被利用。一般来说访问控制就是依据授权，对提出的资源访问请求进行控制，防止未授权的访问，避免未经授权的使用、泄露、销毁和篡改。

除控制访问权限之外，还需要关心数据本身的安全性，包括数据提交的可信性、合规性和质量，避免风险入库，影响线上安全。为了不影响数据的业务应用，通常可以进行数据的脱敏，防止数据外泄。也可以添加水印，进行数据外泄后的溯源处理。

3.4.4 云存储

数据共享中的一个关键问题即为数据存储问题，目前在数据共享中普遍使用的云存储方式带来了许多的便利同时也带来了许多威胁风险。关注云存储技术及其安全是数据共享中的重要一环。

近年来，随着云计算和软件即服务(SaaS)的兴起，云存储成为信息存储领域的一个研究热点。与传统的存储设备相比，云存储不仅仅是一个硬件，而是一个由网络设备、存储设备、服务器、应用软件、公用访问接口、接入网和客户端程序等多个部分组成的系统。云存储提供的是存储服务，存储服务通过网络将本地数据存放在存储服务提供商(SSP)提供的在线存储空间。需要存储服务的用户不再需要建立自己的数据中心，只需向SSP申请存储服务，从而避免了存储平台的重复建设，节约了昂贵的软硬件基础设施投资。

云存储这个概念一经提出，就得到了众多厂商的持续关注。Amazon公司推出弹性块存储(EBS)技术支持数据持久性存储；Google推出在线存储服务GDrive；内容分发网络服务提供商CDNetworks和云存储平台服务商Nirvanix结成战略伙伴关系，提供云存储和内容传送服务集成平台；EMC公司收购BerkeleyDataSystems，取得该公司的Mozy在线服务软件，并开展SaaS业务；Microsoft公司推出WindowsAzure，并在美国各地建立庞大的数据中心；IBM也将云计算标准作为全球备份中心扩展方案的一部分。

（1） 云存储技术

云存储系统与传统存储系统相比，具有如下不同：第一，从功能需求来看，云存储系统面向多种类型的网络在线存储服务，而传统存储系统则面向如高性能计算、事务处理等应用；第二，从性能需求来看，云存储服务首先需要考虑的是数据的安全、可靠、效率等指标，而且由于用户规模大、服务范围广、网络环境复杂多变等特点，实现高质量的云存储服务必将面临更大的技术挑战；第三，从数据管理来看，云存储系统不仅要提供类似于POSIX的传统文件访问，还要能够支持海量数据管理并提供公共服务支撑功能，以方便云存储系统后台数据的维护。基于上述特点，云存储平台整体架构可划分为4个层次，自底向上依次是：数据存储层、数据管理层、数据服务层以及用户访问层。云存储平台整体架构如图所示。

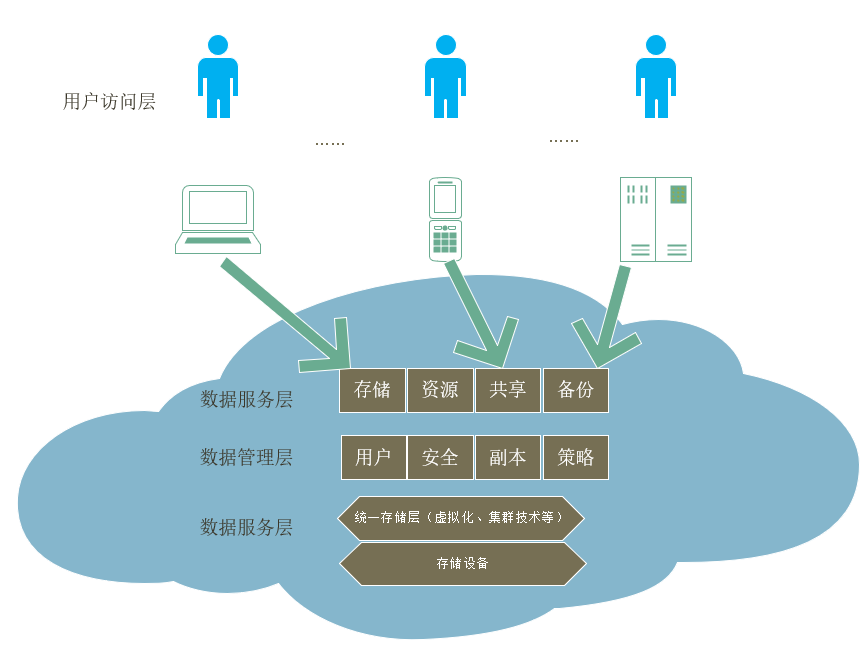


图 3-13 云存储架构

1) 数据存储层：云存储系统对外提供多种不同的存储服务，各种服务的数据统一存放在云存储系统中，形成一个海量数据池。从大多数网络服务后台数据组织方式来看，传统基于单服务器的数据组织难以满足广域网多用户条件下的吞吐性能和存储容量需求。基于P2P架构的数据组织需要庞大的节点数量和复杂编码算法保证数据可靠性。相比而言，基于多存储服务器的数据组织方法能够更好满足在线存储服务的应用需求，在用户规模较大时，构建分布式数据中心能够为不同地理区域的用户提供更好的服务质量。云存储的数据存储层将不同类型的存储设备互连起来，实现海量数据的统一管理，同时实现对存储设备的集中管理、状态监控以及容量的动态扩展，实质是一种面向服务的分布式存储系统。

2) 数据管理层：云存储系统架构中的数据管理层为上层提供不同服务间公共管理的统一视图。通过设计统一的用户管理、安全管理、副本管理及策略管理等公共数据管理功能，将底层存储及上层应用无缝衔接起来，实现多存储设备之间的协同工作，以更好的对外提供多种服务。

3) 数据服务层：数据服务层是云存储平台中可以灵活扩展的、直接面向用户的部分。根据用户需求，可以开发出不同的应用接口，提供相应的服务。比如数据存储服务、空间租赁服务、公共资源服务、多用户数据共享服务、数据备份服务等。

4) 用户访问层：通过用户访问层，任何一个授权用户都可以在任何地方，使用一台联网的终端设备，按照标准的公用应用接口来登录云存储平台，享受云存储服务。

（2） 云存储方式优点

1) 成本低、见效快：传统的购买存储设备或软件定制方式下，企业根据信息化管理的需求，一次性投入大量资金购置硬件设备、搭建平台。软件开发则经过漫长的可行性分析、需求调研、软件设计、编码、测试这一过程。往往在软件开发完成以后，业务需求发生变化，不得不对软件进行返工，不仅影响质量，提高成本，更是延误了企业信息化进程，同时造成了企业之间的低水平重复投资以及企业内部周期性、高成本的技术升级。在云存储方式下，企业除了配置必要的终端设备接收存储服务外，不需要投入额外的资金来搭建平台。企业只需按用户数分期租用服务，规避了一次性投资的风险，降低了使用成本，而且对于选定的服务，可以立即投入使用，既方便又快捷。

(2) 易于管理：传统方式下，企业需要配备专业的IT人员进行系统的维护，由此带来技术和资金成本。云存储模式下，维护工作以及系统的更新升级都由云存储服务提供商完成，企业能够以最低的成本享受到最新最专业的服务。

(3) 方式灵活：传统的购买和定制模式下，一旦完成资金的一次性投入，系统无法在后续使用中动态调整。随着设备的更新换代，落后的硬件平台难以处置。随着业务需求的不断变化，软件需要不断地更新升级甚至重构来与之相适应，导致维护成本高昂，很容易发展到不可控的程度。而云存储方式一般按照客户数、使用时间、服务项目进行收费。企业可以根据业务需求变化、人员增减、资金承受能力，随时调整其租用服务方式，真正做到“按需使用”。

3.4.5 云存储安全

（1） 属性加密

此类技术在应用过程中，需要将云存储系统的属性作为基础，也就是允许用户在系统中进行文件、数据的加密处理和解密处理。该技术将按照具体的属性合理使用加密技术，设计相关的访问系统，将所有的授权机构、授权用户融入其中。授权的一方生成用户属性密钥，对其进行严格的管理，用户可以按照实际情况实现文件的加密操作。在各个层面进行系统加密的情况下，可以维护、保障云存储系统的安全性。

属性密码体制是一种新型的公钥密码体制，它通过多个属性构成的集合模糊地描述身份信息，为数据提供保护的同时还可以实现灵活的访问控制，因此非常适合于构造云上数据安全共享。属性密码体制主要包含属性加密方案和属性签名方案两种密码原语。属性加密方案中包含一个与属性集合相对应的访问策略，当且仅当属性集合满足了访问策略时数据用户才可以解密密文。属性签名方案中则包含一个与属性集合相对应的断言，当且仅当属性集合满足了断言时才可以对消息进行签名。其中访问策略与断言都是一种结构化的声明，它们都规定了什么样的属性集合才是合法的属性集合。

属性密码体制主要包括属性加密方案和属性签名方案，本节将分别对其形式化定义和安全模型进行简要介绍：

1） Hash函数

Hash函数的单向特征和输出数据长度固定的特征使得它可以生成消息或其他数据块的“数据指纹”（消息摘要或hash值），用于消息认证和数字签名等区域。

hash值的生成过程可以表示为h=H（M），其中M是“任意”长度的消息，H是hash函数，h是固定长度的hash值。

H可以用于“任意”长度的消息，“任意”是指实际存在的。H产生的hash值是固定长度的，这是hash函数的基本性质对于任意给定的消息M，容易计算H（M）值，这是要求hash函数的可用性。

2） Hash函数的性质

抗第一原像（单向性）：对于给定的hash值h，要找到M使得H（M）=h在计算上是不可行的。

抗第二原像（抗弱碰撞性）：对于给定的消息M1，要发现另一个消息M1，满足H（M1）=H（M2）在计算上是不可行的。

抗强碰撞性：找任意一对不同消息M1、M2，使H（M1）=H（M2）在计算上是不可行的。

消息对应hash值的每一比特应与消息的每一个比特有关联。当消息原文发生改变时，求得的消息摘要必须相应的变化。

哈希函数的独特性质使其在消息完整性校验中具有重要的应用价值。随着现代密码学的发展，人们发现哈希函数的应用价值远远超出完整性校验的范畴。包括属性密码体制在内，现有的许多密码体制的构建都引入了哈希函数。为了在实际场景下合理化运用哈希函数，同时在安全性证明中便于推导分析，学者提出了随机预言机的概念：

3） 随机预言机

随机预言机是一种仅理论上存在的黑盒函数。这个黑盒函数会不断接收询问，每次询问接收一个输入并给出一个对应的输出。对于每一个输入而言，其输出结果都是在随机预言机值域中随机地、均匀地选择出来的。对于相同输入的询问，随机预言机总会给出相同的输出结果。随机预言机可以理解为一个抽象化的哈希函数，它必须满足以下三个性质：

* 均匀性：输出符号串在[0,2𝑙]上均匀分布。
* 确定性：给定相同的输入符号串，总能获得相同的输出符号串。
* 有效性：给输入定符号串𝑥∈{0,1}∗，在关于𝑥的长度的多项式时间内可以计算出对应的输出符号串ℎ=𝐻(𝑥)。

随机预言机在属性密码体制研究中具有重要的理论意义。在安全性证明中随机预言机代表了一种不确定输入输出状态的哈希函数，而方案的实际设计时则使用实用的哈希函数作为替代。随机预言机的引入不仅有利于构建更高效的密码体制，还能够充分利用可证明安全理论的优势。

4） 属性加密的形式化定义

对于一个属性加密方案而言，其构成主要包含以下四个核心算法：

* 创建：该算法以一个安全参数k为输入，然后输出公钥PP和主密钥MSK。
* 私钥生成：该算法将公钥PP、主密钥MSK以及属性集合S(或访问策略A)作为输入，然后输出对应的私钥SK。
* 加密：该算法将公钥PP、一段消息明文m以及访问策略A(或属性集合S)作为输入，然后输出对应的消息密文CT。
* 解密：该算法将消息密文CT、私钥SK作为输入，当且仅当属性集合S是访问策略A的授权集合时算法输出正确的消息明文m根据以上的形式化定义，如果属性集合S是私钥生成算法的输入而访问策略A是加密算法的输入，则称该属性加密方案为密文策略属性加密方案。相反地，如果访问策略A是私钥生成算法的输入而属性集合S是加密算法的输入，则称该属性加密方案为密钥策略属性加密方案(KeyPolicyAttribute-BasedEncryption,KP-ABE)。

5） 属性签名方案的形式化定义

对于一个属性签名方案而言，其构成主要包含以下四个核心算法：

创建：该算法以一个安全参数k为输入，然后输出公钥PP和主密钥MSK。

私钥生成：该算法将公钥PP、主密钥MSK以及属性集合S(或访问策略A)作为输入，然后输出对应的私钥。

签名：该算法将公钥PP、私钥SK、一段消息明文m以及访问策略A(或属性集合S)作为输入，然后输出对应的消息签名𝜎。

验证：该算法将公钥PP、消息明文m以及消息签名𝜎作为输入，当前仅当属性集合S是访问策略A的授权集合时输出1表示𝜎是消息明文m的合法签名，反之则输出0表示𝜎不是消息明文m的合法签名。

与属性加密方案类似，如果属性集合S是私钥生成算法的输入而访问策略A是签名算法的输入，则称该属性签名方案为签名策略属性签名方案(SignaturePolicyAttribute-BasedEncryption,SP-ABS)。相反地，如果访问策略A是私钥生成算法的输入而属性集合S是签名算法的输入，则称该属性签名方案为密钥策略属性签名方案(KeyPolicyAttribute-BasedEncryption,KP-ABS)。

（2） 增强云存储系统安全功能方法

为了保证云存储系统的安全功能，预防出现病毒入侵、非法攻击的现象，维护系统之内文件和数据信息的安全性，应该合理使用增强安全功能的方法，以此维护整体数据和文件存储的安全性。

主要的方法为：

1） 数据分割的方法：数据分割技术措施，可以对用户的文件进行分割处理，使其形成很多碎片，利用碎片的方式将数据信息提交到相关的系统之内。在此情况下，系统内所存储的数据信息存在位置不明的现象，非法人员攻击期间难以针对所有碎片进行定位处理，在一定程度上能够达到增强安全功能的目的。为降低技术应用成本，在使用数据分割方法期间，可以只进行分割，无须开展编码工作，创建相应的分割模型。但是需要注意的是，对于用户端来讲，应用文件期间需要恢复数据碎片，保持在原本的文件状态，所以此类方式应用期间，需要将用户文件还有碎片的关系记录下来，而非法人员可能会利用盗取记录的形式获取到用户完整的数据和文件，也可能会出现代理服务节点中所存储的代码被篡改或是盗取的现象，导致数据信息和文件出现安全性的问题。因此，在实际的工作中，应该按照具体状况应用数据分割技术措施，一方面，应该着重预防出现代理服务节点方面的代码篡改现象、盗取的现象，可以应用完整性度量技术措施、远程证明技术措施等，针对代码进行可信度的分析，明确有无安全隐患问题，一旦发现存在不足或是缺陷，就要针对性的处理，维护系统数据和文件的安全性。另一方面，在工作中应该确保新设置的安全功能代码不会受到干扰，这样才能确保安全功能的良好应用，建议使用防代码篡改的监测系统、信息安全管理系统等，实时性监控代码的情况，以此保证安全功能的完善。

2） 合理使用数据加密方法：加密属于目前网络环境中经常应用的维护数据信息安全性的方式，为增强云存储系统的安全功能，在工作中可以使用对称加密算法，在代理层的系统中设置密钥生成器，随机性的生成对称性密钥，然后开展用户文件和数据信息的加密处理与解密处理工作。采用此类方法进行数据信息和文件的保密处理，应重视密钥安全管理系统的设计，通过分层次的管理方式，合理开展各个层级密钥的加密存储工作，上级系统可以对下级系统的密钥进行保护，起到良好的安全功能增强的作用。与此同时，为了能够增强用户数据信息的完整性、保密性，还应该按照每位用户的数据文件情况，自动化生成对称密钥，便于开展数据、文件的加密处理工作，预防出现用户数据信息、文件的安全隐患问题。需要注意的是，在增强安全功能期间还应该在存储容器中设置容器密钥，在数据密钥加密之后才能允许数据和文件存储在云端区域，以此来解决机密性方面的问题。建议在设置容器密钥的过程中，为了增强机密性，可以整合用户的容器密钥，将其设置成为容器密钥盒子，利用私密共享算法将其分割成为门限值相同的很多私密共享模块，使用关键词层次派生算法，自动化生成有关共享模块的索引，然后创建索引树，将用户口令当做其中根节点，更好的进行机密性管理。

上述两种方法在应用的过程中，需要注意先使用分割加密的方式进行用户数据、文件的处理，使其成为加密性的内容，分割成为很多碎片。再使用云存储系统存储碎片部分，在存储容器方面设置密钥盒子，将其加密之后存储在云系统中。每个用户都需要使用私密共享方式、层次派送算法等，生成有关的秘密共享模块，然后利用云存储系统分布性的存储。

3.5 本章小结

本章首先介绍了三种不同的企业分发方式：数据中心、被动、主动方式，并介绍了相应的分发架构以及对比，之后主要介绍数据的价值点以及受到的风险和威胁情况。第二部分介绍了电力数据的特点和风险，以及电力数据分发场景和应用类型的解析。第三部分给出了数据外泄的情况和相应的安全需求，最后介绍了云存储技术和优点以及属性加密作为访问控制技术保障云存储的安全以及一些云存储安全措施。

基于区块链的电力数据共享库构建技术

4.1 一般共享数据库描述

4.1.1 数据库结构

共享数据库是一个综合性系统，其结构内容十分复杂，共享数据库业务中不仅包含业务系统数据源，同时也涵盖核心、历史、主题等数据库形式。其中核心数据库是整个数据库系统的中心，其内部还包括很多子数据库，主要包括：

基本数据库：基本数据库作为核心数据库的核心，内部含有企业设备信息、位置信息、生产信息等，这些信息的使用实现了数据库功能，并且可以在平台上共享。

业务数据库：该数据库主要是存储各类信息数据，可以实现信息查询、分析。

统计数据库：该数据库是在基础数据上进行设计分析，根据时间、地域、限制条件等划分为二次数据。数据信息是基本数据的产物，而不是业务系统直接生成的数据，可以反映大量业务报表和统计结果。在生产方式确定下，可以自动推出相关业务数据信息，数据精准性受到业务数据真实性影响。该数据库还可以对每个层次的数据进行统计、分析，从而满足用户的不同要求。

数据字典库：该数据库主要是描述共享数据，包括数据的元数据、来源、用户、应用系统描述等，可以对各个节点信息进行安全验证。

管理信息库：该数据库主要用于存放共享数据的信息，包括用户注册信息、授权信息以及数据适用状况等。

平台日志库：该数据库主要是存储数据交换中的相关信息。

数据字典库对各类信息进行安全验证后，管理信息库通过存储的权限信息、使用信息等来完成控制，最后由平台日志库完成对数据更改与分发过程的记录，如图4-1所示。

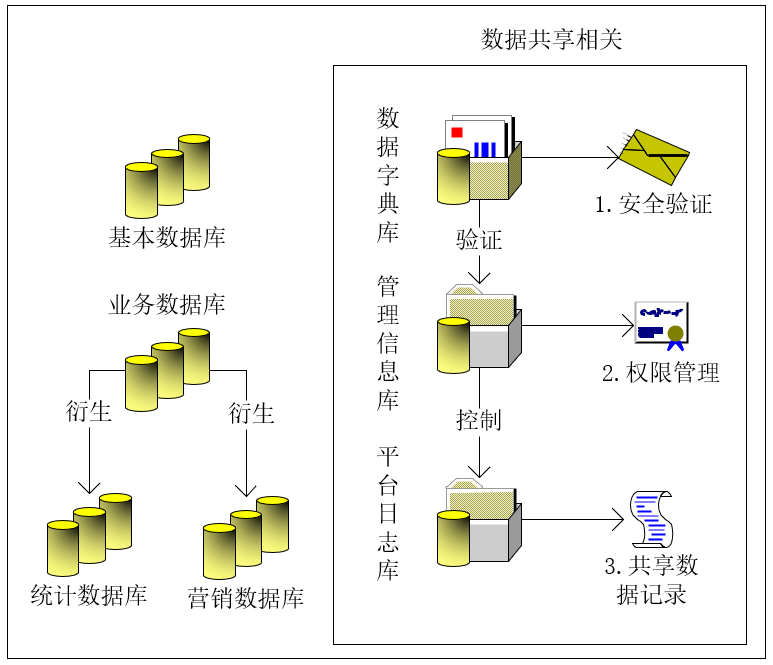


图 4-1 一般共享数据库

4.1.2 共享权限结构

要建设共享数据库，首先就要明确数据的来源和流向。通过使用Ｕ（use）／Ｃ（create）矩阵的方式可以直观、清晰地表现出一个过程对相关数据的使用或者在这个过程中所产生的一些数据从而进一步分析企业哪些部门产生的哪些数据提供给数据中心，以及哪些部门需要从数据中心获取数据。

表 4-1 U/C矩阵举例

|  |  |  |  |
| --- | --- | --- | --- |
|  | 运维部门 | 统计部门 | 营销部门 |
| 售电方案 |  | U | C |
| 设备维护 | C | U |  |
| 活动宣传 |  | C | C |

4.1.3 数据联邦技术

该项技术可以使用客户机方位、使用的数据信息，整合处理多元化信息。具有异构性、透明性、开放性、功能性等特点，在建立、处理信息方面具有很强的作用。数据联邦的核心思想是整合信息，也就是整合各项分散数据，在各项数据不相互干扰情况下，保持数据信息的高度统一。但这种整合并非是强制性，而是对数据进行自由排列组合，让使用者更加灵活的使用数据信息。在数据联邦系统当中，中心数据库可以实现操作功能；数据字典负责编排、记录信息数据，包括各部门业务分类，为用户提供更加精准清晰的信息数据；通过数据映射满足不同使用者需求，让信息处理更加人性化；通信接口负责传递相关信息数据。

4.1.4 信息共平台中共享数据库的使用优势

（1） 加强各部门的信息共享

采用数据联邦、ETL、中间件技术，可以构建性能强、安全性好、稳定性高频的数据库结构模式。构建了比过去更加完备的系统和网络框架，保证了信息共享速率和流通速度，实现了海量电力信息的存储、处理、使用。多元化的数据库结构可以判断不同类型、时间、需求的信息数据，极大的提高了电力信息系统运行效率，也加强了数据使用者的体验。

（2） 强化各部门协作，提高运行能力

本文所涉及的共享数据库应用了多层数据库，各个数据库之间可以相互配合、联通，每个功能模块都可以各司其职。在数据处理当中，应用共享数据库可以有效的降低工作强度，让电力管理进一步朝向信息化、现代化方向发展。在过去，由于电力信息较为分散、难以达成统一、信息之间无交流，各部门需要较长时间进行交流互动，严重影响了电力信息的使用安全性和便捷性。在使用共享数据库之后，各部门之间可以第一时间了解其他部门的情况，加强了信息沟通，最大程度上节省了不必要的时间浪费，同时还可以确保信息的稳定性、精准性。此外，信息共享数据库间还可以自行的对不一致、错误数据进行处理，在电力运行数据产生变化之后，系统会自动记录更新，有效提升了电力运行的稳定性、安全性。

（3） 提高了指挥系统信息化水平

在共享数据库支持下，电力指挥系统也更加智能，让电力技术和信息技术融合，指挥系统可以第一时间掌握电力信息，及时制定相关指导方案，保证电力合理使用。在对信息处理当中，结合不同需求、不同时间对紧急信息进行应急判断，实现指挥系统的快速指挥应急，避免过去由于信息流通不当缺乏实时性造成的电力使用损失。整个共享平台采用了数据系统配合方法，结合电力安全使用信息数据，时刻监管电力使用情况，为电力使用提供安全环境，利用中间件控制，避免或紧急处理突发性的随机事件。闭环的数据库结构，还可以防止信息泄露，确保电力使用安全。

（4） 信息化与人性化建设

构建电力信息共享平台，可以发挥信息数据的最大价值。首先，可以为电力安全使用提供信息保障，第一时间处理业务系统中所传达的指令；其次，对电力、行政管理人员来说，可以实现办公信息化、智能化、现代化，不仅能够提高工作质量和效率，还可以减少人工劳动力投入量；最后，对于用户来说，信息化可以为其提供更加便捷的咨询信息，让电力服务更加人性化，提高电力使用服务质量和用户满意度。

4.2 区块链技术

4.3.1 区块链基础架构

区块链是一种分布式的数据库与计算架构，具有不可篡改、可开放访问以及可审计等优良特性。区块链可以分为：私有链、联盟链以及公有链。私有链中的参与者是相互信任的，由中心化管理者统一管理。私有链中的信息是不公开的，只有少数参与者可以使用私有链中的信息，私有链一般用于组织内部的信息交流。公有链中的参与者是相互不信任的，且公有链中的所有信息都是公开的，这些信息由公有链中的参与者共同使用和维护，公有链现阶段主要以比特币和以太坊为代表。

联盟链介于私有链和公有链两者之间，联盟链中的参与者具有一定的信任基础，联盟链通过控制访问权限，从而确保联盟链中信息的安全性，联盟链现阶段主要以超级账本为代表。区块链本质上是整个交易网络中存储交易记录的分布式数据库。区块链中的参与者在无需建立信任的情况下，通过遵守相同的共识机制，将交易记录存储在区块中。使用哈希函数以链式结构将每个区块依次级联起来，从而确保交易记录的不可篡改和可溯源。每个区块由交易记录和区块头两部分组成。区块头由版本号、父区块哈希值、Merkle树、时间戳、目标哈希以及随机数组成。区块的具体结构，如图4-2所示。

其中，版本号表示所使用的软件和协议的版本信息。父区块哈希值是将各个区块级联起来的哈希值，确保了区块链的链式结构。Merkle树由区块中所有交易记录的哈希值逐级哈希计算得到，可用于检验交易记录的完整性。目标哈希又可以称为难度值，是区块链中数学难题的难度系数。时间戳是每个区块所产生的具体时间。随机数是解决数学难题的答案。

区块的具体形成过程如下：首先，矿工将本地的交易记录存储在区块链的一个区块中，并生成本地交易记录的Merkle树。其次，矿工将区块的哈希值存储在Merkle树根中，交易记录的哈希值存储在Merkle树的叶子节点中，叶子节点连接字符串的哈希值存储在Merkle树非叶子节点中。由于交易记录以这种树状的数据存储模式进行存储，所以只要篡改Merkle树中的一条交易记录，则Merkle树根中的哈希值也会随之发生改变。再次，矿工使用SHA256哈希算法将上一个区块头中的数值转化成一个哈希值，这个哈希值也就是父区块哈希值。矿工将父区块哈希值填入到当前区块的区块头中，从而实现各个区块之间的链接。最后，矿工将当前区块添加到区块链中的具体时间存储在时间戳中。时间戳不仅能够准确的记录区块的创建时间，而且能够反映区块创建的先后顺序，从而确保了区块中交易记录的真实性，进而实现交易记录的不可篡改和可溯源。一旦当前所产生的区块成功添加到区块链中，那么区块链中的所有参与者马上开始组织生成下一个区块。

目标哈希的难度值会根据之前一段时间内区块链中产生区块的平均时间进行调整，以应对整个区块链中不断变化的计算能力。如果计算能力增强了，则区块链会调高难度值，使得生成下一个区块的时间依然在预期的时间范围内。区块资源管理器是区块链的重要组成部分。通过访问区块资源管理器，从而获得存储交易事务的区块的确切信息。

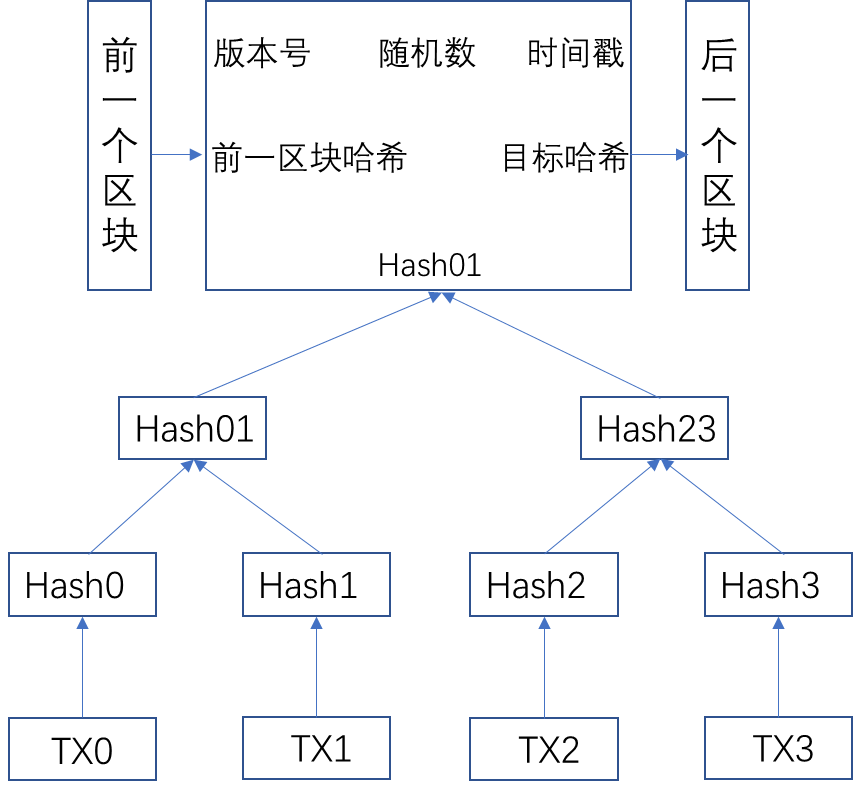


图 4-2 区块链

从广义的角度看，区块链是一种分布式的基础架构技术，主要由六层组成，如图4-3所示，分别为：数据层、网络层、共识层、激励层、合约层以及应用层。

其中，区块链基础结构的最低层是数据层，数据层由区块中所包含的各种数据组成，作用是实现区块上数据的可溯源和不可篡改。网络层由区块链点对点的组网模式和参与者间的通信机制组成，作用是确保区块链参与者间可以点对点的有效通信。共识层由多种共识机制组成，不同的应用场景所使用的共识机制会有所不同，作用是确保区块链上的所有合法参与者数据的最终一致性，解决了多参与者的数据不同步存储问题。激励层由分配机制和发行机制两部分组成，作用是鼓励区块链上的所有合法参与者参与到维护区块链数据安全的工作中，良好的分配机制和发行机制是确保区块链持续安全运转的基础。合约层是区块链中的一种特殊的数据交互方式，主要由智能合约、算法和脚本代码三部分组成，使区块链可以自动的执行预定逻辑，从而使数据交互更加的灵活。应用层由各种实际应用场景组成，用户提供针对不同应用场景的程序和接口，通过部署应用场景，进而实现数据的交互，而不用再考虑区块链的底层技术。但随着区块链的不断发展和改进，区块链的整体架构体系也在不断发生变化，传统架构体系中的一些模块逐渐被弱化甚至被取消。

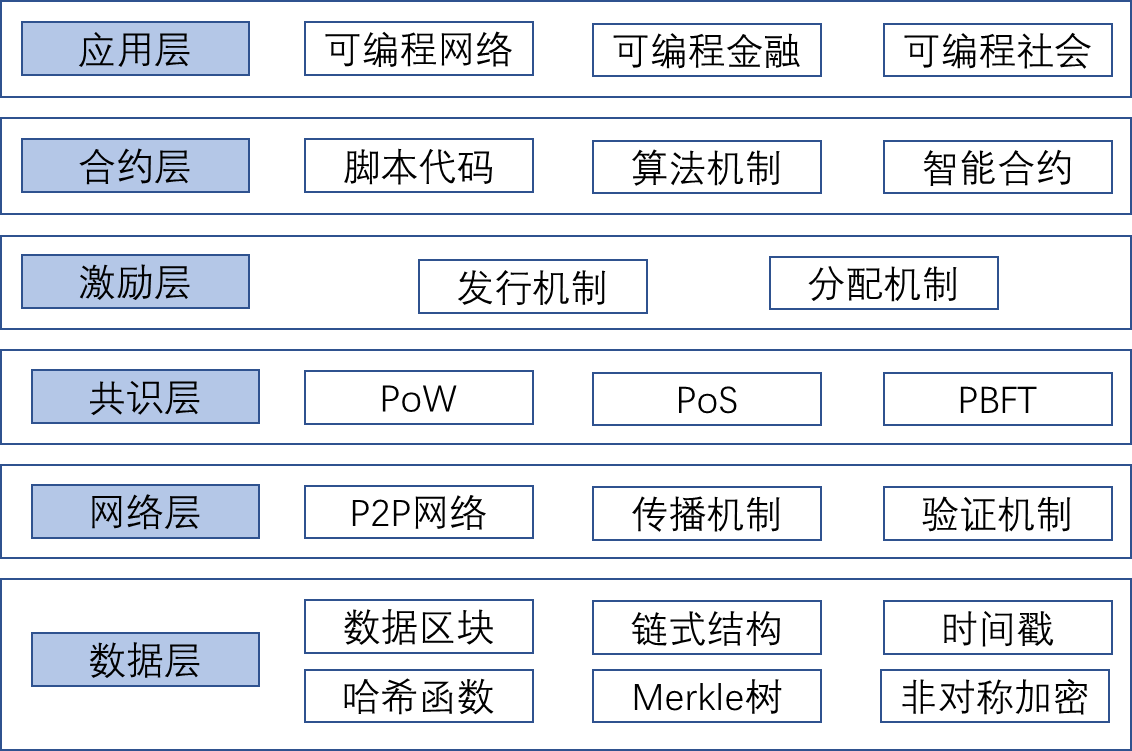


图 4-3 区块链基础结构

4.3.2 共识机制

区块链是一种分布式的数据库账目，各个账目由区块链中的每个参与者各自进行存储。这种特殊的存储方式，可能会导致不同参与者所存储的账目信息不一致。这主有以下两方面原因：一方面，区块链中的单个参与者可能会因为硬件问题或者网络问题宕机，所以不能正常的接收新区块，从而导致单个参与者的账目信息与区块链中其他参与者的账目信息不一致。另一方面，可能有参与者故意传播错误的区块信息，从而造成单个参与者的账目信息与区块链中其他参与者的账目信息不一致。上述问题被称为拜占庭将军问题（ByzantineGeneralsProblem，BGP）。因此为了确保区块链持续正常地运行，区块链中的参与者就需要运行一个一致性协议，使区块可以成功地添加到区块链中，从而确保区块链中参与者账目的一致性，这样的一致性协议被称为共识机制。在区块链中添加区块又被称为挖矿，负责添加区块的参与者又被称为挖矿矿工。为了确保区块链的持续正常运行，需要根据其不同的应用场景设计合适的共识机制。

常见的共识机制有：

（1） 实用拜占庭容错共识机制（PracticalByzantineFaultTolerance，PBFT）

PBFT共识机制是一种可以解决拜占庭将军问题的共识机制。但在公共系统中很难使用PBFT共识机制，主要有以下两方面原因，一方面，PBFT共识机制需要巨大的计算量。另一方面，PBFT共识机制需要首先知道整个区块链中参与者的总数量，之后根据参与者的总数量设置区块链中允许非法参与者存在的最大数量。目前，PBFT共识机制被运用于Ripple和Stellar中。

（2） 工作证明共识机制（ProfofWork，PoW）

PoW共识机制的思想是通过让参与者做一项复杂但又直截了当的工作来证明其具有挖矿能力。一旦完成了这项工作，参与者就可以获得挖矿权。PoW共识机制是基于哈希算法的，可用于发送电子邮件。PoW共识机制的优势在于以下两个方面，一方面，操作简单，容易实现。另一方面，参与者间无需交换额外的信息即可达成共识。但PoW共识机制的挖矿成本较高，且挖矿过程较为复杂。目前，PoW共识机制被运用于比特币中。

（3） 权益证明共识机制（ProfofStake，PoS）

PoS共识机制是PoW共识机制的一种替代共识机制。在PoS共识机制中，拥有权益越多的参与者越容易获得挖矿权，进而获得更多的权益。PoS共识机制的优势在于，它解决了PoW共识机制中存在的资源浪费、效率低等问题。但PoS共识机制同样也存在一些缺点，比如容易造成权益集中化，即参与者的权益越多就越容易获得挖矿权，进而获得更多的权益。再比如容易造成权益的重用，即参与者使用同一权益进行多轮挖矿竞争。最先使用PoS共识机制的区块链项目是2012年产生的peerCoin。目前，以太坊在第四阶段开始使用PoS共识机制。

（4） 行动证明共识机制（ProfofActivity，PoA）

PoA共识机制是PoW共识机制的一种替代共识机制，是PoW共识机制和PoS共识机制的结合。PoA共识机制与PoW共识机制的相同之处在于，都是通过让参与者做一项复杂但又直截了当的工作来证明其具有挖矿能力。PoA共识机制与PoW共识机制的不同之处在于，PoA共识机制所挖掘的区块只有区块头和挖矿的奖励地址，不包含任何交易信息。区块头中的信息将被用于选择参与者对区块进行挖掘。一旦挖掘了这个几乎为空的区块，PoA共识机制就会使用PoS共识机制中的思想，即所持权益越大被选中挖掘新区块的机率也就越大。PoA共识机制的优势在于，它比PoW共识机制和PoS共识机制都更安全。但PoA共识机制同样继承了PoW共识机制和PoS共识机制的缺点，比如容易出现权益集中化。目前，PoA共识机制被运用于Decred加密货币和Espers加密货币中。

（5） 燃烧证明共识机制（ProfofBurn，PoB）

PoB共识机制是目前相对流行的区块链共识机制，被认为是未来替代PoS共识机制的备选方案之一。PoB共识机制的核心思想是通过销毁权益来证明参与者对区块链的投入，从而获得挖矿权。PoB共识机制的优势在于，资源消耗少且安全性高。但PoB共识机制存在以下两方面的缺点，一方面，会造成很多不必要的资源浪费。另一方面，挖矿权会被那些愿意燃烧更多权益的参与者所掌控。目前，PoB共识机制被运用于Slimcoin加密货币中。

（6） 可检索证明共识机制（ProfofRetrievability，PoR）

PoR共识机制类似于PoB共识机制，但考虑到了共识机制的带宽和可检索性。PoR共识机制通过定期的检查区块，来确定在某一时刻挖矿矿工是否进行了正常的挖矿。PoR共识机制可以防止女巫攻击、外源攻击和生成攻击。女巫攻击是指通过创建多个同样物理存储空间的账号来竞争挖矿权，以赚取不合理的奖励。外源攻击是指凭借带宽优势和速度优势，获得超出自己物理挖矿能力的挖矿权。生成攻击是指恶意的矿工声明挖掘了超过实际数量的区块，从而获取不合理的奖励。PoR共识机制可以运用于分布式文件存储的费用支付。目前，PoR共识机制被运用于Filecoin加密货币中。

（7） 时间消逝证明共识机制（ProfofElapsedTime，PoET）

PoET共识机制通过使用可信执行环境来提高PoW共识机制的效率。PoET共识机制通常运用于联盟链的挖矿矿工选择。PoET共识机制是基于彩票系统公平原则的，即每个参与者都有均等的机会成为挖矿矿工。PoET共识机制要求区块链中的每个参与者在定时器上设置一个随机等待时间，第一个完成指定等待时间的参与者就可以成为挖矿矿工。PoET共识机制的优势在于，比较节能且不需要昂贵的硬件。但在PoET共识机制中，确定完挖矿矿工后，参与者还需要等待区块被添加到区块链上。目前，英特尔公司发起的SawtoothLake项目中使用了PoET共识机制。

（8） 委任权益证明共识机制（DelegatedProfofStake，DPoS）

DPoS共识机制的原理是让每一个权益相关者进行投票，产生一定数量的权益相关代表，由这些权益相关代表进行挖矿。DPoS共识机制相比于PoS共识机制优势在于，减少了挖矿所需的参与者数量，提高了达成共识的效率。但DPoS共识机制仍依赖于权益的分配。目前，DPoS共识机制被运用于EOS中。EOS（Enter prise Operation System）是为商用分布式应用设计的一款区块链操作系统。

表 4-2 共识机制对比

|  |  |  |  |
| --- | --- | --- | --- |
| 共识机制 | 适用场景 | 优点 | 缺点 |
| 实用拜占庭容错共识机制 | 内网分发 | 通信复杂度O(n^2)、能容纳一定数量故障节点、作恶节点 | 计算量巨大、需要知道参与者总数量 |
| 工作证明共识机制 | 外网分发 | 简单、容易实现，无需交换额外信息 | 完成工作量成本高 |
| 权益证明共识机制 | 内/外网分发 | 解决资源浪费、效率低问题 | 权益集中问题 |
| 行动证明共识机制 | 外网分发 | 不获取交易信息、更安全 | 权益集中问题 |
| 燃烧证明机制 | 内/外网分发 | 资源消耗少且安全性高 | 资源浪费 |
| 可检索证明共识机制 | 外网分发 | 防止女巫攻击、外源攻击、生成攻击 | 资源浪费 |
| 时间流逝证明共识机制 | 内/外网分发 | 节能且对硬件要求不高 | 额外的区块上链成本 |
| 委托权益证明共识机制 | 外网分发 | 减少参与者数量 | 权益集中 |

4.3 基于区块链的电力数据分发记录模式

4.3.1 数据溯源

数据溯源是一个新兴的研究领域，诞生于20世纪90年代，由“data provence”翻译而来。某些文献将其称为数据志或数据档案，之后大部分文献将其命名为数据起源，有追踪数据的起源和重现数据的历史状态之意。而我们在本文中将其称为数据溯源，从应用的角度出发，强调追踪的过程和方法。数据溯源是一种溯本追源的技术，根据追踪路径重现数据的历史状态和演变过程，实现数据历史档案的追溯。

建立一个有效的数据模型是数据溯源技术关键所在，根据模型可以初步确定数据溯源的大体步骤，以及数据溯源的基本思路。

（1） 主要的数据溯源模型介绍

1） 流溯源信息模型

由6个相关实体构成，主要包括流实体(变化事件实体、元数据实体和查询输入实体)和查询实体(变化事件实体、接收查询输入实体，包括元数据实体)。实体间关系密切，通过这种密切的关系可以根据数据的溯源时间来推断数据溯源。

2） Time-ValueCentric(TVC)模型

又称时间值中心溯源模型，是一种简单有效的溯源模型。由于过去的溯源模型无论是基于标注的还是基于过程的溯源模型都用于面向交易的系统中，并不适合高容量特定需求以及连续的医疗流。于是，提出支持医疗领域数据源特点的TVC模型专门处理医疗事件流的溯源信息。根据数据中的时间戳和流ID号来推断医疗事件的序列和原始数据的痕迹。

3） 四维溯源模型

此模型将溯源看成一系列离散的活动集，这些活动发生在整个工作流生命周期中，并由四个维度(时间、空问、层和数据流分布)组成。四维溯源模型通过时间维度区分标注链中处于不同活动层中的多个活动，进而通过追踪发生在不同工作流组件中的活动，捕获工作流溯源和支持工作流执行的数据溯源。

4） 开放的数据溯源模型OPM

在首届InternationalProvenaceandAnnotationWorkshop(IPAW)会议中，与会者对数据溯源的描述产生了一些共同的观念，并提出了一种原始的数据模型。后来，南安普顿大学等组织整理了会议的主要思想并发表了题为“The Open Provenance Model”文章，文中提及的OPM模型基本形成业界信息交换标准，定义一些具体的格式和协议就能应用到实际当中。

5） PROV模型

在2013年W3C对OPM进行了重大修改并将其作为溯源标准公开发布，命名为PROV。PROV定义了3个核心数据类型(Entity，Activity和Agent)及其之间的关系，数据关系都具有属性，由一个文档将所有的数据类型、关系、属性整合在一起。PROV针对OPM模型面临的问题和困难，提供一组定义文档来提高在异构环境下溯源信息互操作性，将OPM模型带入了新的境界。PROV从数据建模的观点出发，并且考虑到了信息表示和数据共享领域现有的技术状况。但关于PROV的定义文档过于细致复杂，而许多研究往往关注特定领域，在特定领域的PROV实现难以应用到其他领域中。

6） ProVOC数据溯源模型标准

《信息技术数据溯源描述模型》是现行的数据溯源模型标准，该标准提出了一种数据溯源模型ProVOC模型，模型定义了三个核心类型（数据、活动、执行实体）及其之间的关系。同时在核心类型内部进行了新的构件设计（包括数据集、参数等）。能较为方便根据标准确定溯源大致内容，但标准本身没有更加具体细致的针对现实问题的设计，且没有具体的溯源内容构件模型的方法介绍。

（2） 传统的数据溯源方法

除数据溯源模型以外，还有一些传统的数据溯源方法。数据溯源追踪的主要方法有标注法和反向查询法。

1） 双向指针追踪法：利用图论思想和专用查询语言追踪法，以位向量存储定位。标注法是一种简单且有效的数据溯源方法，使用非常广泛。通过记录处理相关的信息来追溯数据的历史状态，即用标注的方式来记录原始数据的一些重要信息，如背景、作者、时间、出处等，并让标注和数据一起传播，通过查看目标数据的标注来获得数据的溯源。事先标记并携带溯源信息完成数据溯源的模型，被称为eager方法。采用标注法来进行数据溯源虽然简单，但存储标注信息需要额外的存储空间。

2） 反向查询法：有的文献也称逆置函数法。由于标注法并不适合细粒度数据，特别是大数据集中的数据溯源，学者提出了逆置函数反向查询法。此方法是通过逆向查询或构造逆向函数对查询求逆，或者说根据转换过程反向推导，由结果溯到原数据的过程。这种方法仅在需要时才计算，所以又叫lazzy方法。反向查询法关键是要构造出逆向函数，逆向函数构造的好与坏直接影响查询的效果以及算法的性能，与标注法相比，它比较复杂，但需要的存储空间比标注法要小。下面将标注法与查询法进行比较，列出其优缺点。

标注法的优点是实现简单，容易管理；缺点是只适合小型系统，对于大型系统而言很难为细粒度的数据提供详细的数据溯源信息，因为对细粒度数据的实现可能导致元数据比原始数据还多，需要额外的存储空间，对存储造成很大的压力，且效率低。

逆置函数反向查询法的优点：追踪比较简单，只需存储少量的元数据就可实现对数据的溯源追踪，不需要存储中间处理信息、全过程的注释信息。其缺点是用户需要提供逆置函数(并不是所有的函数都具有可逆性)和相对应的验证函数．构造逆置函数具有一定局限性，实现相对比较复杂。

逆置函数在数据库中应用具体举例：在数据仓库定义和创建阶段，为了跟踪完整的数据集，永久存储所有中间转换结果；根据转换的性质为每个转换定义相应的跟踪查询过程。在数据志跟踪阶段，调用预定义的跟踪查询过程进行逆向跟踪，对输入数据中的输入元组进行验证。

4.3.2 数据模型

以上模型是比较经典的模型，对于这三种模型而言，流模型和时间值模型没有明确指出对W7模式的支持（基于数据对象生命周期的本体模型），只有四维模型支持动态构建数据溯源图，能根据一系列溯源事件以及数据结点和服务结点所构成的数据流边来构建。存在的不足之处在于形成过程不直接，难以理解。PROV从数据建模的观点出发，并且考虑到了信息表示和数据共享领域现有的技术状况。但关于PROV的定义文档过于细致复杂，而许多研究往往关注特定领域，在特定领域的PROV实现难以应用到其他领域中。 ProVOC模型在原来PROV基础上添加了参数、数据集等子分类，能较为方便的对溯源内容进行归类，但溯源模型本身存在一些定义模糊之处（如执行实体、数据集与数据）以及冗余的构件关系（构件之间定义的关联、具有、触）。本文的溯源模型正是在ProVOC的优点基础上，进行了模型构件定义和关系的改良得到。

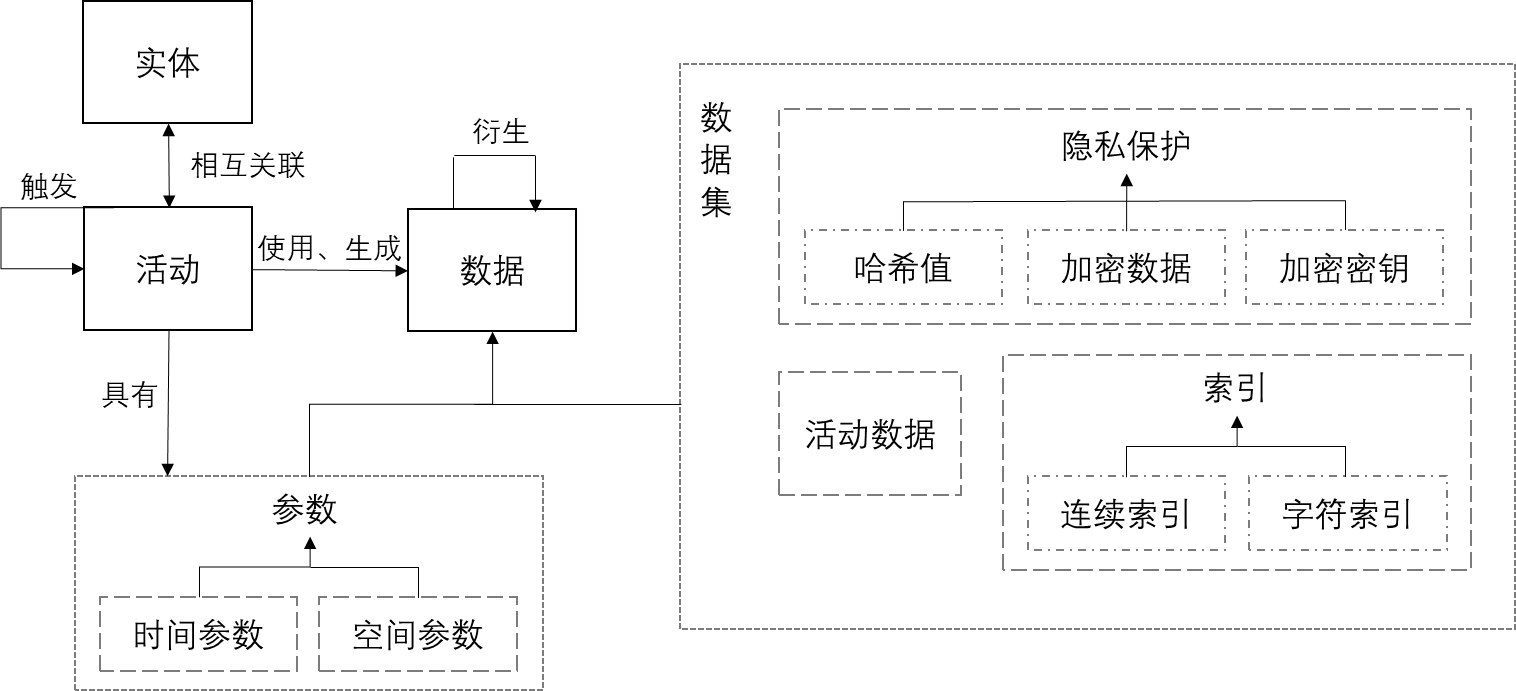


图4-4 溯源模型

本文提出的溯源模型由四级构件组成，即一级类构件、二级类构件、三级类构件、四级类构件。如表2所示，各构件内容如下：

一级构件包括活动、执行实体、数据三部分。活动：指需要进行数据溯源的活动；执行实体：指产生活动的机构、人等；数据：为了进行溯源需要记录的数据内容；其中的活动和执行实体不再划分子类构件，不进行结构化表征，仅使用一个字段来存储具体内容，可直接用于检索。

二级构件包括参数和数据集两个子类构件。参数包括时间参数、空间参数两个子类构件。时间参数为时间戳，用于指定数据的分发时间、使用时间等。空间参数为数据的存储位置。

数据集包含三个三级构件，分别为隐私保护、活动数据、索引。

其中隐私保护包含三个四级构件，分别为加密数据、加密秘钥、哈希值，用于对隐私保护数据的加密和哈希处理的保存和记录。

索引包括连续索引和字符索引两类子构件，用于在大规模溯源情况下的溯源记录，节约存储空间。当涉及到大规模溯源，如一批连续文件的索引时，无需对每个文件进行溯源记录，只需根据文件之间不同的溯源内容来构建索引，其余相同溯源内容如存储位置等仅需记录一次即可。当进行溯源时根据索引来寻找这一批文件的溯源记录即可。

活动数据指的是数据集内非隐私保护、非索引的数据条目。

4.3.3 模型构建及上链过程

我们基于上述溯源模型提出了一种快速构建溯源模型方法。在进行数据溯源工作时，首先要面临的问题是选取哪些数据作为溯源内容。尤其是现在的许多公司、单位数据库中的数据内容繁杂，而对于溯源而言，全部选取的方式会降低存储和溯源效率。如果通过专家来进行手工挑选，首先效率不高，人工成本高，且数据库内容会随时间增多，溯源内容也应当及时更新，人工方式不利于及时对溯源内容进行迭代更新。因此我们提出了一种根据数据库表键值以及列名频率来确定溯源内容的方法，快速从繁杂的数据库表内容中筛选出溯源内容，构建溯源模型。

据已有的数据库表内容快速筛选得到溯源模型所需的构件内容。如下图所示，工作流程由数据库表内容开始，通过对列名集合的筛选、融合得到更合理的溯源列集合；对隐私相关数据的处理，以及对部分列名的特征提取，得到用于隐私保护场景和大规模分发场景下的相关内容，最后根据上述内容构成最终的溯源内容。该工作流程描述如下。

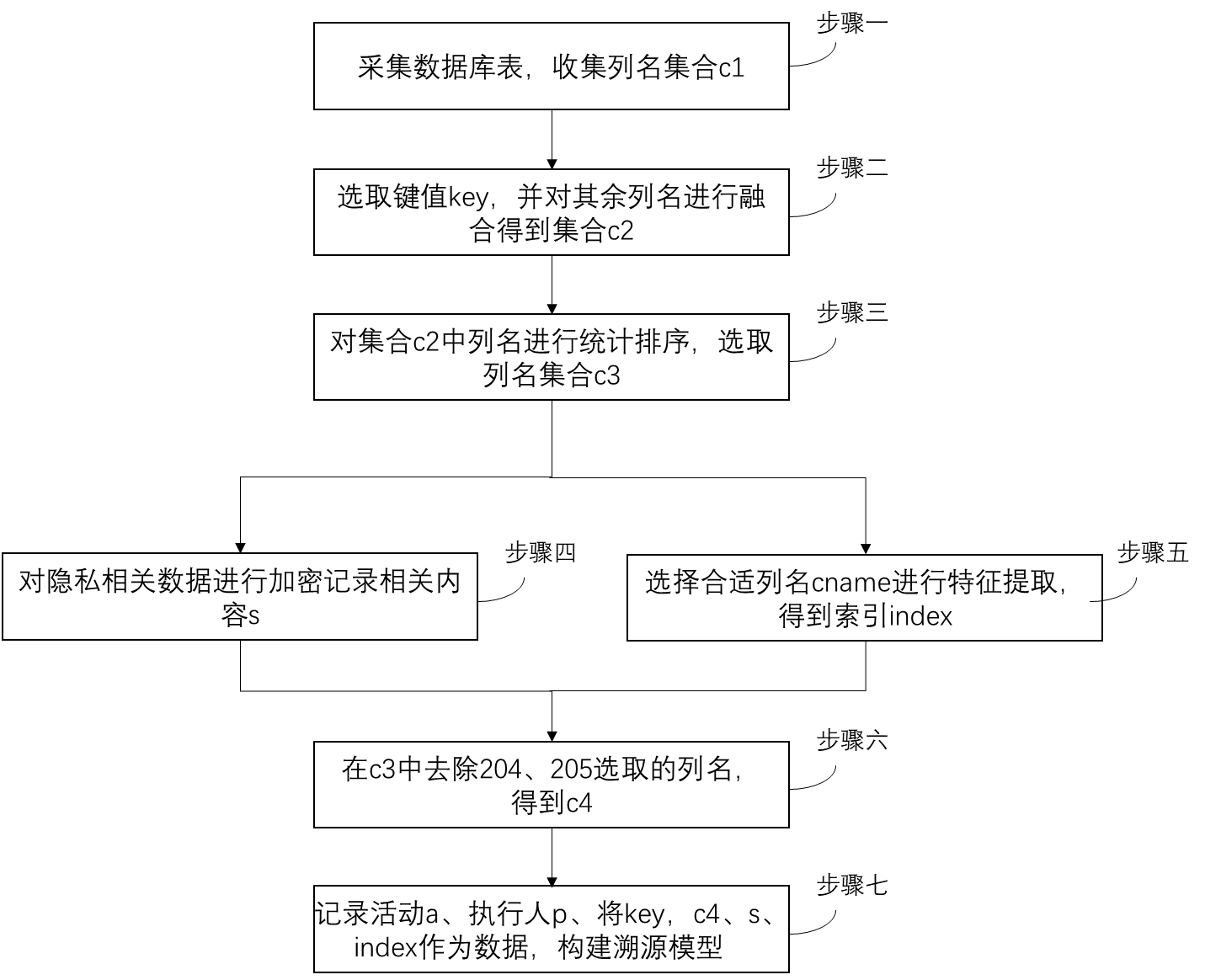


图4-5 构建方法

（1）采集数据库表，收集列名集合c1；

（2）选取键值key，并对其余列名进行融合得到集合c2；

（3）对集合c2中列名进行统计排序，选取列名集合c3；

（4）对隐私相关数据进行加密或哈希，记录相关内容s；

（5）选择合适列名cname进行特征提取，得到索引index；

（6）去除c3在上述步骤中选取的列名得到C4；

（7）记录活动a、执行人p、将key、c4、s、index作为数据，构建溯源模型；

在步骤（2）选取键值key，并对其余列名进行融合得到集合c2时。我们首先提取键值作为溯源内容，数据库表中键值是天然的重要的溯源特征，可以直接确定作为溯源内容一部分。列名融合主要针对当不同列名指向同一内容的情况，通过对内容是否一致来进行列名融合，减少列名冗余，提高溯源效率。

在步骤（3）对集合c2中列名进行统计排序，选取列名集合c3；显然列名出现次数越多，说明重要程度越高，越有可能起到溯源作用。

在步骤（4）对隐私相关数据进行加密或哈希，记录相关内容s；当使用加密时我们记录加密秘钥和加密内容，当时用哈希时我们记录哈希值。通过这一步来防止溯源信息泄露导致的隐私泄露。

在步骤（5）选择合适列名cname进行特征提取，得到索引index；特征提取根据列数据内容类型进行构建不同的索引：当数据内容为连续型数字等时，选取数字上下界作为索引；当数据内容为不连续如字符串时，选取最大匹配字串进行索引。

数据通过溯源模型得数据结构进行划分，并以活动为中心进行数据上链，根据活动类型进行上链。以活动为核心进行上链，是因为活动部分包含信息方便检索。

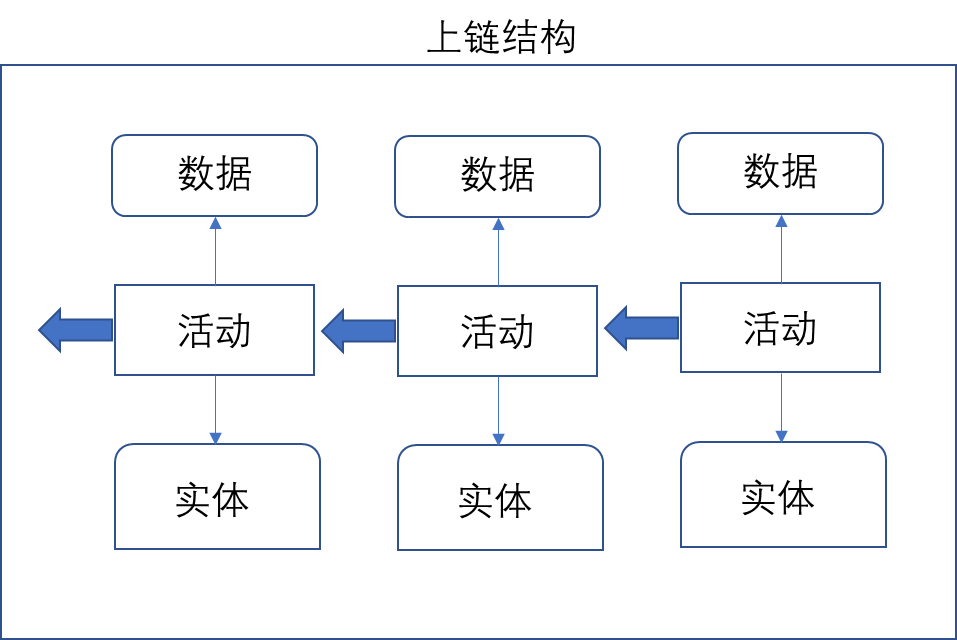


图4-6 上链结构

4.3.4 模型实例

（1）模型构建实例实施

为了尽可能体现我们方法的优越性，我们选择了来自电力系统的一组数据库表。

电力系统数据量庞大，仅以江苏省为例，为4600万用户提供电力资源，仅报修一项就有二十万报修数据[11]。如此庞大的服务量和电力系统提供的多种服务意味着一个庞大的数据库系统以及相应的繁杂的数据库表内容，直接将所有类型数据都作为溯源内容是不合理的，因而需要一种快速有效的方法来进行内容筛选。

电力系统对其他领域具有极大参考价值，不可避免涉及到大规模的对外数据分发和溯源。如电力数据对工业生产部门，尤其是大型工厂等单位的意义主要是指导这类机构优化用电策略，开展绿色生产，降低企业运营成本等。电力数据对售电方最直接的作用就是通过分析数据提升企业效益。用电数据可以帮助政府对工业生产状况、城区住房空置率、电价补贴政策影响等涉及经济宏观走势和人民日常生活的方方面面进行分析。

电力服务是国民经济的命脉，电力数据一旦遭到篡改将导致极其严重的后果。构建一个合理的溯源流程，能帮助对威胁发生时的责任处理，也能对威胁事前起到警告保护作用。

电力数据中存在大量不宜公开的隐私信息，如居民的地址、电话号码等，因此在进行溯源时，如何避免居民隐私信息被泄露是重要的溯源环节。

本文提供的快速溯源工作内容流程,如图1所示。下面我们将对具体的步骤使用实际的数据进行描述展示。我们使用了三张实际数据库表作为方法操作对象，如表三、表四、表五所示。

步骤一，采集数据库表，收集列名集合c1；对于步骤201来说，收集的数据库表如表3、4、5所示，得到一个列名集合c1。

步骤二，选取键值key，并对其余列名进行融合得到集合c2；对于步骤202来说，选取其中的键值ORG\_NO单位代码、CONS\_NO户号组成键值集合key。对剩余列名进行融合,其中的ELEC\_ADDR（用电地址）和CONS\_ADDR（用户地址）指代内容一致，进行融合只保留ELEC\_ADDR，得到C2。

步骤三，对集合c2中列名进行统计排序，选取前k个条目（此处假定k为5）得到列名集合c3；对列名出现频率进行统计排序，选取若干项列名组成集合C3。溯源方可以根据自身需求和存储要求选择相应数量的列名。

步骤四，对隐私相关数据进行加密或哈希，记录相关内容s；对与用户隐私相关的如PHONE\_NUMBER联系方式等进行加密或哈希后得到相关内容s。

步骤五，选择合适列名cname进行特征提取，得到索引index；对于步骤205来说，当需要对某地区（如表中\*\*县城）数据进行溯源记录时，选取列名ELEC\_ADDR（用电地址）中“江苏省\*\*市\*\*县”字段作为索引，仅记录列名用电地址以及该索引即可。当要对一批连续的交易进行溯源记录时，选取列名TRADE\_CODE（交易代码），记录列名交易代码，和代码起始值4710，代码结束值4766即可。该步骤得到内容index

步骤六，去除c3在步骤204，205中选取的列名得到C4，对于步骤206来说，在C3中去除步骤204、205选取的列名得到C4。对于步骤207来说，记录活动如电表收费计算，执行人某某电力公司，将key、c4、s、index作为数据，构建如图2所示溯源模型。

表4-3数据表实例（一）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 单位代码 | 户号 | 户名 | 用电地址 | 用电类型 | 线路号 | 交易代码 | 上限 |
| 30408300\*\*\* | 3620128\*\*\* | 江苏省  \*\*公司 | 江苏省\*\*市\*\*县\*\*镇\*\* | 402 | 19520000 | 4710 | 315 |

表4-4数据表实例（二）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 户号 | 户名 | 用电地址 | 用电类型 | 线路号 | 交易代码 | 伏特代码 | 上限 |
| 3620128\*\*\* | 江苏省\*\*公司 | 江苏省\*\*市\*\*县\*\*镇\*\* | 402 | 19520000 | 4710 | AC001\*\* | 315 |

表4-5数据表实例（三）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 户号 | 户名 | 用户地址 | 联系方式 | 电量变化 | 费用 | 用电类型 | 所属公司代码 |
| 36201  28\*\*\* | 江苏省\*\*公司 | 江苏省\*\*市\*\*县\*\*镇\*\* | 1879795\*\*\*\* | 19520000 | 423\* | 402 | Z\*\*\* |

（2） 数据共享实例

根据PROV模型，设计数据共享时记录在区块链上的数据结构。

实体：文件ID、文件存储位置。

活动：文件新ID，新拥有者信息，新存储位置等。

代理：发起人ID、组织ID、签名等。

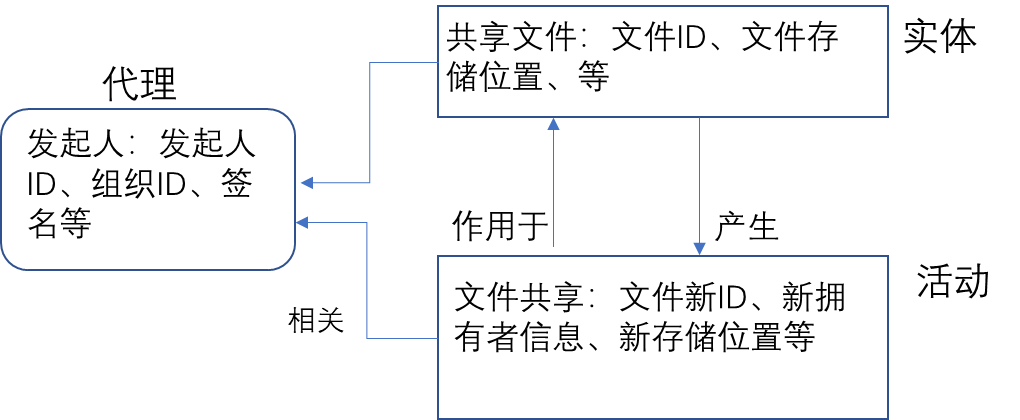


图 4-7 模型实例

（3） 业务模型实例

根据PROV模型，设计记录在区块链上的数据结构。以电力公司常规产生电力收费工单活动为例：实体为电表，活动为产生工单，代理为实体所属或执行活动的电力公司。

实体：姓名、电话、联系地址、户名（consname）户号（cons\_no）单位代码（ORG\_NO）。

活动：产生工单时间、电量差值、用电类型、费用、执行公司代码。

代理：所属电力公司代码、电力公司层级（如省、市、区等）。

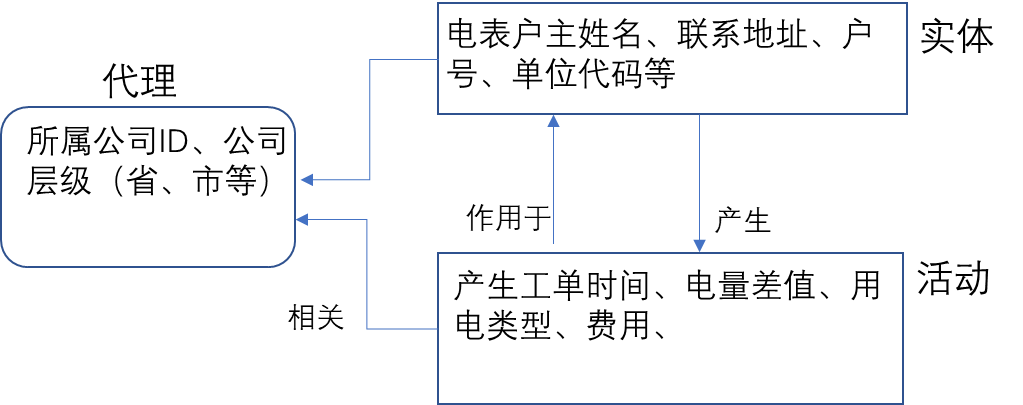


图 4-8 业务实例

4.4 基于区块链的共享方案

4.4.1 共享方案主体思想

（1） 区块链技术层

基于企业组织结构的特点，区块链技术选择以联盟链的形式实现。各部门或机构通过审核后加入区块链网络，节点之间通过P2P网络协议进行数据传播与验证；相比较公有链基于证明的共识机制，选择低延迟、无分叉、基于投票的RBFT共识机制更加合适。将智能合约以代码的形式嵌入系统，通过智能合约自动处理相关的事务。

（2） 智能合约层

将智能合约以数字化的形式写入区块链，借助区块链技术，发挥了智能合约在成本效率方面的优势，而且避免了作恶行为对合约执行的干扰，保障合约公开透明、不可篡改可追踪的公平执行。

智能合约不仅代表区块链上所有组织或用户共同认可的一种协议，更重要的是它不受某一中心机构的控制，只要满足预定的执行条件，系统随即触发相应的协议自动执行。

4.4.2 共享方案溯源流程

PROV系统模型的单个执行周期如图所示，包括6个步骤：

第一步，更改发起人修改数据文件并将其上传到云中。数据文件的不同版本均保存在云中，以便在更改被拒绝的情况下还原。

第二步，更改发起人通过客户应用程序将更改请求和押金一起提交给投票合约。更改请求包含文档ID、数据文件先前和当前版本的哈希加密形式、云存储库中的文件位置、时间戳以及发起人签名。

第三步，客户端模块提交更改，然后启动投票周期。在投票期间，授权用户使用驻留在云存储中的验证程序脚本验证更改，如果更改有效，则返回TRUE，否则返回FALSE。

第四步，用户使用投票合约根据验证结果对此次更改进行投票(该过程是自动的)。

第五步，投票合同记录用户所投的票数。在投票结束时，如果多于一半的用户投反对票，则拒绝更改，更改发起者会受到罚款，区块链系统将该罚款在参与投票的用户中分配;如果在投票结束后，投反对票的人数少于一半，则接受变更，并向变更发起人退还押金。

第六步，如果在表决后接受更改，则表决合约会将此更改记录在文档溯源跟踪器中。更改的日志条目包括负责更改的作者、当前文档的哈希值和文档先前版本的哈希值、当前更改的高级PROV表示形式以及用于将来验证的数字签名。

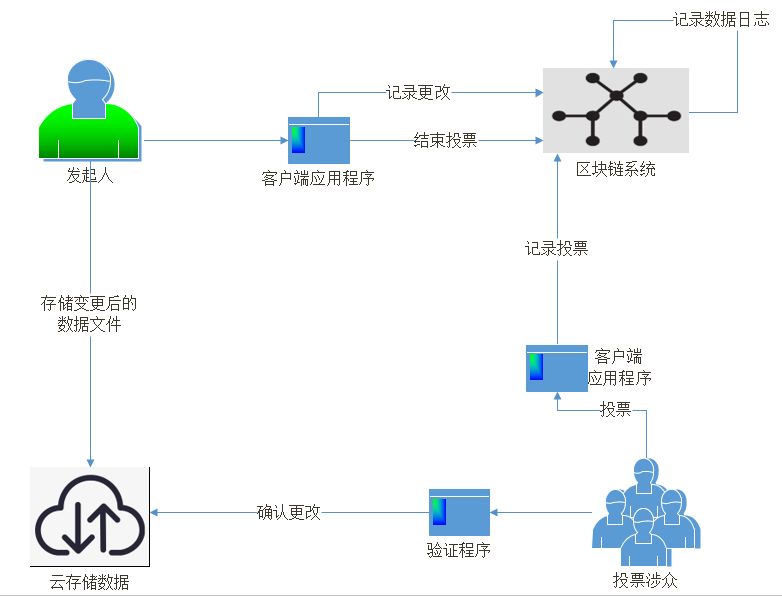


图 4-9 文档追踪合约

4.4.3 文档追踪合约

文档追踪合约用于跟踪特定文档的变化轨迹，实施访问控制策略，维护所有用户对文档的访问信息，提供用于生成特定文档的溯源轨迹的方法。所生成的数据文档溯源信息作为事件存储在文档追踪合约的事件日志中。

更改事件日志格式为：change(docid，agent，EK(docid，H(Xo)，H(Xn)，link，ts)，PROV，signk)其中，docid为数据文件的唯一标识，agent为更改发起人的地址，EK为加密文本，H(Xo)为先前版本的文件哈希值，H(Xn)为修改后的文件版本的哈希值，link为数据文档在云存储中的位置，ts为最新的时间戳，PROV为PROV溯源模型，signk为基于加密文本的发起人签名。文档追踪合约支持访问控制管理的所有基本功能，例如创建文档、用户权限管理。

系统不会在明文链中存储任何敏感信息，因为明文中存储的包括智能合约代码在内的任何信息都是公开可访问的。此外，由于存储成本和区块链存储限制，实际数据存储在区块链以外的云中。将文档添加到系统时，文档所有者会生成溯源记录的初始迭代。智能合约强制执行以下约束，即该文档溯源信息的访问权限严格由文档所有者控制，获得授权的用户禁止转让其权限。除主要方法外，文档追踪合约还包含用于检查用户对文档访问权限的帮助程序，以及用于更新文档所有者的方法。所有对溯源的更改事件都必须通过投票合约的投票过程来批准。

本方案选择投票来核实所提交的更改信息主要基于两方面的考虑：首先是要有效地防止明显违反数据使用限制的恶意更改其次是不希望验证过程泄漏任何敏感信息。另外，如果验证是在合约中完成的，则需要以不透露任何信息的方式进行此操作(如使用零知识证明，因为合同源代码和执行是公开可见的)，然而，现有零知识技术还不够通用。并且，通用零知识验证技术的效率不足以实现溯源捕获。因此，系统允许每个参与者通过客户程序在链外运行验证代码，并使用链上合约对数据更改进行投票。

4.4.4 投票合约

执行投票协议，有两种类型的投票协议，即简单多数投票和阈值投票。发起人以加密形式将所做的更改及其签名与文档ID一同提交给投票合约。投票合约接收更改，并在验证后生成日志事件以启动更改投票。投票周期设为t1，在此期间参与者可以对此更改投票(赞成或反对)。

在投票期结束时，投票合约统计票数，根据投票协议类型，决定拒绝或者接受更改。如果当前投票期的总票数未达到协议规定的最低阈值，则投票合约将开启新一轮的投票。在投票阶段结束后，如果表决结果为接受更改，则投票合约会将更改提交给文档追踪合约来生成溯源文档。投票合约仅接受特定文档的单个未完成更改，以确保数据溯源链的连续性和一致性。

（1） 多数表决投票

在多数表决投票中，文档的利益相关者对该文档的更改进行投票，接受或拒绝更改的决定基于简单多数策略，即如果多数用户对更改投了反对票，则更改被拒绝，否则接受更改。这种投票策略的缺点是，要求所有授权用户都进行投票，当文档的用户较少(少于5个)时，此策略结果非常理想，但在用户数量巨大的情况下，要求所有利益相关者为每项更改投票的代价相当昂贵。因此，当文档的涉众数量少于5个时，本方案实施简单的多数表决;而对于用户量较大时，采用随机阈值表决策。

（2） 随机阈值投票

对于包含大量更改和用户的系统，用户对每个更改进行投票的效率都不高。在此情况下，本文采用随机阈值表决策略。在随机阈值投票中，投票合约要求以最小票数接受或拒绝更改。假设文档有n个用户，要接受或拒绝更改，则投票合约的阈值为s。为了确保更改的每个投票阶段都获得s票，合约会尝试在t＞s的情况下获得预期的t票。阈值t确保每次更改都收到最少的赞成票或反对票s。

为了确定是否参与变更投票，每个客户端会根据以下公式获得一个随机Rs：

Rs=Hash(Bno，Etext，Glim，Addr)modn

Rs是通过散列发起更改信息的哈希而得到的随机数，更改信息包括Bno(当前块号)，Etext(更改事件中的加密文本)，Glim(当前gas限制)和Addr(发起者的地址)。如果生成的随机数低于投票合约设置的阈值t(即Ks＜t)，则客户端将根据验证脚本的结果进行投票。提交表决后，投票合约将以相同的方式为每个投票生成随机数，并验证所提交的表决是否有效。在随机阈值表决策略中，对变更的投票基于安全的伪随机数，由于每个哈希表的哈希函数输入几乎都是随机的，因此无法知道哪个客户对哪个变更进行投票。在投票期结束时，如果投票合约发现投票总数低于阈值s，则投票合约重新启动投票过程。如果在预定义的最大重新启动次数之后依然未收到所需的票数，则更改被拒绝，并将押金退还给更改发起人。

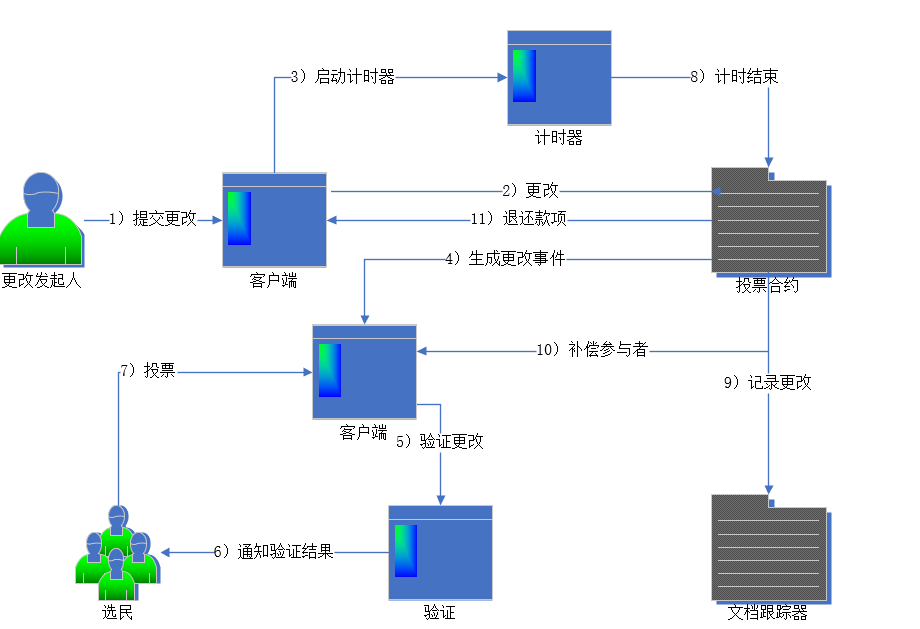


图 4-10 投票合约

4.4.5 方案安全与隐私分析

（1） 安全分析

外部攻击者可以根据数据文档ID提交无效的更改请求来尝试攻击当前系统，方案通过对文档实施访问控制策略来阻止此类攻击。文档追踪合约仅接受来自文档所有者授予访问权限的用户的更改请求，拒绝所有其他权限用户的更改请求。通过跟踪攻击尝试，预扣更改的押金将使外部攻击者受到惩罚。外部攻击者可能使用早期的更改请求签名来发起反复攻击，文档跟踪器通过跟踪特定文档的最新更改时间戳来防止这种攻击，任何迟于该文档最新时间戳的消息都将被忽略。内部用户可以是文档的所有者，也可以是所有者授予其访问文档权限的用户。后者可能通过提交不实的更改来尝试破坏数据溯源轨迹。由于系统要求每个更改都必须由最少数量的用户批准，因此，只有当其可以控制文档所允许的用户总数的一半以上时，攻击才能成功。随机阈值投票进一步确保了攻击者无法提前知道所有选民中哪些可以参与当前投票，因此这类攻击很难成功。如果内部攻击者与其他利益相关者串通并投票赞成更改，则可能破坏系统。所有者是唯一可以授予访问权限的用户，如果所有者选择一组忠实于他并有意破坏溯源轨迹的用户，那么该系统可能处于不利地位。尽管这种攻击可能是成功的，但它仍在区块链上留下了可追踪的痕迹，可用于检测攻击。

（2） 隐私保护分析

通过使用哈希散列和加密来实现对溯源数据轨迹的隐私保护。外部用户只能通过查看事件日志来推断文档ID和对特定文档ID进行的更改次数。所有更改事件对有效信息进行加密，外部攻击者可以获得的信息无外乎文档ID、密文和签名。指向实际文件的云存储的链接位置也是加密的。外部用户从合约交易记录中可以得到的信息包括哪些用户与特定的文档ID相关联，该信息可通过查看投票合约的迭代推导出来。系统平台通过使用随机公共地址为用户提供匿名服务。Prov的用户不会在环境中透露其身份，而是使用公共地址在系统中执行操作，只有文件所有者才知道文档用户的身份，查看到多个投票迭代的用户最多可以推断出与每个文档关联的公共地址。

4.5 本章小结

为了应对数字时代的不同需求，许多大型的公司采用大数据处理的新技术，以及面向协同式数据分析的分布式基础设施。本章首先给出了一般企业数据库的介绍，主要包括数据库结构、共享权限结构、数据联邦技术等。之后本文介绍了数据溯源模型，以及相关的共享方案，对 ProVOC 模型进行了改良，减少了冗余的附属关系和结构，添加了新的用于上述问题的新的构件，根据实际业务情况提出了PROV实例以及上链的过程。改善了数据溯源中的隐私泄露和溯源内容繁杂问题，提高了大规模溯源情况下的溯源效率。

基于区块链的电力数据安全分发技术研究

典型数据分发架构

5.1.1 中心化数据分发架构

(1) 分发图：

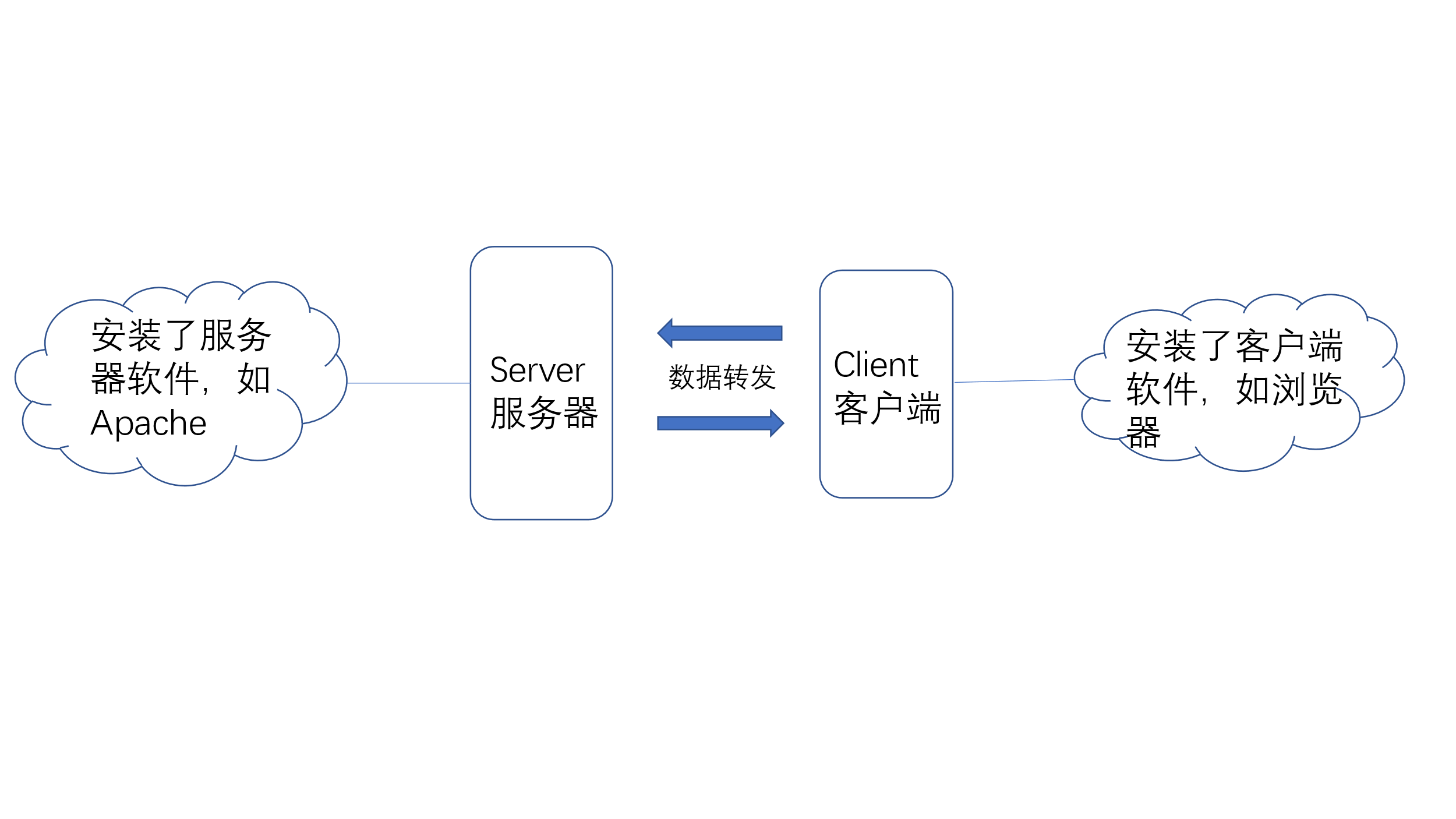


图 5-1 中心化数据分发图

(2) 分发过程：

用户向中心服务器数据库请求数据，如果数据存在，就采用直接寻址的方式下载下来，否则将用户提交的数据请求交给协同子系统进行处理，协同子系统通过信息传递、队列排队、任务调度等，将请求数据推送到指定的FTP服务器上，用户再从FTP服务器下载数据。通常利用到的CDN技术，是一种构建在网络之上的内容分发网络技术，它的基本原理是把数据从源站推送到离用户最近的服务器上，然后用户直接从离自己最近的服务器获取数据，从而获得最好的用户体验。依靠部署在各地的边缘服务器，通过中心平台的负载均衡、内容分发、调度等功能模块，使用户就近获取所需内容，降低网络拥塞，提高用户访问响应速度和命中率。

(3) 风险分析：

隐私泄露，服务器管理员有权限从服务器平台上查看和删除用户上传的文件。

中心化服务器是黑客入侵的首要目标，服务器的安全直接影响用户信息安全。

数据传输过程中可能会被监听，篡改。

服务提供商可能因为管理不善而关闭服务。

(4) 优缺点：

统一管理：由中心化机构统一管理，物理介质集中布放，信息资源集中，只要数据中心没问题，数据安全性就能得到保障。

带宽受限：大量用户同时进行访问，会造成服务器压力过大，用户体验差。

易受攻击：容易受到攻击，是恶意黑客的首要目标。

脆弱性：主机出现故障时可能使整个系统停止工作。

5.1.2 分布式数据分发架构

(1) 分发图：

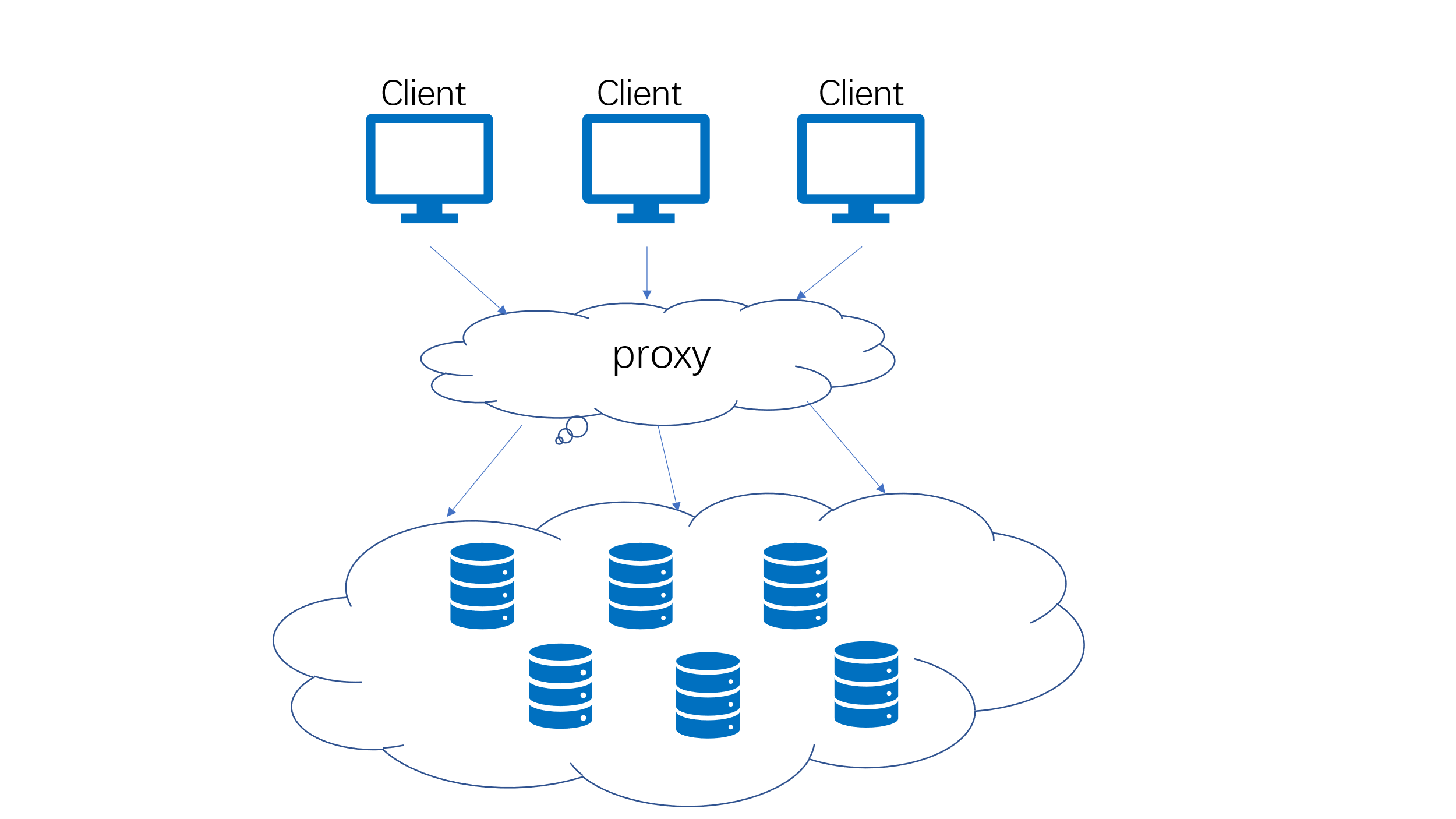


图 5-2 分布式数据分发图

(2) 分发流程：

分布式系统结构将大量普通服务器结合成整体，利用服务器定位技术来存储信息。从客户端发来的数据访问请求转发到不同的服务器上，从相应的数据库服务器中取回数据文件，有利于负载均衡，缓解网络压力，增加可靠性。

(3) 风险分析：

实际场景中，由于网络丢失数据包、故障、节点停机，可能导致数据副本不一致。

数据传输过程中可能会被监听，篡改。

(4) 优缺点：

高性能：一个具有高性能的分布式云存储能够高效地管理读、写缓存，配合高速存储来明显改变整体存储的性能，先将数据写入高速存储，来提高系统响应速度，再在适当的时间进行同步落盘。

负载均衡：每台主机缓存本地最常用的数据，不需要频繁的访问中心服务器，减轻了服务器的负担，克服计算机资源紧张的问题。

多副本备份：分布式存储采用了多副本备份机制，即在存储数据之前，分布式存储对数据进行了分片，之后数据保存在集群节点上，通过副本写入方式和多个副本读取的一致性技术，在读取数据失败的时候，系统可以通过从其他副本读取数据，重新写入该副本进行恢复。

网络环境依赖性高：分布式存储非常依赖网络环境和带宽，如果网络发生抖动或者故障，都可能影响分布式存储系统运行。

数据一致性不可靠：对于数据一致性要求比较高的应用场景，分布式存储的性能可能就稍弱了，虽然技术在进步，但对于数据同步问题，分布式不如集中式的存储方式可靠。

5.1.3 P2P数据分发架构

(1) 分发图



图 5-3 基于P2P数据分发

(2) 分发流程：

每一个节点既是客户端也是服务器，当一个节点需要获得数据时，可以从多个具有数据文件的节点同时下载数据，速度快，同时也能作为服务器向其他节点提供下载功能。

(3) 风险分析：

广域网用户直连通讯传输而没有确保安全，故传输的文件数据可能会有危害性或失真。

P2P网络是一种比较脆弱的网络，用户直接面临黑客的攻击。

(4) 优缺点：

速度快：P2P技术优势很明显，在多节点上复制数据，可实现加速下载。

节省带宽：由于多个节点互相连接，用户所在的带宽将会被最大程度的利用，提高网络用户的利用率。

健壮稳定：一个节点出现故障，并不会影响到其他节点的工作。

可扩展：将以服务器为中心的服务分散到各个网络节点，避免出现服务器性能瓶颈。

安全性低：缺少监管，各个节点身份合法性不能得到确认，可能存在恶意节点。

5.1.4 区块链数据分发架构

(1) 分发图

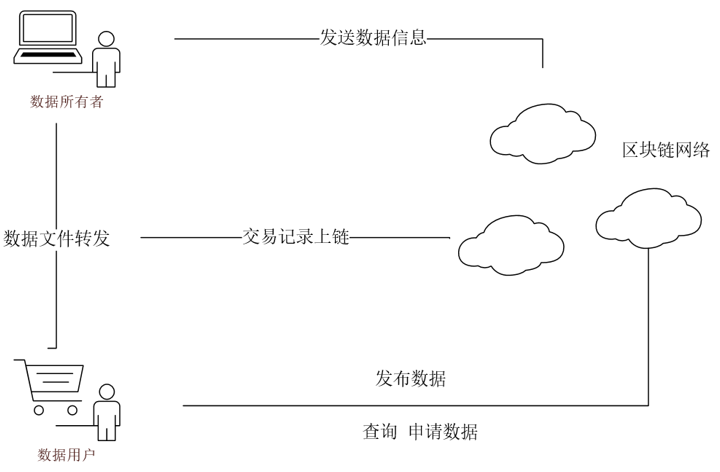


图 5-4 基于区块链数据分发

(2) 分发流程：

数据所有者向区块链网络广播自己所拥有数据文件的信息，数据用户通过区块链网络查询自己想要的数据文件相关信息，通过智能合约向数据所有者发送请求，在满足相应的访问条件后，数据所有者将数据文件发送给数据用户，并将交易记录上链。

(3) 风险分析：

不可篡改，既是优点也是缺点，一旦交易信息填错，对区块链的数据变动几乎无能为力。

交易数据必须是公开透明的，每个人可以看到其他全部的数据交易记录。

数据量变大后性能会下降。

(4) 优缺点：

安全稳定：在去中心化的区块链网络中，无中心节点可攻击。

交易可靠：去中心化的交易方法便捷而简单，无第三方介入，不需要担心信息泄露。

简约便利：由于去中心化处理方式较传统处理方式更为简单和便捷，因此在大数据量交易同时进行时，去中心化的方式会节约资源。

自主高效：去中心化的区块链技术，无需第三方介入，点对点直接交互，使得高效率、无中心化代理、大规模的信息交互方式成为现实。

不可篡改：既是优点也是缺点，对区块链记录的数据变动不可撤销，一旦被记录的区块链上，无法更改和弥补操作上的失误。

交易公开：交易数据是公开透明的，如果知道某个人的账户，就能知道他所做的每一笔交易。

性能降低：随着时间推移，交易数据量变大的时候，需要追溯历史记录就会产生性能问题。

5.2 电力数据分发风险分析及防护点

5.2.1 电力数据分发安全需求分析和风险描述

(1) 近年来，大多数电力企业虽在持续加强信息安全建设，但由于自身网络复杂、业务特殊、系统繁多等特性，依然面临严峻的安全威胁与挑战，而数据安全正是其中的重要一环。

(2) 置身互联网、大数据的时代下，电力企业、尤其是规模巨大的电网公司，其业务的高效运转越来越依赖于对数据的传输和使用。电力企业的营销、人资、财务、资产、协同、综合等核心系统中也存储着大量的业务往来、用户隐私等重要敏感数据，如若发生盗用、泄露、篡改、删除等安全事件，不仅会对电力企业自身的业务、信誉和经济利益造成严重损害，甚至可能影响能源供应，导致社会恐慌，威胁国家安全。

(3) 与此同时，国家在保障敏感数据安全方面积极制定相关法律法规。如《网络安全法》、《中央企业商业秘密安全保护技术指引》、《电力行业网络与信息安全管理办法》等，都在对电力企业开展数据安全保护提出更高、更严的要求。一方面，电力企业需要通过不断挖掘数据价值以支撑自身服务质量、工作效率和发展需要；另一方面，又要保证数据在复杂的场景、系统之间被安全、合理的访问和使用。很显然，这不是一项轻轻松松就能够实现的目标，需要电力企业看清问题所在，并有针对性的加以解决。

5.2.2 电力数据安全分发管控不足与防护点

(1) 不足：

敏感数据管理不足

随着电力行业信息化建设的持续推进，电力企业内部各部门以及跨组织、跨区域之间的电力数据传输与共享场景日渐普遍，需要通过对数据进行脱敏来实现“用、护”结合。但是，部分电力企业仍存在采用脚本或人工脱敏的情况，脱敏规则也不统一，从而导致脱敏效率低下，以及脱敏后数据质量差、数据间关联关系被破坏等一系列问题。

风险行为监控不足

电力企业规模庞大、系统繁杂、人员众多，日常工作中发生的越权访问、下载或篡改数据等违规操作行为难以及时被发现和定位，对内部数据安全事件的预防和调查造成困扰；此外，根据国内风险评价机构调研情况显示，逾两成的电力企业数据库处于直接暴露在互联网中的风险状态，且大多使用的版本十分陈旧，很多在新版本中已经得到修复的安全漏洞依然存在，甚至可能成为外部黑客入侵内网的跳板。

数据库运维管控不足

电力企业网络复杂、业务特殊、数据库众多，在运维专区中，通常是使用堡垒机来对运维人员进行管理，但这种管理方式在数据安全防护上存在一定问题：运维人员不按操作规范或既定方案进行数据库运维操作、非法导出敏感数据、数据库操作行为没有细粒度的审计记录。

(2) 防护：

梳理与脱敏

在数据共享场景下，电力企业应完成对自身数据资产的系统梳理，并根据数据的敏感程度进行分级分类，制定出数据库共享和分发的处理流程。同时，在执行数据共享的操作过程中，应遵循业务角色最小化原则，对数据进行有针对性的脱敏处理，做到安全、合理的使用数据。

考虑到电力企业的数据错综复杂，各业务数据流转通道各不相同，应按照电力企业数据的分级分类标准，对存储在内网以及临时存储在外网的数据进行发现及标记。在对数据进行跨部门共享或外发到政府部门时，应针对重要数据进行脱敏降级，确保数据接收方不会对数据内容进行二次扩散。

审计与稽核

通过部署数据库安全审计产品，电力企业能够对数据库的访问及其他操作行为进行细粒度审计与分析，从而全程监控、记录包括非法访问、数据库违规操作、数据批量导出或篡改在内的一系列风险行为，实现对所有数据访问行为进行审计记录，然后通过数据分析技术结合电力企业数据操作审计典型策略要求，对风险行为进行挖掘和预警，并在安全事件发生后，做到准确、高效的溯源定责。

运维与管理

对于电力企业而言，应在确保不影响正常开展运维工作的前提下，建立数据库运维操作的审批机制和技术措施。通过数据库运维管理系统对所有涉及敏感数据的操作进行限制，强化对数据库运维操作的监管力度，及时阻断越权操作行为的发生，令运维工作实际操作与计划操作保持一致。为此，电力企业应对数据运维操作的关键动作进行划分，将那些敏感操作梳理出来并默认禁止。当检修期间确实存在数据运维需求时，通过发起运维审批流程，根据审批小组的审批意见，有序、安全的执行运维操作。

(3) 对电力分发数据，要做到：

数据含义识别：根据既定的数据标准，通过数据语义识别等技术对数据内容进行自动化识别。

数据分类分级：按业务进行数据分类，按数据敏感程度进行分级。

数据关系梳理：根据数据特征分析，形成数据关系图谱。

数据打标：根据分析和识别结果，为数据打上类别及等级标记。

数据访问确权：采用零信任思想，对每一次数据访问的主体进行身份和权限校验。

数据共享脱敏：通过高保真、高效数据脱敏策略，快速为数据共享提供数据保护。

数据分发水印：为分发的数据提供水印保护能力，确保在数据泄露后可追溯泄露责任主体。

数据存储加密：对存储中的数据进行透明加密保护，并保障业务与安全间的平衡。

数据流向分析：监控数据从数据库到应用到终端的全程流向。

数据访问分析：监控敏感数据被哪些人员、应用访问，执行过哪些操作。

风险行为分析：针对既定的风险模型进行数据操作风险行为分析。

5.3 基于区块链的数据分发过程

5.3.1 传统数据分发服务与区块链分发的对比

受益于网络技术与智能设备的发展，网络数据信息与人们工作生活的联系愈加紧密，人们在网络上分发分享数据信息的需求越来越大。

传统数据分发通常是用户发出请求，经网络传播给服务器，服务器接收到请求后给出相应响应，完成一次数据分发过程。但是，当前的分发服务中存在许多不足之处。首先，在数据存储与传输的过程中，无论是本地存储还是逐渐普及的云存储，都存在内部泄密或外部攻击的问题，这些问题会造成数据的泄密与丢失，导致用户隐私泄露或利益损失；其次，数据共享过程中数据所有者很难精确的选择将数据分享给谁，无法做到细粒度的访问权限控制；同时，当前的数据分发服务中数据所有者往往需要将数字内容交给第三方的中心化分发平台，从而失去了对分发数据的控制与访问授权，并且第三方中心化平台拥有的权利过大，这对数据分发双方的权益存在潜在威胁；最后，传统的数据分发服务中，对于用户的数据传输过程没有进行专门的处理，只是单纯的将数据依次重复发送给每个订阅者来实现数据的发布过程。因此，在网络中存在大量的冗余数据，造成了链路带宽资源的浪费，增加了用户数据的传输时延，降低了用户体验。并且，数据平台中数据的流转和逻辑过程复杂，难以追溯数据来源，许多企业目前没有统一的数据资产标准，各业务系统中数据质量参差不齐，存在信息孤岛现象，不同部门同一名称数据可能有不同的含义，同一个数据可能又有不同的命名，数据有效交互和共享存在问题。存在部分系统数据更新不及时的问题，核心业务数据无法溯源，数据的准确性和及时性较低，现有报表在建模时几乎每个报表都要重复建模，人为参与工作过多且层次复杂，无法高效的对流程及指标进行精确监控及分析，数据的利用效率和模型重复使用率较低。

在数据分发服务中需要对数据进行有效的安全防护，建设可信的公共平台，保障用户的数据隐私，以确保数据被合法访问。数据分发服务中的主要需求可以分为三个方面：第一，保证数据传输与存储过程中的安全性，防止发生数据泄密；第二，对数据的获取设置访问权限控制。第三，在该公共平台中要保障数据分发双方的利益，在保障数据所有者对分发数据的控制权与收益权的同时，也要保障数据消费者的合法权益。

区块链是解决上述难题的有力工具，它是一种分布式的数据库，采用了P2P网络架构，不存在中心化节点，各节点共同参与区块链网络的维护、管理和运行工作。区块链网络中的节点遵循相同的规则生成新的区块，并写入本地的区块序列，数据一旦写入区块链中就很难更改。区块链网络中的数据通过加密与签名保障了数据的安全性与可验证性。区块链具有的一致性、去中心化、不可篡改和可验证等特点，是数据分发服务中可靠数据共享的技术保障。利用区块链不可篡改和可验证的特点，系统将数据分发服务中的数据信息与分发信息存储到链上，来保障数据文件的真实可靠，并利用区块链的去中心化机制，保障了数据所有者对于数据信息的控制权与收益权，同时在数据发生泄露的时候，可根据链上的信息进行追溯。

5.3.3 基于区块链分发的主体

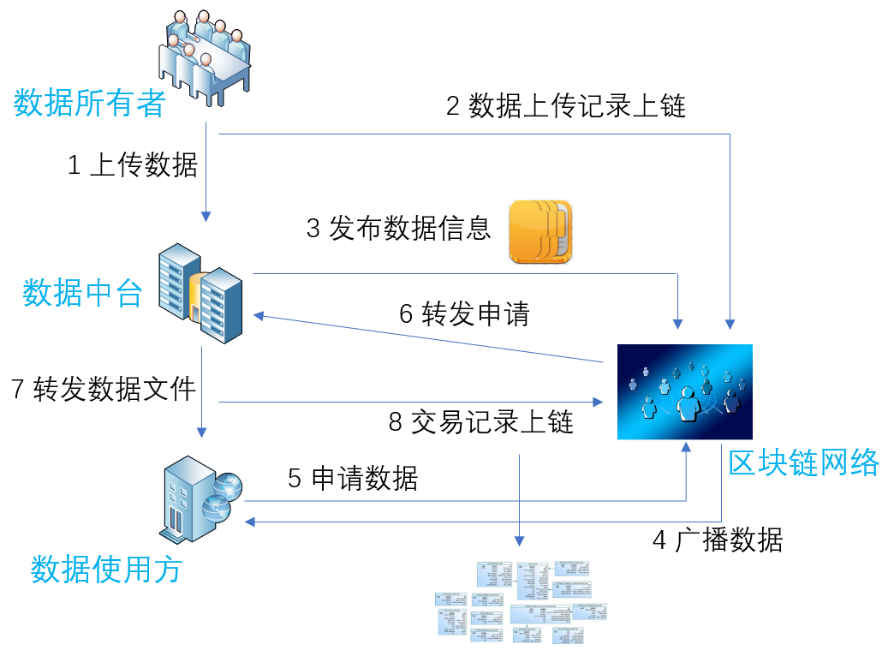


图 5-6 区块链数据分发流转图

数据所有者：是产生原始数据的单位。通常是指业务部门，由业务部门产生的业务数据经区块链网络传输到数据中台，由数据中台进行数据统一分发。

数据中台：控制数据对外开放的单位。定义数据文件访问权限的策略，并且将数据的相关信息与访问策略发送到区块链网络中，给满足访问策略的第三方单位分发数据文件，在我们的方案中数据中台享有数据的控制权。

数据使用方：也就是数据用户，用户从区块链网络中获取到数据的相关信息后可以根据自己的兴趣、与访问策略的匹配度来决定是否向数据所有者申请数据文件。数据使用方通常是公司内业务部门或者外部单位。

区块链网络：区块链网络中部署着有关数据分发的智能合约，该智能合约在链上起到分发平台的作用。合约中记录分发数据与访问策略的相关信息，并会判断申请文件的用户属性是否满足数据所有者的访问策略，如果满足则为用户与数据所有者提供分发文件的交易，否则就会终止执行合约。

5.3.2 电力区块链架构的选取

区块链根据不同场景有如下分类：

按照原创程度可划分为原链和分叉链；按照独立程度可划分为主链和侧链；按照应用划分为基础链和行业链；按照开放程度可划分为公链，私有链，联盟链。

(1) 公有链：

定义：公有链是指任何人都能参与的区块链。公有链是去中心化程度最高的区块链，不受机构控制，整个账本对所有人公开透明。任何人都能在公有链上查询交易、发送交易、参与记账。加入公有链不需要任何人授权，可以自由加入或者离开，所以公有链又称为非许可链。公有链的任何节点都是向任何人开放的，每个人都可以参与到这个区块链中的计算，而且任何人都可以下载获得完整区块链数据，即全部账本。

特点：在公有链中程序开发者无权干涉用户，所以区块链可以保护使用它进行开发的程序的用户；任何拥有足够技术能力的人都可以访问，也就是说，只要有一台能够联网的计算机就能够满足访问的条件；尽管所有关联的参与者都隐藏自己的真实身份（这种现象十分的普遍，他们通过他们的公共性来产生自己的安全性），但每个参与者都可以看到所有的账户余额和其所有的交易活动。

公有链优势:链上所有数据完全公开透明，链上任何一个成员可查看链上任何信息。

公有链缺点:交易速度慢，公链上由于要确保交易信息真实性，防止分叉攻击，需要等到此交易所在区块后的产生6个区块的时间(比特币为例)才能确认该交易真实有效，这一过程会影响交易速度。

应用：公链主要有以太坊、比特币、EOS等。

(2) 私有链：

定义：在某些区块链的应用场景下，开发者并不希望任何人都可以参与这个系统，因此建立一种不对外公开、只有被许可的节点才可以参与并且查看所有数据的私有区块链，私有链一般适用于特定机构的内部数据管理与审计。和公有链的账本对所有人公开透明和人人皆可记账的情况相反，私有链是指区块链记账权限仅在一个人或者一个机构手里，并且参与记账的权限由机构内部制定，读取权限可以为对方开放也可以任意程度地限制。

特点：交易效率高，私有链的交易速度很快，毫不夸张的说，其速度可超过任何其他的区块链。这是因为即使是少量的节点，其也具有高信任度，所以交易的进程不需要每一个节点都来验证，所有造就了私有链独一无二的交易速度。其速度之快，甚至接近了常规数据库（非区块链数据库）；保障隐私，私有链上不必处理访问权限等繁琐进程，个人数据不会被网络上任何人获得；成本低，私有链上完成的交易成本通常十分廉价。这是因为私有链上的交易速度十分之快，各个节点间不需要完全的协议，以至于它们不会为任意一个交易而工作，如此一来大大降低了交易成本。

私有链优势：交易速度更快，由于交易不需要所有节点确认，所以大大的提升了交易效率；隐私性更好，私有链只对需要开放权限的节点授权查看权限，否则节点是看不到其他成员信息的；安全性更高，链上成员都是经过审核授权的，所以恶意攻击的可能性相对较小。

私有链缺点：权限被少数节点控制，不能从根本解决作弊问题，背离了去中心化的初衷。

应用：私有链则有超级账本、Hashgraph、Corda等。

(3)联盟链：

定义：联盟链是介于公有链和私有链之间、实质上仍属于私有链范畴的区块链。联盟链与公有链的差别在于它只对特定的组织团体开放，因此在联盟链中，每个参与者都可以查阅和交易，但不能验证交易，也不能发布智能合约，简单来说，联盟链上的信息对每个人都是只读的，只有节点有权利进行验证或发布交易，这些节点组成了一个联盟。普通用户如果想发布或者验证交易，则需获得联盟的许可。因此，联盟链更类似一种分布式的数据库技术。

特点：目前联盟链使用的群体主要在金融业，主要是银行、保险、证券业。未来3年，工商业将成为下一个重要的应用领域，再之后，随着社会各界对区块链的普及与认可程度的提高，政府系统全面采用联盟链技术提供公共服务指日可待。在公有链中，一个新的区块是否能够上链，得由区块链中所有的节点来决定，所以一笔交易的真伪至少要得到全网51%的节点验证才能被确定，导致公有链对交易的处理速度很慢。而对于联盟链来说，一个新的区块是否能够上链，只要其中几个权重较高的节点进行确定就可以了，这就意味着，一笔交易不需要所有节点的确认就可以进行，大大的降低了交易处理时间；隐私保护强。联盟链上的信息并不是所有有访问条件的人都可以访问的，联盟链上的信息，只有该联盟链上的节点才可以进行读取修改和访问等活动。联盟链中的每个节点都有属于自己的一个私钥，每个节点自己产生的数据信息只有该节点自己知道，如果节点与节点之间需要进行信息交换和数据交流，就必须知道对方节点私钥，这样一来，既能够保证信息流通，又避免了节点隐私泄露的问题。

联盟链优点：节点少，处理速度快，隐私保护强。

联盟链缺点：由于节点少，如果权限设计不周则容易出现权力集中和安全问题。

应用：典型的联盟链比如Ripple区块链(为属于联盟成员的银行类金融机构提供跨境支付服务，希望取代SWIFT跨境转账平台，打造全球统一的网络金融传输协议)、超级账本(HyperLedger)。

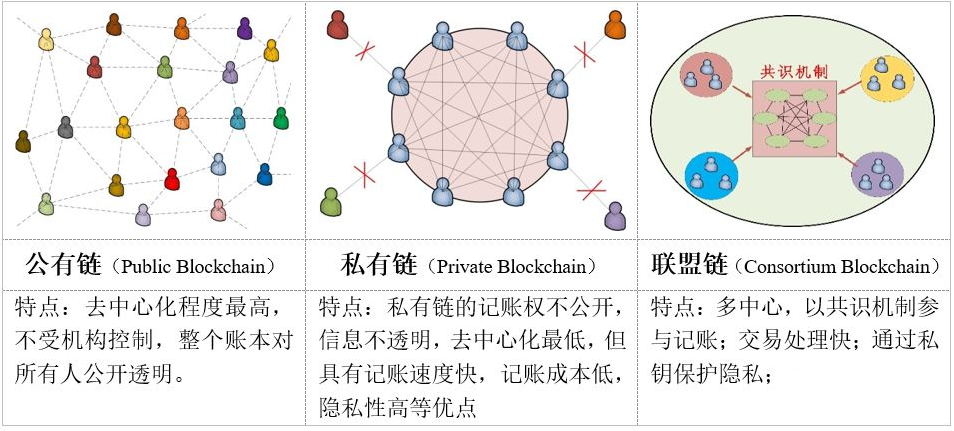


图 5-5 区块链架构的对比

本项目拟搭建一条联盟链，通过智能合约来构建电力数据系统架构，支持电网企业内外业务可信服务的开展。联盟链节点由国网内部各部门以及外部单位构成，采用联盟链的方式以适应电力业务系统数据的实时与隐私性要求。

5.3.4 分发中用到的智能合约

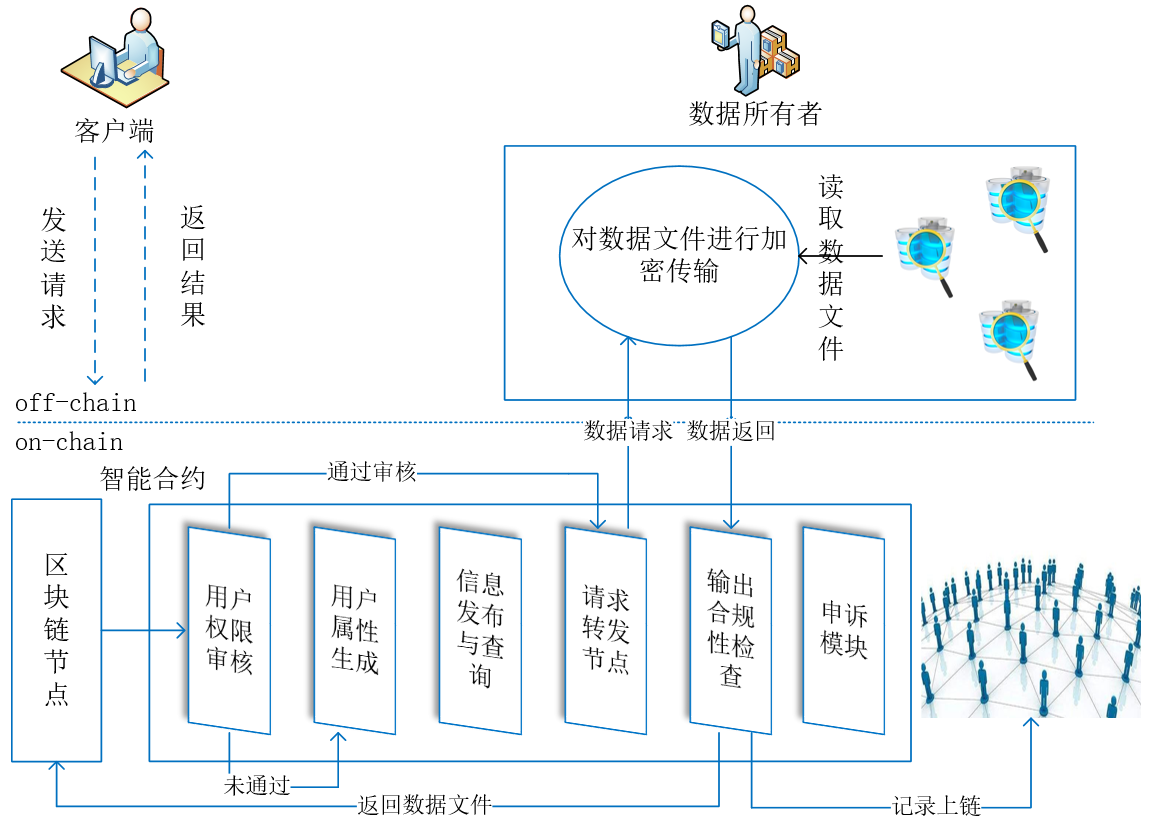


图 5-7 一次区块链数据分发流程图

(1) 用户权限审核：判断申请数据文件的用户是否满足数据所有者的访问策略。

(2) 用户属性生成：如果用户属性不满足访问策略，可以通过购买等方式获取相应属性，以满足访问策略。

(3) 信息发布查询：数据所有者将数据信息以及访问控制策略发布在区块链网络中，供用户调用智能合约查询。

(4) 请求转发：对满足访问策略的用户请求转发给数据所有者，请求数据所有者进行数据转发。

(5) 输出合规检查：检查返回的数据文件是否做了加密处理，是否将此次交易记录上链保存。

(6) 申诉模块：若数据用户解密后发现数据文件与描述信息不符，可以申诉此次交易的不合法性。

5.4 基于区块链的数据流转过程

5.4.1 数据流转过程中的数据信息

大致可分为四种数据信息：

(1) 数据文件，加密后存储在云服务器或本地，如果存储在云服务器中，需要保存文件在云服务器中的地址。

(2) 系统用户信息，包括系统注册信息，以及用户所做的交易记录信息，存储在数据库之中。

(3) 消息摘要：包括文件简介、原始密文哈希值、以及存储于云服务器中的地址等信息，被发布在区块链上，通过智能合约供用户查询和调用。

(4) 记录凭据：包括数据发送方、数据接收方、时间、数据文件id、使用的智能合约等，供溯源追责。

5.4.2 数据分发的一个典型过程

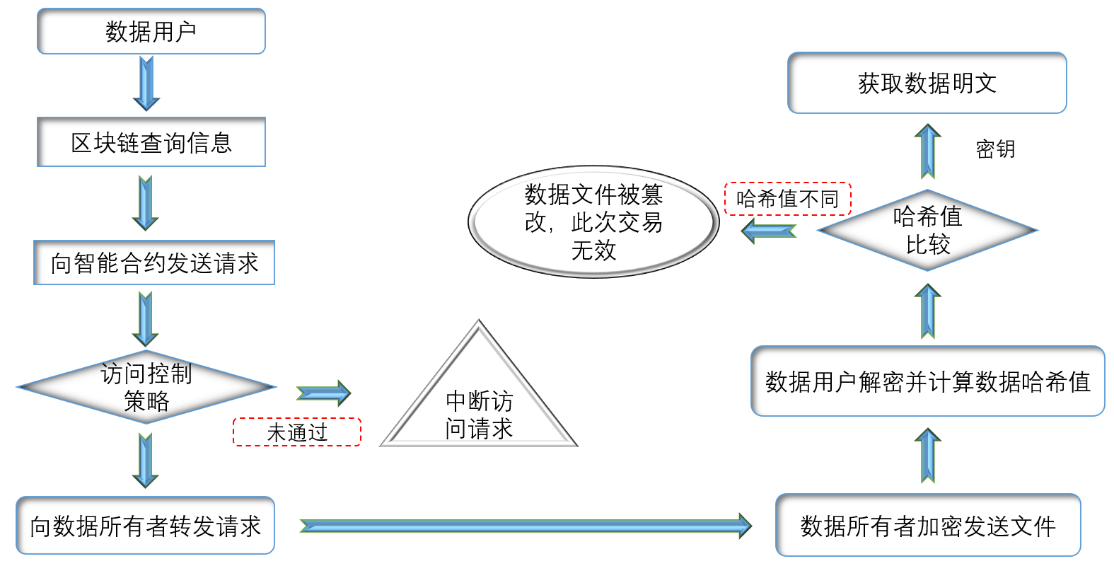


图 5-8 数据分发逻辑框图

(1) 数据中台在系统中广播分发数据的信息，信息包括数据中台在区块链网络中的地址、数据信息所属类型、数据信息简介、原始数据哈希值等。

(2) 数据使用者需要获取数据时，可以在系统中查询相关数据，根据数据信息决定是否向数据中台申请数据，当数据使用者决定申请此数据时，向数据中台发出访问请求，该请求包括数据使用者的地址和数据使用者的公钥。

(3) 如果数据中台认为该用户有权限查看数据，则将经数据使用者公钥加密的数据文件发送给数据使用者，并将此次交易记录上链。

(4) 数据使用者收到加密后的数据，用自己的私钥得到数据文件。

(5) 比较数据文件哈希值与数据中台发布的哈希值是否相等，若是不同，证明数据文件被篡改，需要重新发送。

5.4.3 转发过程中链上记录的信息

主要有以下两种：

(1) 智能合约中记录着发布的数据信息包括原始文件哈希值、数据文件简介、数据所有者、发布时间、数据文件编号，制订的访问控制策略等。

(2) 交易记录中包含发送时间、数据发送者、数据用户、用到的智能合约、数据文件简介、用户权限等。

5.5 本章小结

针对传统的数据分发集中化程度高、权限分配不清晰、分发效率低且溯源条件复杂等状况。本章主要研究了基于区块链的数据分发的相关内容，包括数据分发的几种典型架构及其流程描述和风险分析；电力数据分发所面临的风险和防护点；分析了具体的数据分发过程，包括需要记录在链上的用于溯源的信息内容等。使用联盟链构建区块链服务，构建电力数据分发的去中心化模式,实现电力分发记录上链,电力数据存储以及电力数据安全分享，建立了可追溯，不可篡改，公开透明的共享溯源模型，保证了电力数据的安全传输以及数据的溯源可追踪问题，实现对数据发送方和数据接收方的有效监管。

基于加密的数据安全分发技术研究

6.1 常见加密方式

6.1.1 加密的概念流程

计算机与计算机之间进行通信时，发送方的计算机通常通过加密将明文消息变成密文消息，然后通过网络将加密的密文消息发送到接收方。接收方的计算机对加密的密文消息采用相反的过程，即通过解密还原成明文消息。注：要加密明文消息，发送方进行加密，即采用加密算法。要解密收到的加密消息，接收方要进行解密，即采用解密算法。显然，加密算法和解密算法相同，否则无法通过解密得到原先的消息。其中每个加密与解密过程都有两个方面：加密与解密的算法与密钥。一般来说，加密和解密过程中使用的算法通常是公开的，但加密与解密的所使用的密钥能够保证加密过程的安全性。

根据加解密时是否采用同一个密钥，可以将加密算法划分为对称加密体制和非对称加密体制。

(1) 对称加密体制：

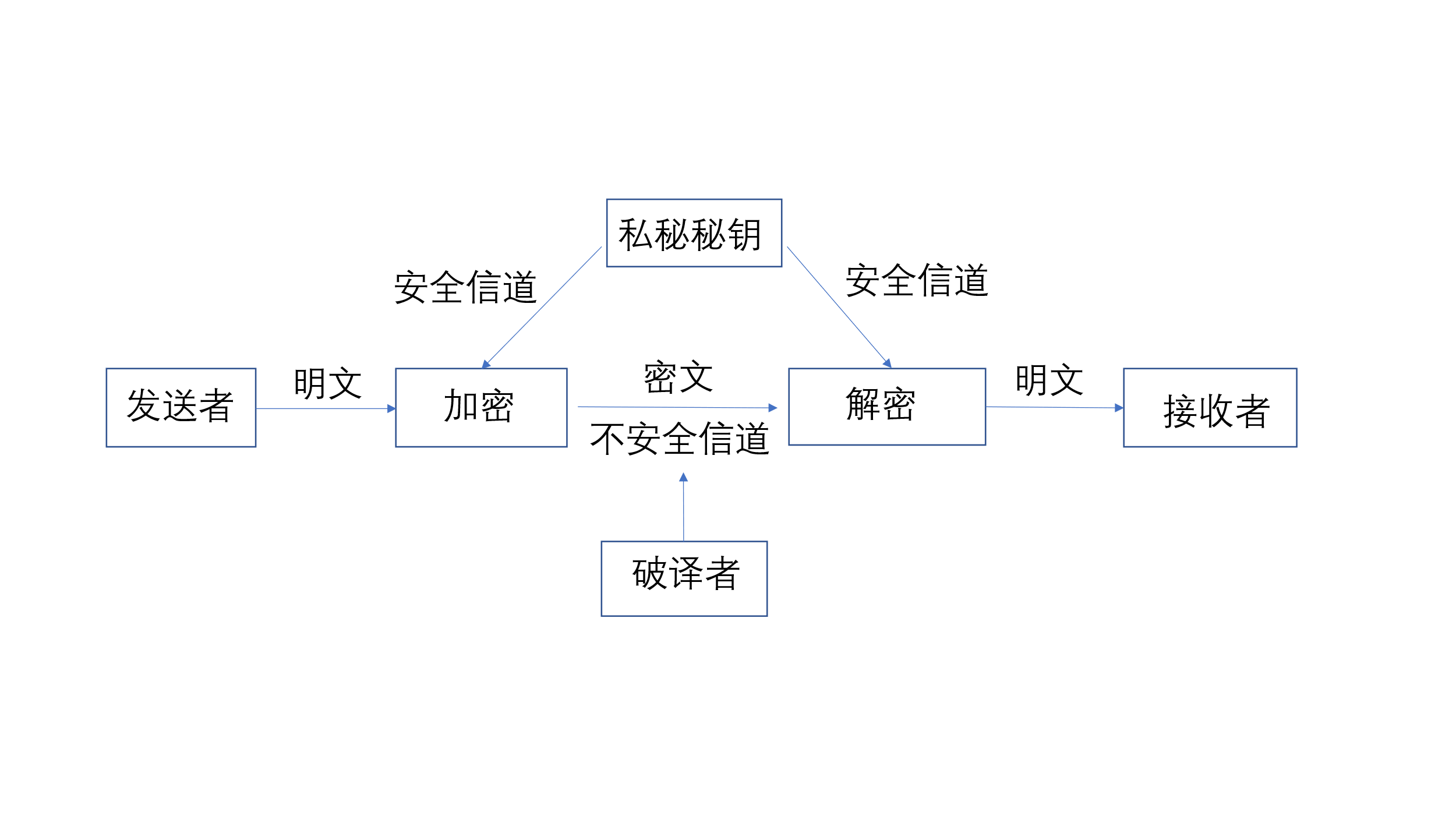


图 6-1 基于对称加密的信道传输

发送者首先通过安全信道和接收者共享相同的密钥k，在数据发送前，将明文经密钥k加密后通过不安全信道发送给接收方，接收方拿到密文后用相同的密钥和解密算法得到明文数据。

(2) 非对称加密体制：

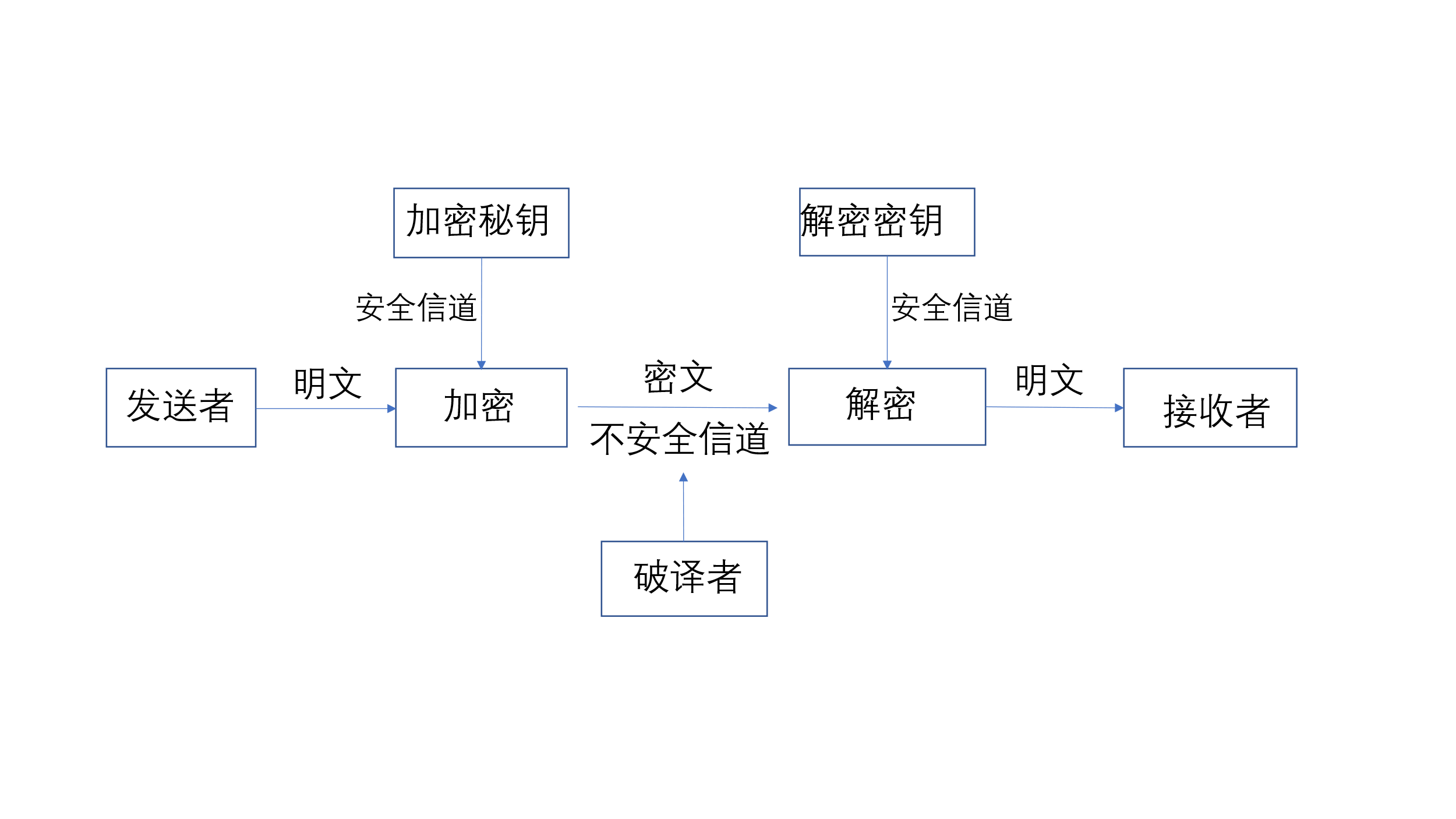


图 6-2 基于非对称加密的信道传输

加密密钥与解密密钥不相同，密钥分为公钥和私钥，发送方将明文数据经接收者公钥加密后由不安全信道传递给接收方，接收方私钥解密得到数据原文。

6.1.2 对称加密体制算法

(1) 定义：所谓对称加密，就是它们在加密时使用的密钥e和解密时使用的密钥d一样(e=d)，我们就将其统称为密钥k。因此加密的安全性不仅取决于加密算法本身，密钥管理的安全性更是重要。因为加密和解密都使用同一个密钥，如何把密钥安全地传递到解密者手上就成了必须要解决的问题。

(2) 工作过程：发送端和接收端首先要共享相同的密钥k(即通信前双方都需要知道对应的密钥)才能进行通信。发送端用共享密钥k对明文p进行加密，得到密文c，并将得到的密文发送给接收端，接收端收到密文后，并用其相同的共享密钥k对密文c进行解密，得出明文p。在传输过程中，即使攻击者截获了密文，在不知道密钥的情况下也无法破解。

(3) 常见算法：SM4、SM1、[DES](https://baike.baidu.com/item/DES)、[3DES](https://baike.baidu.com/item/3DES)、AES、RC2、RC4等。

DES算法：

DES是一个分组加密算法，典型的DES以64位为一组进行数据加密，加密和解密用的是同一个算法。密钥长64位，密钥事实上是56位参与DES运算（第8、16、24、32、40、48、56、64位是校验位），分组后的明文组和56位的密钥按位替代或交换的方法形成密文组。

三重DES：

为了克服DES密钥空间小的缺陷，人们又提出了三重DES这种变形方式。三重DES是为了增加DES的强度，将DES重复3次所得到的一种密码算法，通常缩写为3DES。

AES算法：



图 6-3 AES加密传输

这里以AES-128为例，也就是密钥长度为128位，加密轮数为10轮。加密的第1轮到第9轮的轮函数一样，包括4个操作：字节代换、行位移、列混合和轮密钥加。最后一轮迭代不执行列混合。另外，在第一轮迭代之前，先将明文和原始密钥进行一次异或加密操作。

SM1算法：

对称加密算法，加密强度与AES相当，但是算法不公开，调用该算法时，需要通过加密芯片的接口进行调用。

SM4算法：

基本方案：

2012年3月，国家密码管理局正式公布了包含SM4分组密码算法在内的《祖冲之序列密码算法》等6项密码行业标准。与DES和AES算法类似，SM4算法是一种分组密码算法，其分组长度为128bit，密钥长度也为128bit。加密算法与密钥扩展算法均采用32轮非线性迭代结构，以字（32位）为单位进行加密运算，每一次迭代运算均为一轮变换函数F。SM4算法加/解密算法的结构相同，只是使用的轮密钥相反，其中解密轮密钥是加密轮密钥的逆序。

整体结构：

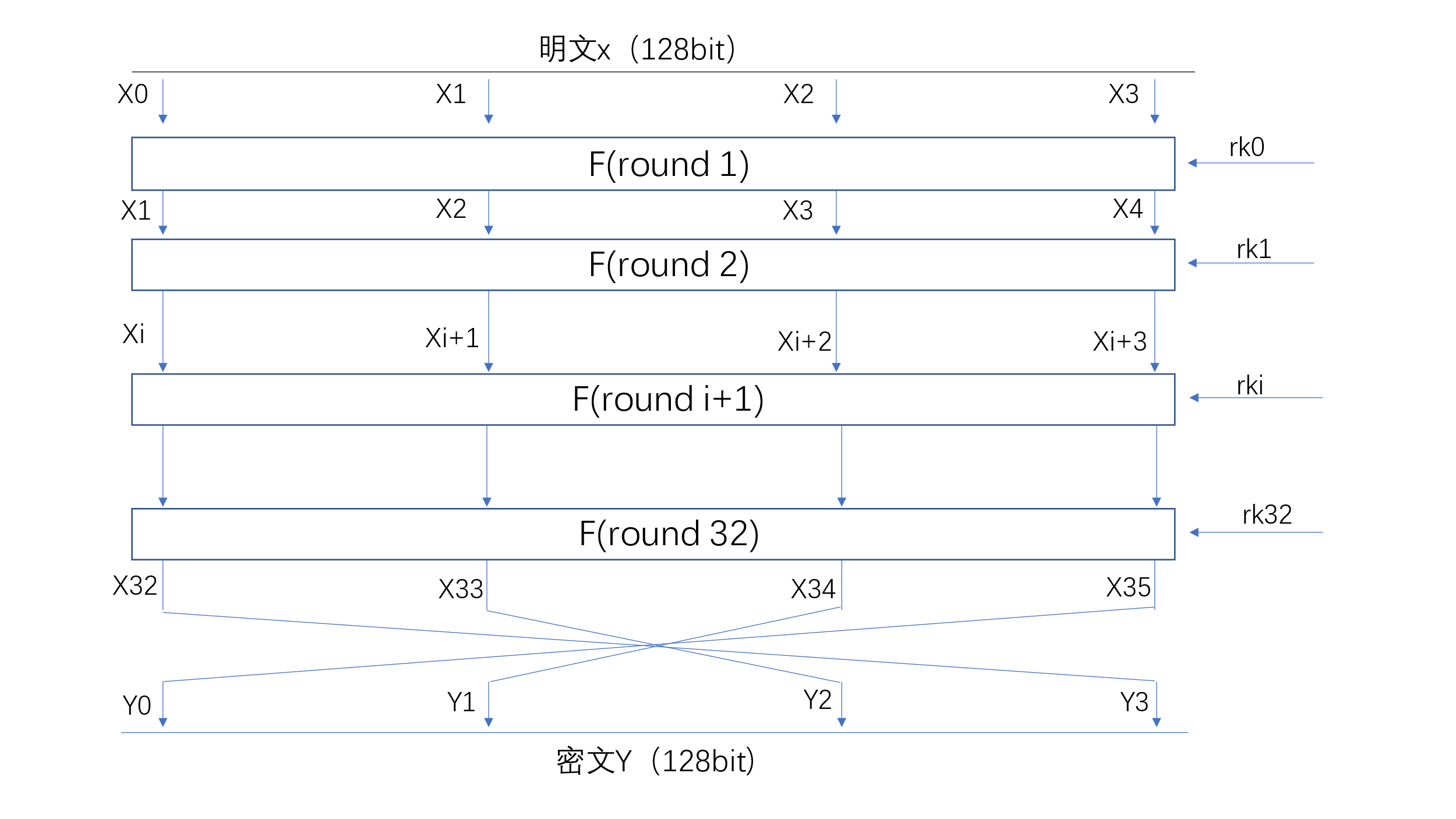


图 6-4 SM4的加密流程

参数产生：

字节由8位2进制数表示，字由32位2进制数表示；

S盒为固定的8bit输入和输出置换；

加密密钥长度为128bit，表示为MK = (MK0,MK1,MK2,MK3)，其中MKi (i=0,1,2,3)为字。轮密钥表示为rki（i=0,1,2.....,31）为字。FK=(FK0,FK1,FK2,FK3)为系统参数，CK=(CK0,CK1,.....,CK31)为固定参数，都为字。

轮函数：

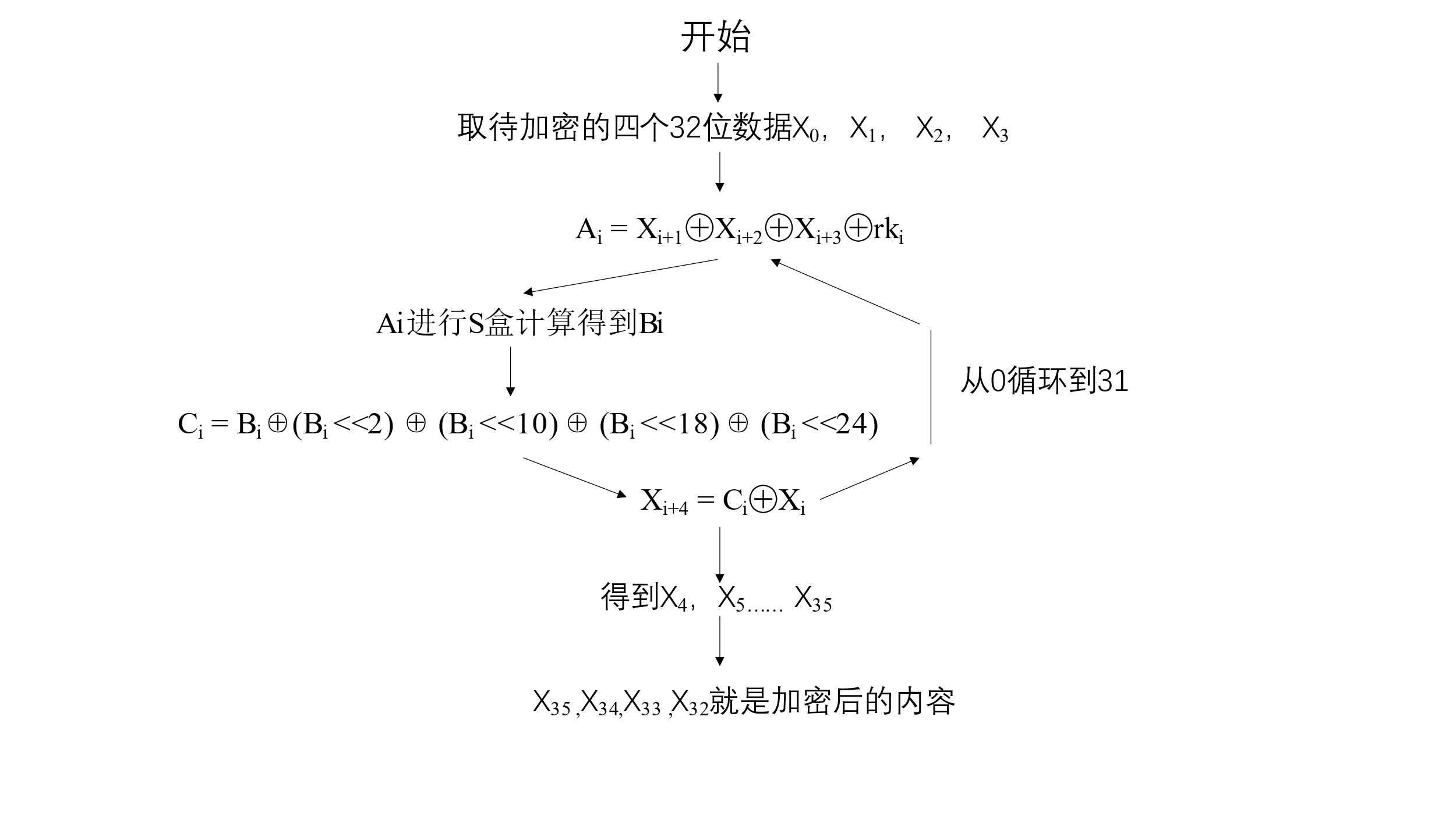


图 6-5 SM4轮函数计算过程

整体的加密函数为：

Xi+4 = F(Xi,Xi+1,Xi+2,Xi+3,rki) = Xi ⊕ T(Xi+1⊕Xi+2⊕Xi+3⊕rki)，其中T为一个合成置换，由非线性变换和线性变换复合而成。非线性变换由4个平行的S盒构成，S盒的数据均采用16进制。线性变换公式如下，其中B为非线性变换得到的字。C = L(B) = B⊕(B<<2) ⊕ (B<<10) ⊕ (B<<18) ⊕ (B<<24)。

密钥扩展：

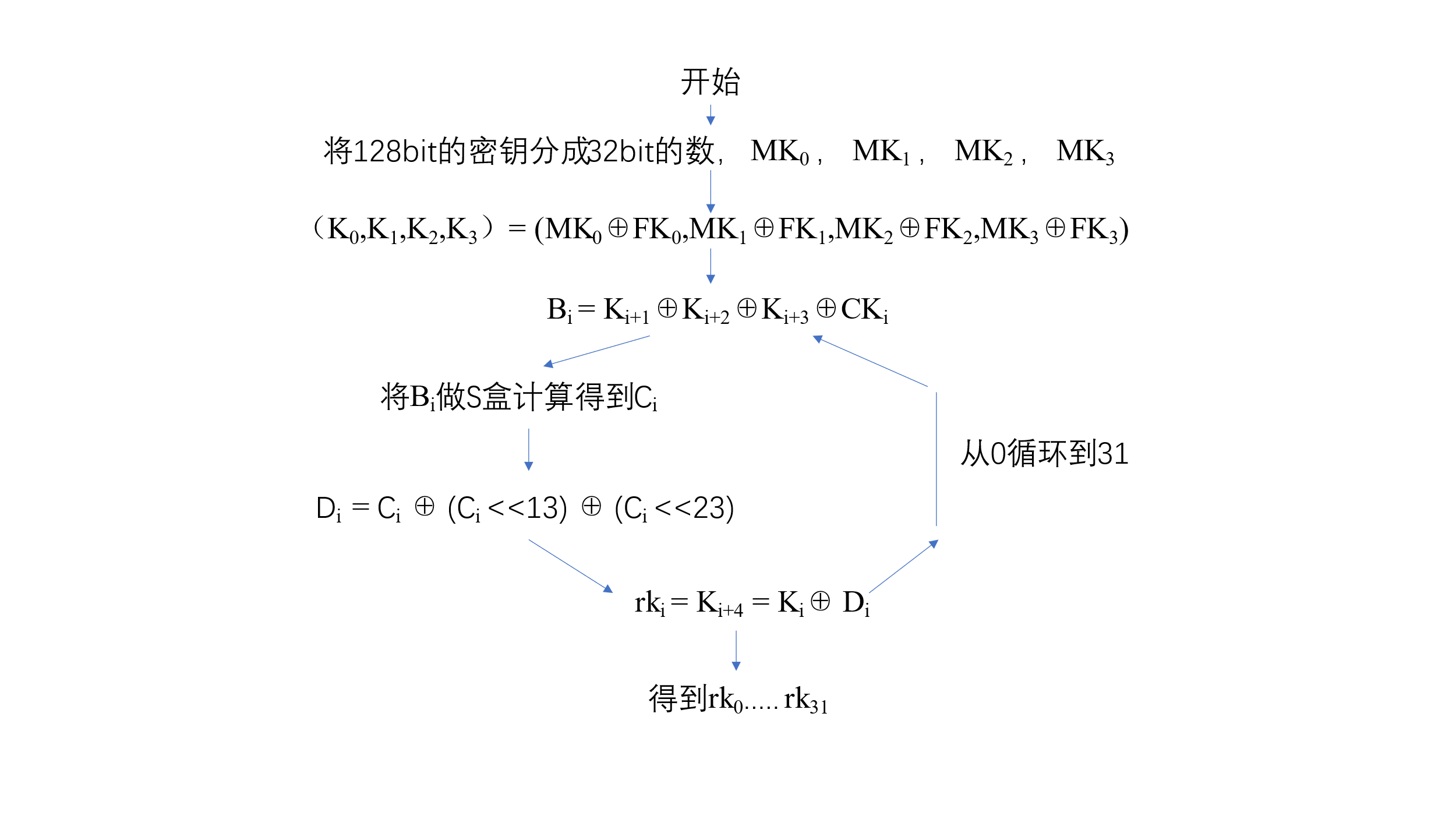


图 6-6 SM4密钥扩展过程

已知加密密钥MK = (MK0,MK1,MK2,MK3)，系统参数FK=(FK0,FK1,FK2,FK3)， 固定参数CK=(CK0,CK1,.....,CK31)。rki为轮密钥，轮密钥由加密密钥生成。

首先，（K0,K1,K2,K3）= (MK0⊕FK0,MK1⊕FK1,MK2⊕FK2,MK3⊕FK3),然后对i=0，1…31：rki = Ki+4 = Ki⊕T1(Ki+1⊕Ki+2⊕Ki+3⊕CKi),该变换与加密中的T变换基本相同，只是将其中的线性变换改为：L1(B) = B ⊕ (B<<13) ⊕ (B<<23),由于系统参数和固定参数是已知的，轮密钥即可求得。

加解密过程：

加密最后一轮变换时，输出为(Y0,Y1,Y2,Y3) = R(X32,X33,X34,X35) = (X35,X34,X33,X32),最后输出是加密的反序，解密时只是将轮密钥的使用顺序逆向进行。

6.1.3 非对称加密体制算法

(1) 定义：所谓非对称加密技术是指加密的密钥e和解密的密钥d是不同的（e!=d），并且加密的密钥e是公开的，叫做公钥，而解密的密钥d是保密的，叫私钥。与[对称加密算法](https://baike.baidu.com/item/%E5%AF%B9%E7%A7%B0%E5%8A%A0%E5%AF%86%E7%AE%97%E6%B3%95/211953)不同，[非对称加密算法](https://baike.baidu.com/item/%E9%9D%9E%E5%AF%B9%E7%A7%B0%E5%8A%A0%E5%AF%86%E7%AE%97%E6%B3%95)需要两个[密钥](https://baike.baidu.com/item/%E5%AF%86%E9%92%A5)：[公开密钥](https://baike.baidu.com/item/%E5%85%AC%E5%BC%80%E5%AF%86%E9%92%A5)（public key）和私有密钥（private key）。[公开密钥](https://baike.baidu.com/item/%E5%85%AC%E5%BC%80%E5%AF%86%E9%92%A5)与私有密钥是一对，如果用公开密钥对数据进行加密，只有用对应的私有密钥才能解密；如果用私有密钥对数据进行加密，那么只有用对应的公开密钥才能解密。因为加密和解密使用的是两个不同的[密钥](https://baike.baidu.com/item/%E5%AF%86%E9%92%A5)，所以这种算法叫做[非对称加密算法](https://baike.baidu.com/item/%E9%9D%9E%E5%AF%B9%E7%A7%B0%E5%8A%A0%E5%AF%86%E7%AE%97%E6%B3%95)。

(2) 工作过程：加密一方找到接收方的公钥e(如何找到呢？大部分的公钥查找工作实际上都是通过数字证书来实现的)，然后用公钥e对明文p进行加密后得到密文c，并将得到的密文c发送给接收方，接收方收到密文后，用自己保留的私钥d进行解密，得到明文p，需要注意的是：用公钥加密的密文，只有拥有私钥的一方才能解密，加密的各方统一使用一个公钥即可。在传输过程中，即使攻击者截获了传输的密文，并得到了接收方的公钥，也无法破解密文，因为只有接收方的私钥才能解密密文。同样，如果接收方要回复加密信息给发送方，那么需要发送方先公布自己的公钥给接收方用于加密，然后用自己的私钥进行解密。

(3) 常见算法：SM2、ECC、EIGamal、RSA等。

RSA算法：

算法原理：根据数论，寻求两个大素数比较简单，而将它们的乘积进行因式分解却极其困难，因此可以将乘积公开作为加密密钥。RSA算法的密钥很长，加密的计算量比较大，安全性较高，但是加密速度比较慢。由于进行的都是大数计算，使得RSA最快的情况也比DES慢上好几倍，无论是软件还是硬件实现。速度一直是RSA的缺陷。采用DES与RSA相结合的应用，使它们的优缺点正好互补，即DES加密速度快，适合加密较长的报文，可用其加密明文；RSA加密速度慢，安全性好，应用于DES 密钥的加密，可解决DES 密钥分配的问题。

ECC算法：

椭圆曲线难题：K=kG，其中K，G为Ep（a,b）上的点，k为小于n的整数，n是点G的阶，给定k和G，计算K容易，但是给定K和G，求k就很难了。ECC 的主要优势是在某些情况下它比其他的方法使用更小的密钥(比如RSA)，提供相当的或更高等级的安全。ECC 的另一个优势是可以定义群之间的双线性映射，基于Weil对或是Tate对；双线性映射已经在密码学中发现了大量的应用，例如基于身份的加密。ECC被广泛认为是在给定密钥长度的情况下，最强大的非对称算法，因此在对带宽要求高的连接中十分有用。

EIGamal算法：

密钥生成：（1）随机选择一个满足安全要求的大素数p，并生成有限域Zp的一个生成元g∈Zp\*；（2）选一个随机数x（1<r<p-1），计算y≡gx(mod p)，则公钥为（y，g，p），私钥为x。ELGamal算法既可以用来加密，也可以用来签名。哪怕使用相同的密钥，对相同的密文进行加密，得到的签名也不相同，能有效防止网络中的重放攻击。ELGamal算法的缺点就是它的计算量特别大，而且密文会成倍的扩张。

SM2算法：

定义：SM2算法由国家密码管理局于2010年12月17日发布，全称为椭圆曲线算法。椭圆曲线并不是椭圆，之所以称为椭圆曲线是因为它们是用三次方程来表示的，并且该方程与计算椭圆周长的方程相似。一般而言，椭圆曲线的三次方程形为：y2+axy+by=x3+cx2+dx+e（其中a,b,c,d和e是满足某些条件的实数，因为方程中的指数最高是3，所以我们称之为三次方程，或者说方程的次数为3）。SM2算法使用的方程为：y2= x3 + ax + b。

算法实现：

1. A选定一条椭圆曲线Ep(a，b),并取曲线上一点作为基点G。

2. A选定一个私钥k，并生成公钥K=kG，k为小于G的阶n的素数。

3. A将Ep(a，b)和K，G发送给B。

4. B收到后将明文编码到Ep(a，b)上一点M，并产生一个随机数r。

5. B计算点C1=M+rK，C2=rG。

6. B将C1，C2传给A。

7. A计算C1-kC2=M+rkG-krG=M。

8. A对M解码得到明文。

攻击者只能得到Ep(a，b)，G，K，C1，C2，没有k就无法得到M。

注意点：

并不是所有椭圆曲线都适合加密，SM2算法采用的方程是一类可以用来加密的椭圆曲线，也是较为简单的一类。这条曲线定义在Fp上，x,y属于0到p-1(p为素数)间的整数，将这条曲线定义为Ep(a，b)。

参数p的选取：p当然越大越安全，但是计算速度会变慢，200位左右会满足一般安全需求。

基点的确定：随着a，b，p确定，曲线定下来了。先随机产生0到p-1间的整数作为基点x坐标，然后根据曲线确定基点y坐标。

私钥的确定：随机选取1到p-1之间的素数作为私钥k。

G的阶n是指满足nG为无穷远点的最小n值。

6.1.4 不同加密算法的比较

(1) 国密与国际算法之间的比较：

对称密码算法国密SM4与国际DES之间的比较：

表 6-1 DES与SM4算法的比较

|  |  |  |
| --- | --- | --- |
|  | DES算法 | SM4算法 |
| 计算基础 | 二进制 | 二进制 |
| 算法结构 | 使用标准的算术和逻辑运算，先替代后置换，不含非线性变换 | 基本轮函数加迭代，含非线性变换 |
| 加解密算法是否相同 | 是 | 是 |
| 计算轮数 | 16轮（3DES为48轮） | 32轮 |
| 分组长度 | 64位 | 128位 |
| 密钥长度 | 64位（3DES为128位） | 128位 |
| 有效密钥长度 | 56位（3DES为112位） | 128位 |
| 实现难度 | 易于实现 | 易于实现 |
| 实现性能 | 软件实现慢、硬件实现快 | 软件实现和硬件实现都快 |
| 安全性 | 较低（3DES较高） | 算法新，还未经过现实检验 |

总结：从算法上看，国产SM4算法在计算过程中增加非线性变换，理论上能大大提高其算法的安全性，并且由专业机构进行了密码分析，民间也对21轮SM4进行了差分密码分析，结论均为安全性较高。

公钥密码算法国密SM2和国际RSA的比较：

表 6-2 SM2与RSA算法之间的比较

|  |  |  |
| --- | --- | --- |
|  | RSA算法 | SM2算法 |
| 计算结构 | 基于特殊的可逆模幂运算 | 基于椭圆曲线 |
| 计算复杂度 | 亚指数级 | 完全指数级 |
| 相同的安全性能下所需的公钥位数 | 较多 | 较少（160位的SM2与1024位的RSA具有相同的安全等级） |
| 密钥生成速度 | 慢 | 较RSA算法快百倍以上 |
| 加密解密速度 | 一般 | 较快 |
| 安全性难度 | 基于分解大整数的难度 | 基于椭圆曲线的离散对数难题 |

哈希算法国际SHA、MD与国密SM3的比较：

表 6-3 SHA、MD与SM3算法之间的比较

|  |  |  |
| --- | --- | --- |
|  | SHA、MD | SM3 |
| 算法结构不同 | 消息填充方面，几个hash算法基本相同，但SHA-256压缩函数每一轮使用1个消息字 | SM3算法的压缩函数和SHA-256具有相似的结构。但SM3设计更加复杂，比如压缩函数每一轮都使用2个消息字 |
| 安全性不同 | 现今MD5和SHA-1算法不再是安全的算法 | SM3是在SHA-256基础上改进的算法，消息分组长度为512位，摘要值长度为256位，安全性相对较高 |
| 效率不同 | 效率较低 | 效率较高 |

(2) 对称与非对称算法之间的比较：

表 6-4 对称与非对称算法之间的比较

|  |  |  |
| --- | --- | --- |
|  | 优点 | 缺点 |
| 对称加密算法 | 加解密的速度快 | 1.共享密钥的分发困难  2.若一方要跟n方通信，则需要维护n对密钥 |
| 非对称加密算法 | 1. 不存在密钥分发的问题，解码方可以自己生成密钥对  2.多个加密方都可以使用一个已知的公钥进行加密，不需要n对密钥 | 加解密的速度没有对称加密快 |

6.2 常见的数字签名算法

6.2.1 数字签名的相关概念

(1) 定义：

数字签名（又称公钥数字签名）是只有信息的发送者才能产生的别人无法伪造的一段数字串，这段数字串同时也是对信息的发送者发送信息真实性的一个有效证明。它是一种类似写在纸上的普通的物理签名，但是使用了公钥加密领域的技术来实现，用于鉴别数字信息的方法。一套数字签名通常定义两种互补的运算，一个用于签名，另一个用于验证。数字签名是非对称密钥加密技术与数字摘要技术的应用。

(2) 原理：

简单地说，所谓数字签名就是附加在[数据单元](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E5%8D%95%E5%85%83)上的一些数据，或是对数据单元所作的[密码](https://baike.baidu.com/item/%E5%AF%86%E7%A0%81)变换。这种数据或变换允许[数据单元](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E5%8D%95%E5%85%83)的接收者用以确认数据单元的来源和数据单元的完整性并保护数据，防止伪造。基于[公钥](https://baike.baidu.com/item/%E5%85%AC%E9%92%A5)[密码体制](https://baike.baidu.com/item/%E5%AF%86%E7%A0%81%E4%BD%93%E5%88%B6)和对称密码体制都可以获得数字签名，主要是基于公钥密码体制的数字签名。包括普通数字签名和特殊数字签名。普通数字签名[算法](https://baike.baidu.com/item/%E7%AE%97%E6%B3%95)有RSA、ElGamal、Fiat-Shamir、Guillou- Quisquarter、Schnorr、Ong-Schnorr-Shamir[数字签名算法](https://baike.baidu.com/item/%E6%95%B0%E5%AD%97%E7%AD%BE%E5%90%8D%E7%AE%97%E6%B3%95/12724298)、DES/DSA，椭圆曲线数字签名算法和[有限自动机](https://baike.baidu.com/item/%E6%9C%89%E9%99%90%E8%87%AA%E5%8A%A8%E6%9C%BA/8700995)数字签名算法等。特殊数字签名有[盲签名](https://baike.baidu.com/item/%E7%9B%B2%E7%AD%BE%E5%90%8D)、代理签名、[群签名](https://baike.baidu.com/item/%E7%BE%A4%E7%AD%BE%E5%90%8D)、不可否认签名、公平盲签名、门限签名、具有消息恢复功能的签名等，它与具体应用环境密切相关。显然，数字签名的应用涉及到法律问题，[美国联邦政府](https://baike.baidu.com/item/%E7%BE%8E%E5%9B%BD%E8%81%94%E9%82%A6%E6%94%BF%E5%BA%9C)基于有限域上的离散对数问题制定了自己的[数字签名标准](https://baike.baidu.com/item/%E6%95%B0%E5%AD%97%E7%AD%BE%E5%90%8D%E6%A0%87%E5%87%86)([DSS](https://baike.baidu.com/item/DSS))。

(3) 特点：

网络的安全，主要是[网络信息安全](https://baike.baidu.com/item/%E7%BD%91%E7%BB%9C%E4%BF%A1%E6%81%AF%E5%AE%89%E5%85%A8/10031875)，需要采取相应的[安全技术措施](https://baike.baidu.com/item/%E5%AE%89%E5%85%A8%E6%8A%80%E6%9C%AF%E6%8E%AA%E6%96%BD/9490902)，提供适合的[安全服务](https://baike.baidu.com/item/%E5%AE%89%E5%85%A8%E6%9C%8D%E5%8A%A1/1176923)。数字签名机制作为保障[网络信息安全](https://baike.baidu.com/item/%E7%BD%91%E7%BB%9C%E4%BF%A1%E6%81%AF%E5%AE%89%E5%85%A8/10031875)的手段之一，可以解决伪造、抵赖、冒充和篡改问题。数字签名的目的之一就是在网络环境中代替传统的手工签字与印章，发挥以下重要作用：

防冒充(伪造)：私有密钥只有签名者自己知道，所以其他人不可能构造出伪造的签名。

可[鉴别](https://baike.baidu.com/item/%E9%89%B4%E5%88%AB/10755551)身份：由于传统的手工签名一般是双方直接见面的，身份自可一清二楚。在网络环境中，接收方必须能够[鉴别](https://baike.baidu.com/item/%E9%89%B4%E5%88%AB/10755551)发送方所宣称的身份。

防[篡改](https://baike.baidu.com/item/%E7%AF%A1%E6%94%B9/9659251)(防破坏信息的完整性)：对于传统的手工签字，假如要签署一份200页的合同，是仅仅在合同末尾签名呢？还是对每一页都签名？如果仅在合同末尾签名，对方会不会偷换其中的几页? 而对于数字签名，签名与原有文件已经形成了一个混合的整体数据，不可能被篡改，从而保证了数据的[完整性](https://baike.baidu.com/item/%E5%AE%8C%E6%95%B4%E6%80%A7/949221)。

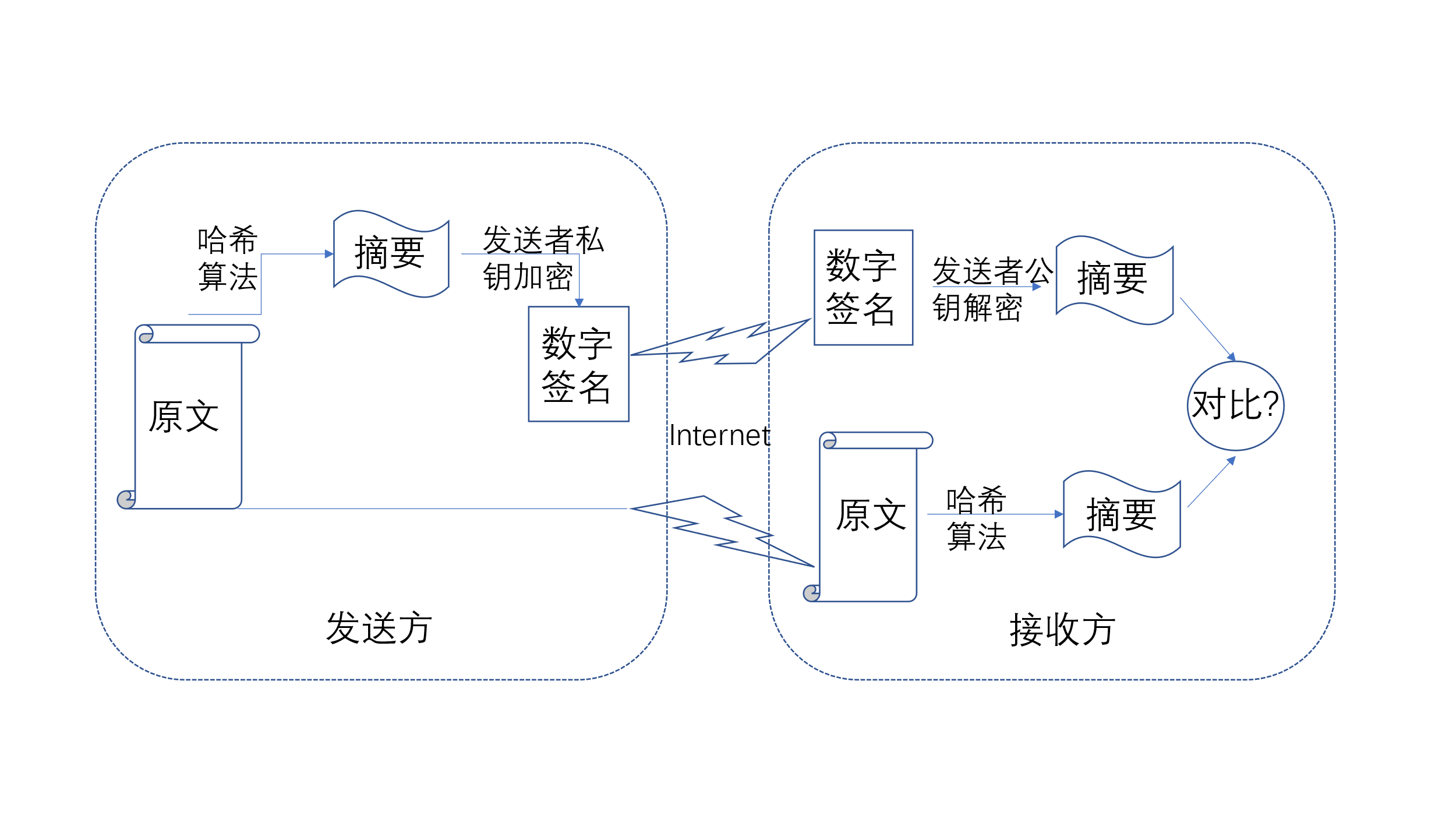
防[重放](https://baike.baidu.com/item/%E9%87%8D%E6%94%BE/1344212)：如在日常生活中，A向B借了钱，同时写了一张借条给B，当A还钱的候，肯定要向B索回他写的借条撕毁，不然，恐怕他会再次用借条要求A还钱。在数字签名中，如果采用了对签名报文添加[流水号](https://baike.baidu.com/item/%E6%B5%81%E6%B0%B4%E5%8F%B7/5623476)、[时间戳](https://baike.baidu.com/item/%E6%97%B6%E9%97%B4%E6%88%B3/6439235)等技术，可以防止[重放](https://baike.baidu.com/item/%E9%87%8D%E6%94%BE/1344212)攻击。

防[抵赖](https://baike.baidu.com/item/%E6%8A%B5%E8%B5%96/9852743)：如前所述，数字签名可以鉴别身份，不可能冒充伪造。那么，只要保护好签名的报文，就好似保存好了手工签署的合同文本，也就是保留了证据，签名者就无法抵赖。那如果接收者确已收到对方的签名报文，却抵赖没有收到呢?要预防接收者的[抵赖](https://baike.baidu.com/item/%E6%8A%B5%E8%B5%96/9852743)。在数字签名体制中，要求接收者返回一个自己签名的表示收到的报文，给对方或者第三方。如此操作，双方均不可抵赖。

机密性(保密性)：有了机密性保证，截收攻击也就失效了。手工签字的文件(如同文本)是不具备保密性的，文件一旦丢失，其中的信息就极可能泄露。数字签名可以加密要签名的消息，当然，如果签名的报文不要求机密性，也可以不用加密。

6.2.2 数字签名的过程及应用

(1) 数字签名过程：

图 6-7 数字签名一般过程

1） 使用消息摘要算法将发送数据加密生成数字摘要。

2） 发送方用自己的私钥对摘要再加密，形成数字签名。

3） 将原文和加密的摘要同时传给对方。

4） 接收方用发送方的公钥对摘要解密，同时对收到的数据用消息摘要算法产生同一摘要。

5） 将解密后的摘要和收到的数据在接收方重新加密产生的摘要相互对比，如果两者一致，则说明在传送过程中信息没有破坏和篡改。否则，则说明信息已经失去安全性和保密性。

(2) 数字签名的作用：

数字签名有两种功效：一是能确定消息确实是由发送方签名并发出来的，因为别人假冒不了发送方的签名。二是数字签名能确定消息的完整性。因为数字签名的特点是它代表了文件的特征，文件如果发生改变，数字摘要的值也将发生变化。不同的文件将得到不同的数字摘要。一次数字签名涉及到一个哈希函数、接收者的[公钥](https://baike.baidu.com/item/%E5%85%AC%E9%92%A5)、发送方的[私钥](https://baike.baidu.com/item/%E7%A7%81%E9%92%A5)，用于保证信息传输的完整性、发送者的[身份认证](https://baike.baidu.com/item/%E8%BA%AB%E4%BB%BD%E8%AE%A4%E8%AF%81)、防止交易中的抵赖发生。

(3) 数字签名的应用：

金融业：以P2P网络借贷业务为例，当借款方通过P2P平台确认投资后，接入电子合同与数字签名应用的P2P平台可生成投资协议，经投资方与借款方确认后实现快速签署。

政府：政府部门可通过将办公系统与电子合同接口集成，实现文件高流转率、高到达率。例如将数字签名技术应用于通知发布，发送人可通过系统在线批量发送文件到指定接收人，实现文件快速流转与签字回传。

旅游业：旅行社可直接应用电子合同签署平台收发合同，或将电子合同应用接入在线旅游平台，游客与旅行社可通过在线平台快速订立出行协议。

医疗保健：使用数字签名的电子病历不仅将医生从病历书写工作中解放出来，而且，经可靠的数字签名技术签署的电子病历，具备书写内容不可篡改、签署身份可识别等特性。此外，电子病历查询快捷，为医生准确实施救治提供可靠依据。

在线教育：在线教育网站可通过接入契约锁电子合同服务接口的方式实现课程在线选择、购买、签署协议等环节。学员（或学生家长）可直接在教育网站同平台签订、购买课程。

大宗商品电商：应用电子合同与数字签名服务的电商平台可实现买卖双方安全、快捷的交易。买卖双方经电商平台达成交易意向后，平台将根据交易信息生成交易协议，经双方确认无误后签署。经销商确认收款后通过物流即可将商品发送给买方。

汽车行业：汽车产业链中最常见的环节主要有零部件供应商、整车生产制造商、汽车经销商以及终端车主。通过将电子合同与数字签名应用接入汽车线上销售平台中，消费者可直接通过平台线上预定并签署试驾协议。线下门店试驾后，经销商和车主之间也可通过电子合同应用快速完成购车手续办理。

人力资源：将电子合同签署服务部署到人事系统中是企业普遍选用的应用方式。HR可在接入电子合同与数字签名应用的EHR系统中发起人事合同签署流程。通过选定相应的合同模板或创建新的合同并发送给签署人签署。

6.2.3 数字签名的常见哈希算法

(1) SM3国密算法：

定义：SM3密码杂凑（哈希、散列）算法是一种摘要算法，是国家密码局认定的国产商用密码算法。此算法适用于商用密码应用中的数字签名和验证，消息认证码的生成与验证以及随机数的生成，可满足多种密码应用的安全需求，在SM2，SM9标准中广泛使用。此算法对输入长度小于2的64次方的比特消息，进行填充和迭代压缩处理，生成长度为256比特的杂凑值，中间使用了异或，模，模加，移位，与，或，非运算，算法由填充，迭代，消息扩展和压缩函数所构成。

算法简介：

1) 常数与函数

初始值：IV = 7380166f 4914b2b9 172442d7 da8a0600 a96f30bc 163138aa e38dee4d b0fb0e4e。

常量：Tj = { 79cc4519 0≤j≤15；7a879d8a 16≤j≤63}。

布尔函数：FFj(X,Y,Z) = {X^Y^Z  0≤j≤15；(X & Y)|(X&Z)|(Y&Z) 16≤j≤63}，GGj(X,Y,Z) = {X^Y^Z  0≤j≤15；（X & Y）|（~X&Z）16≤j≤63}，X、Y、Z为字（32bit）。

置换函数：P0（X）= X^(X<<<9)^(X<<<17)，P1（X）= X^(X<<<15)^(X<<<23)，X 为字。

2) 算法流程

填充过程：设消息m的长度为l 比特，首先将比特"1"添加到消息末尾，再加k个“0”，k是满足l+1+k =448 mod 512的最小非负整数。然后再添加一个64比特串，该串是l的二进制表示，填充后的消息m'长度为512的整数倍。

迭代压缩：m'按照512bit进行分组：m'=B(0)B(1)...B(n-1)，n=(l+k+65)/512。FOR i=0 to n-1，V(i+1) = CF(V(i),B(i))，ENDFOR。CF为压缩函数，V(0)为初始值IV，迭代压缩的结果为V(n)。

迭代压缩中的消息扩展：

a) 消息分组B(i)划分为16个字W0，W1...W15；

b) FOR j=16 to 67，Wj=P1(Wj-16 ^ Wj-9 ^ (Wj-3 <<<15)^(Wj-13<<<7)^Wj-6，ENDFOR；

c) FOR j=0 to 63，Wj'=Wj^Wj+4，ENDFOR；

d) 消息分组B(i)最终扩展生成132个字W0,W1…W67,W0'…W63'；

迭代压缩中的压缩函数：A/B/C/D/E/F/G/H为字寄存器，SS1/SS2/TT1/TT2为中间变量，压缩函数V(i+1) =CF(V(i),B(i))计算过程如下：

ABCDEFGH = V(i)；FOR j=0 to 63：

SS1 = ((A<<<12)+E+(Tj<<<j))<<<7；SS2 = SS1^(A<<<12)；

TT1 = FFj(A,B,C)+D+SS2+Wj'；TT2 = GGj(E,F,G)+H+SS1+Wj；

D = C；C = B<<<9；B = A；A = TT1；

H = G；G = F<<<19；F = E；E = P0(TT2)；

ENDFOR，V(i+1) = ABCDEFGH^V(i)。

运算过程中，字按照大端格式存储。

(2) MD2：1989年，著名的非对称算法RSA发明人之一麻省理工学院教授罗纳德·李维斯特教授开发了MD2算法。这个算法首先对信息进行数据补位，使信息的字节长度是16的倍数；再以一个16位的检验和作为补充信息追加到原信息的末尾；最后根据这个新产生的信息计算出一个128位的散列值，MD2算法由此诞生。

(3) MD4：1990年，罗纳德·李维斯特教授开发出较之MD2算法有着更高安全性的MD4算法。在这个算法中，我们仍需对信息进行数据补位。不同的是，这种补位使其信息的字节长度加上448个字节后成为512的倍数（信息字节长度mod 512 = 448）。此外，关于MD4算的处理和MD2算法有很大的差别，但最终仍旧会获得一个长度为128位的散列值。MD4算法对后续消息摘要算法起到了推动作用，许多比较有名的消息摘要算法都是在MD4算法的基础上发展而来的，例如MD5、SHA-1、RIPE-MD和HAVAL算法等。

(4) MD5：1991年，继MD4算法后，罗纳德.李维斯特教授开发了MD5算法，最终将MD系列算法推向了成熟。MD5算法经MD2和MD4算法发展而来，算法复杂程度和安全强度也大大提高， MD5算法是输入任意长度字符，输出固定长度128位的算法。经过程序流程，生成四个32位数据，最后联合起来成为一个128位Hash值，主要方式是通过求余、取余、调整长度、与链接变量进行循环运算进而得出结果。但这些MD算法的最终结果都是产生一个128位长度的信息摘要，这也是MD系列算法的主要特点。从安全性上说：MD5 > MD4 > MD2，MD5算法安全性最高。

(5) SHA-1：由NIST设计为同DSA一起使用的，SHA-1设计时基于和MD4相同原理，并且模仿了该算法，SHA-1抗穷举(brute-force)性更好，它产出160位的Hash值，对于非线性运算、移位和加法运算也与MD5类似。SHA-1也应用于包括TLS和SSL、PGP、SSH、S/MIME和IPSEC等多种协议中，曾被视为是MD5的后继者。SHA-1的如今已经明确不具备安全性可言了。

(6) SHA-2：SHA-224、SHA-256、SHA-384和SHA-512并称为SHA-2，发布于2001年，目前比较广泛应用的SSL数字证书和代码签名证书签名算法均采用SHA-256算法，相较于SHA-1算法而言，至今SHA-2算法还未被破解，从某种意义上SHA-2延用了SHA-1算法，所以目前是安全的。目前顶级CA和Google、苹果等公司都采用基于SHA-256算法作为SSL证书和代码签名证书的主流签名算法。

(7) SHA-3：SHA-3算法正式发布于2015年，SHA-3并不是要取代SHA-2，因为SHA-2目前并没有出现明显的弱点。由于对MD5、SHA-0和SHA-1出现成功的破解，NIST感觉需要一个与之前算法不同的，可替换的加密Hash算法，也就是现在的SHA-3。SHA-3选用keccak算法为标准，以太坊便是采用keccak算法。Keccak采用了创新的“海绵引擎”散列消息文本。它设计简单，方便硬件实现，具有广泛的安全边际。至目前为止，第三方密码分析已经显示出Keccak没有严重的弱点。SHA-3只在软件中的运行速度较慢。但在硬件中，其速度优于SHA-1和SHA-2。目前，越来越多的加密程序是由硬件组件负责处理，并且预计未来会更为广泛。

6.3 基于加密的数据流转图

6.3.1 分发过程介绍

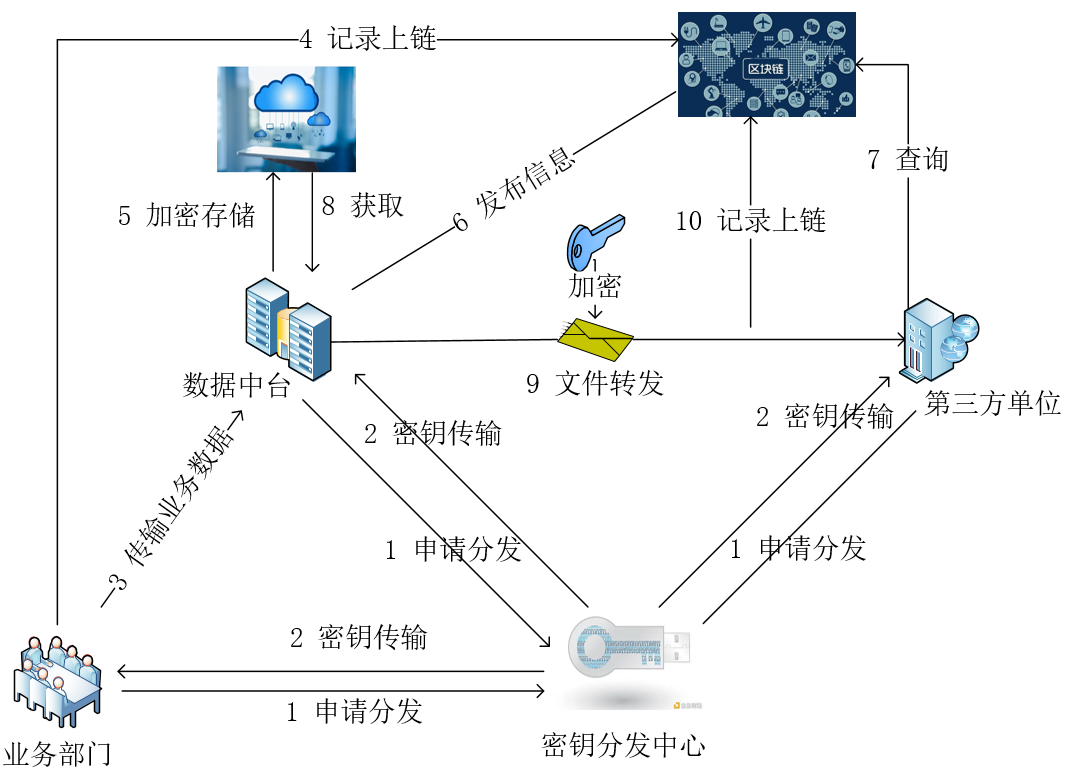


图 6-8 基于加密的区块链数据流转图

(1) 业务部门、数据中台和第三方单位申请密钥，密钥中心对申请分发的密钥进行记录。

(2) 业务部门向数据中台传输业务数据，并将数据传输记录在区块链上。

(3) 数据中台加密存储数据，并向区块链网络中发布信息。第三方单位查询到相应的信息后，发送请求。

(4) 智能合约验证用户权限满足访问策略后，转发请求，数据中台加密传输数据文件密文，并将此次交易记录上链。

(5) 第三方单位用自己的私钥解密文件，将文件哈希后与区块链中发布的文件哈希值比较，若哈希值相同，则表示数据文件未得到篡改。用第三方单位私钥解密得到数据文件密钥，从而获取数据明文。

6.3.2 数据流转过程中的加密应用

(1) 数据文件尽量不应明文存储，应加密存储在本地或者云服务器中。

(2) 数据转发过程中，应对数据文件用接收者的公钥进行加密处理，在接收方拿到加密的数据文件后，用私钥解密并将原文哈希后比较摘要是否相等。

基于国密SM2的加密应用：

第一步：数据用户查询到想要的数据文件之后，向智能合约发送请求，智能合约判断用户满足其访问权限之后，将用户请求转发给数据所有者，请求中包含（p,a,b,G,B）等公钥信息，其中B = kG，B为公钥，k为私钥。

第二步：数据所有者将数据明文m编码成椭圆曲线上一个点Pm，然后选择随机数r，计算密文Cm = {rG，Pm + rB}。

第三步：数据用户收到密文后，用私钥解密，具体来说，计算Pm + rB – k(rG) = Pm ，因为rB = rkG，得到编码后的数据明文Pm，解码即可获得明文。

6.3.3 数据流转过程中的签名应用

在数据用户将数据文件的简介信息发布到区块链网络中时，需要将原始数据哈希值发布出去，让数据用户在接收到数据文件后对该数据文件进行完整性检查。

基于国密SM3的哈希算法应用：

假设用户要传输的数据明文为m，SM3就是给数据加一个固定长度的指纹，这个指纹大小就是256bit。

第一步：填充，使填充后的数据长度是512bit的整数倍，具体来说，先在数据m的尾巴最后加一个1，然后把原始数据m的长度用64bit表示，放在最后面；再看看现在的数据的长度值距离512的整数倍还差多少个比特，差多少个就填多少个0。添加位置在之前添加的1和64bit长度信息之间。

第二步：分组，把填充后的信息按照512bit一个组进行分组，如果分成了n组，就是b(0)、b(1)…b(n)。

第三步：迭代压缩得到最后的杂凑值，IV(n) = CF(IV(n-1),b(n-1))，如果消息分为N组，那么256bit的IV(n)就是最后得到的杂凑值。在这个过程中，每个b(i)要先被扩展成132个字大小才能用于迭代压缩，64轮压缩过后生成数据文件的哈希值IV(n),该哈希值作为数据文件消息简介的一部分发布到区块链网络中，供数据用户得到文件后的完整性校验。

6.4 基于明文和密文的两种数据安全分发方案

6.4.1 两种方案流程

(1) 数据文件明文存储：

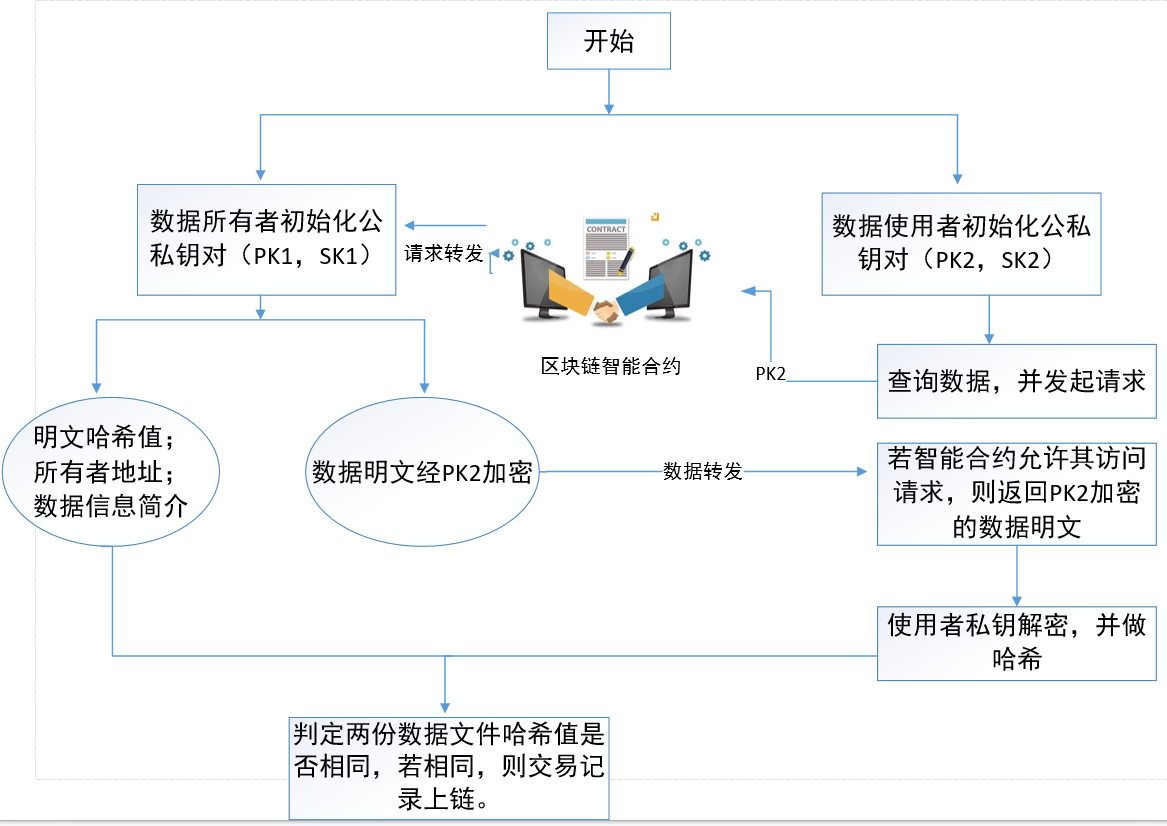


图 6-9 明文存储转发过程

数据中台将数据明文保存在本地或云服务器中，并将数据明文信息简介发布到区块链网络中，其中包括数据文件的相关内容信息，摘要，访问控制策略。

数据用户若想访问数据文件，在满足访问控制策略后，智能合约将数据用户的公钥发送给数据所有者。

数据中台将经过公钥加密的数据文件发送给数据用户。

数据用户用自己的私钥解密文件，将文件哈希后与区块链中发布的文件摘要进行比较。

若哈希值相同，则表示数据文件未得到篡改。

(2) 数据文件密文存储：

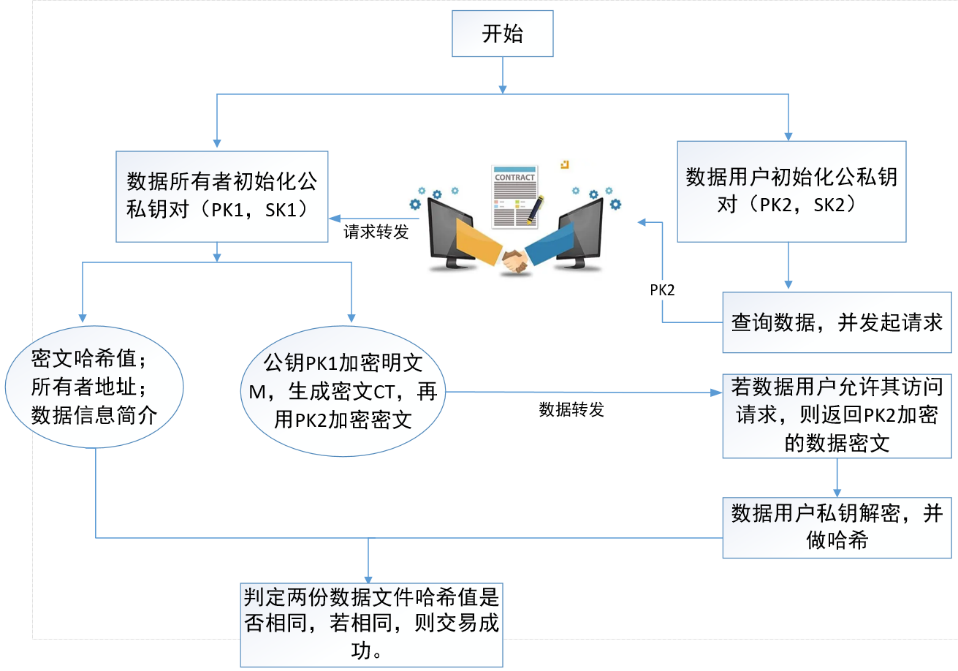


图 6-10 密文存储转发过程

数据中台用自己的密钥对数据明文加密，保存在本地或云服务器中，并将数据明文信息简介发布到区块链网络中，其中包括数据文件的相关内容信息，密文摘要，访问控制策略。

数据用户查询到想要的数据文件，智能合约判断满足相应的访问控制策略后，将数据用户的公钥转发给数据中台。

数据所有者将数据密文和密文对应的解密密钥用数据用户公钥加密后发送给数据用户。

数据用户用自己的私钥解密数据密文对应的密钥和密文本身，将数据密文哈希后与区块链中发布的文件哈希值比较。

若哈希值相同，则表示数据文件未得到篡改。

用数据文件对应的密钥解密文件得到明文数据。

6.4.2 两种方案的不同点

不同点：

(1) 在云服务器或本地存储数据文件时，明文存储不加密，密文存储需要先用数据所有者公钥加密。

(2) 传输过程中，明文存储只需传递数据用户公钥加密后的密文，而密文存储时除了原始密文外还需将解密数据文件的私钥一起加密传输。

(3) 解密过程中，明文存储需要比较解密后明文的哈希值是否相同，密文存储则比较解密后原始密文的哈希值是否相同，且要得到数据明文需要用接收到的私钥解密。

6.5 基于区块链的数据溯源方案

6.5.1 常见溯源方案

对数据进行溯源追踪，是用户对委托到云存储的数据安全保护的重要需求。在数据外包到不完全可信的云服务提供商（CSP）后，用户对数据脱离自己的控制的担忧，自然引发了对云上数据进行溯源追踪的需求。

首先提出的方法就是基于日志记录的溯源追踪。利用CSP提供的系统日志服务，从日志中提取数据操作事件，实现对数据的追踪。这一方法本质上是服务器的系统日志分析，目标是提取数据操作记录，且被Amazon等当前大多数主流云服务提供商普遍支持。但这一方法存在两大问题：一是系统日志记录条目繁多、数据量巨大，而其中的数据操作记录较少，会造成事件提取操作的效率低下；二是系统日志的高度结构化，易被篡改伪造，作为溯源追踪的证据来说，存在可靠性的问题；如果生成系统日志的CSP不完全可信，那么日志甚至存在真实性的问题。

为了使用户能够像使用本地设备一样使用云存储，发展云存储的公开溯源至关重要，这使得用户可以求助于第三方溯源者（TPA）对数据进行验证而无需担心，但同时溯源过程也不应该为用户的数据隐私带来漏洞，现有的研究基于此提出了一个支持隐私保护的公开溯源安全云存储模型，并且对TPA进行了扩展使其能够同时高效的为多个用户执行溯源。利用环签名计算溯源所需的验证信息，共享数据中每个数据块上签名者的身份信息作为隐私对第三方溯源者不可见，第三方溯源能在不检索整个文件的情况下验证数据的完整性。针对云存储中多个数据中心的多台服务器上保存的副本可能不一致的情况，研究者们提出了一种有效的多副本可证明数据持有（PDP）方案，该方案中利用完全同态加密（FH）生成数据块副本，数据拥有者会先对数据分块加密后再利用同态加密生成块标签，并且该方案支持数据拥有者对数据块的动态更新操作和第三方溯源。

在大多数研究中，为了使溯源结果公平公正，使用第三方溯源是一种可行的选择。但同时也带来了问题，如何找到一个用户和云存储提供方都信任的第三方机构是一个挑战，因为在现实中完全可信的第三方几乎不存在。但是随着2009年比特币首次亮相，其底层的区块链技术引起了学术界的广泛关注。区块链技术去中心化的分布式架构和去信任化的运行机制使得建立一个不依赖于可信第三方的去中心化溯源架构成为可能。

综上，已有的数据进行溯源追踪方法的研究主要在于对数据完整性的溯源以及获得不可抵赖的溯源证据，如真实的不可篡改的操作记录。但现有的方法为了实现溯源证据的不可抵赖性，都不可避免的引入了CSP的协助产生操作记录。用户对CSP信任度的担忧，是上述方法均无法完全避免的问题。另外，目前对数据的溯源追踪，出现了溯源证据海量，溯源效率不足的隐患。而区块链技术提供了不可篡改与去中心化两大特点，对溯源证据的不可抵赖性与溯源效率提高提供了新的研究思路。

本报告中根据泄露的数据文件密文，利用密钥分发中心中记录的密钥信息，根据特定条件进行穷举搜索，找出泄露者的真实身份。

6.5.2 溯源场景及方法

数据溯源场景：

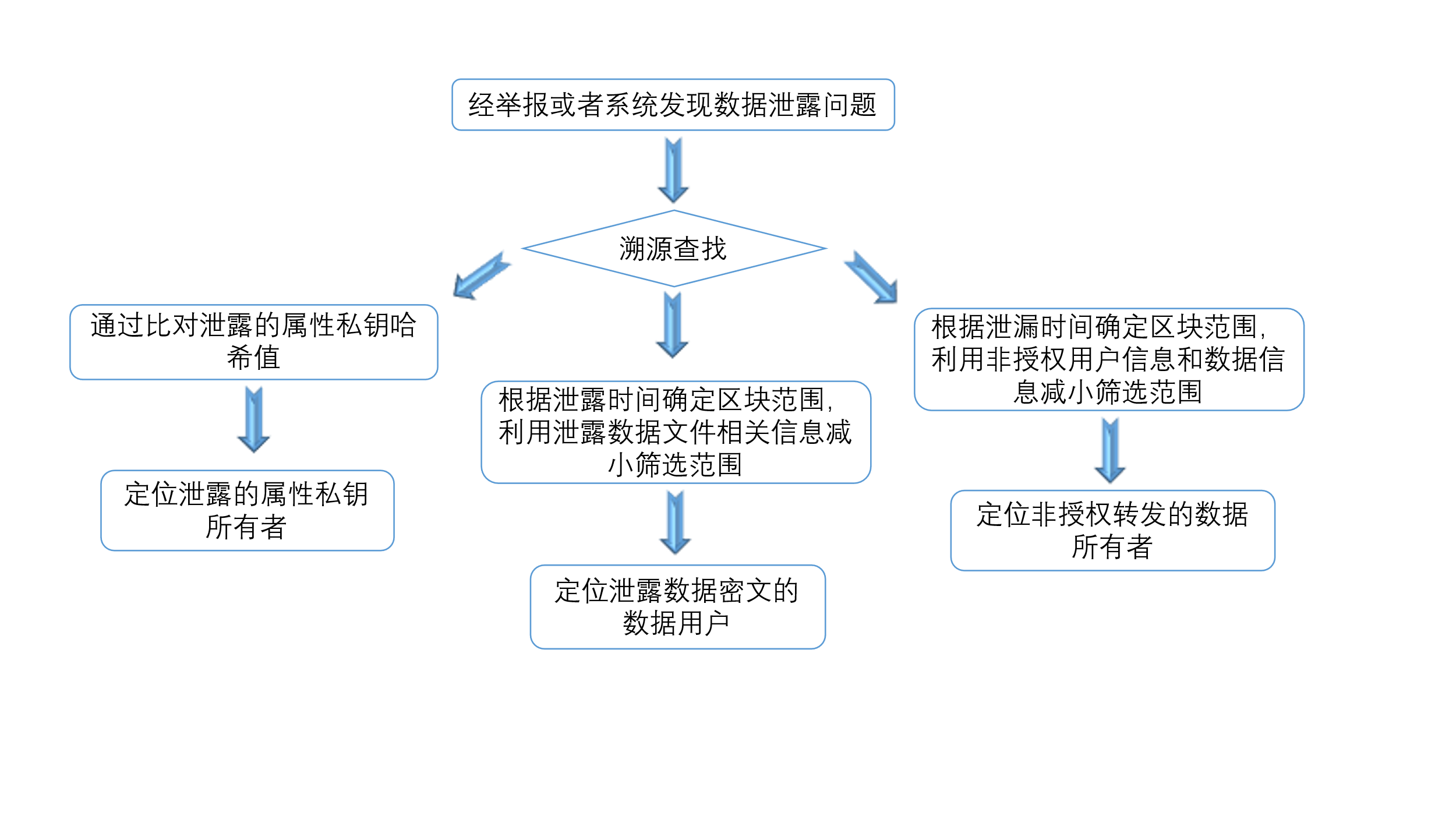


图 6-11 三种溯源场景描述

场景一：对于存储在云服务器上的数据文件，因为存在属性私钥故意泄露用于获利的情况，针对这种情形，由于属性私钥的生成过程中加入了随机数确保其唯一，可以将泄露的属性私钥作哈希运算通过比对区块链中交易记录中的属性私钥哈希值找到对应的数据用户。由于在数据分发的过程中数据发送方会将生成的属性私钥作哈希并打包进交易记录里最终存储于区块链中，可以先通过泄露时间确定离发生泄露时时间最接近的区块的时间戳，根据区块中记录的每一条记录，逐个对比。如果确定了部分泄露的数据文件的信息以及数据所有者信息，可以进一步筛选交易记录，减少要测试的记录范围，提升溯源追责效率。

场景二：如果用数据用户公钥加密的属性私钥或者对称密钥发生泄露，则需利用数据用户的私钥进行穷举试探。具体来说，从确定外泄的那个时间点开始算起，找到离外泄时间最近且小于外泄时间的区块的时间戳，从该区块开始往前找，逐个区块进行尝试，凡是区块中交易所包含的数据用户，用其私钥尝试对泄露的属性私钥或者对称密钥解密，能成功解密的私钥所对应的用户就是责任方。特别地，如果部分数据文件的信息也已泄露，确定了数据文件的种类或是对应的访问控制权限，则可以根据交易记录中的数据信息和访问权限进行筛选。即先利用时间，确定相应记录所在的区块范围，再根据其它已知条件进一步的筛选，减少需要尝试的解密次数。

场景三：发生了非授权转发，即数据用户没有相应的权限去请求数据文件，但数据发送方通过区块链网络将数据共享给了其它用户。对于此类情况，通常有两种可能，一种是业务部门向数据中台传输数据过程会发生非授权转发；另一种是数据中台向有数据需求的业务部门共享数据时的非授权转发。同以上两个场景一样，先根据泄露时间确定发生非授权转发的区块记录范围，逐个交易记录进行查询，如果获得了泄露的数据文件以及相关非授权用户信息，则可以根据相关信息进一步筛选交易记录，减少筛选的范围，提升溯源效率。

数据溯源方法：

(1) 基于时间的溯源方案策略：

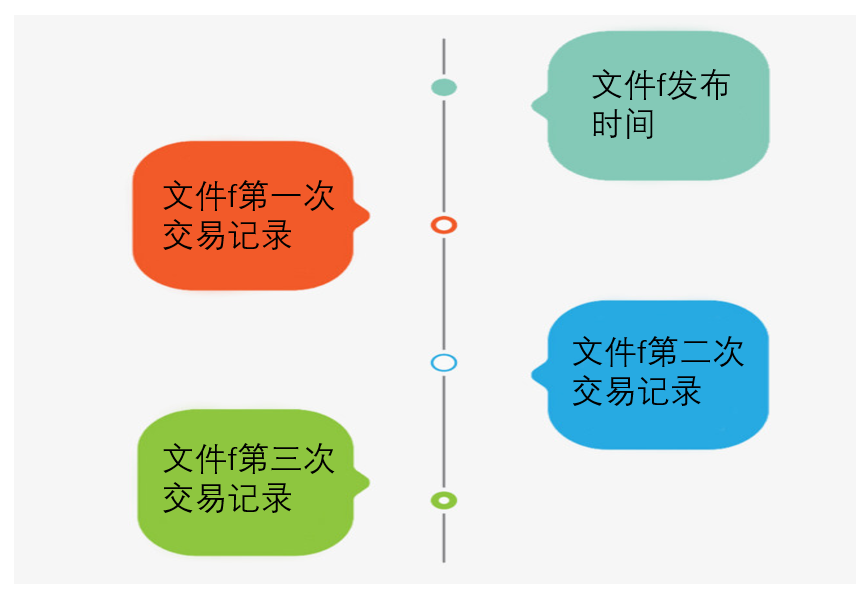


图 6-12 基于时间的数据溯源

根据要溯源的数据文件的发布时间和最后一次在区块链上的交易时间设定范围，可以只用这段时间内交易的用户私钥进行尝试解密，缩小搜索范围，降低搜索时间，加快搜索效率。

基于时间的搜索方案是一种简单有效的方法，具体来说，通过数据所有者发布的数据简介信息的时间，得到该份文件的“出生日期”,故对此文件所做的一切交易记录的时间都是在这之后；同时，通过溯源记录找到该文件最后一次转发的时间或者发生了泄露的时间作为溯源终点时间，而在此之后就不会有关于该文件的任何操作或者泄露与之后的转发无关。

这样，控制了要溯源文件的生命周期，有效缩小了要搜寻的范围，利用这段时间内有关于此数据文件的操作用户私钥进行尝试，便可对泄露的文件的属主做出有效的判断。

(2) 基于用户权限的溯源方案策略



图 6-12 基于权限的数据溯源

可以根据各部门各用户的权限，只对具有特定权限的用户私钥进行尝试，从而缩小搜索范围，加快搜索效率。

由于数据文件可能只在公司内部进行转发，公司内部各部门之间业务差异分明，赋予的权限各不相同。对于泄露的数据文件而言，查看其访问控制权限，这样，首先最有可能泄露数据文件的数据用户往往在具有相应完整权限的部门，用相应部门人员的私钥尝试数据文件的解密，从而减少了搜寻的目标范围。

然而，可能存在不具备权限的部门人员通过申请相应权限来访问数据文件，这时候需要结合其它溯源方案，增加可能泄露数据密文的目标人选，找出泄露者的真实身份。

(3) 基于多条件结合的溯源方案策略



图 6-13 基于多条件结合的数据溯源

将前两种策略结合起来，更有针对性的筛选目标，从而缩小搜索范围，加快搜索效率。

区块链中记录的溯源信息包括数据发送方、数据用户、交易时间、访问控制权限、用到的智能合约等，这样通过记录可以直接找到参与分发的人员信息。结合前两种溯源方案，一方面找到时间上符合条件的用户群体，另一方面找到直接具有相应权限的用户群体，通过两者的并集可以找到同时具有权限和时间的重点查询目标，极大提高了溯源效率，降低了查找时间。

假如数据所有者A拥有一份数据文件，并将该数据文件的相关简介信息发送到区块链网络之中，数据用户B浏览信息，发现A所拥有的数据文件正是自己需要的，于是向A发送了请求，A接收到智能合约转发来的请求之后，用B的公钥加密了数据文件并发送给B，同时将此条交易记录上链。在之后某个时间发现了一份加密的数据文件遭到泄露，查明该数据文件对应的信息后，首先查明该数据文件在区块链网络中发布的时间和发现泄露的时间，在这个事件段内找到所有关于这份数据文件的收发记录信息，然后根据用户的权限进一步筛选出能够直接获取数据文件的潜在目标，对这一部分对象的私钥先进行尝试数据文件解密；若全部不匹配，则用其他申请权限的用户私钥尝试解密，找到泄露者身份。

6.6 本章小结

本节主要做了基于加密的数据分发研究，结合常用加密、签名算法的利弊分析，提出一种基于区块链技术的数据分发和溯源技术，针对于数据文件在发送方不同存储方式，对应制定了不同的加密数据转发方案，并根据各种转发方案中数据泄露的不同场景适配了不同的数据泄露溯源策略，建立了可追溯,不可篡改,公开透明的共享模型，实现对数据发送方和数据 接收方的有效监管。与其他安全共享方案比较,数据共享安全系数更高，溯源效果更好，为用户或部门安全使用电力数据提供更便捷的服务与更可信的网络环境。

7 结论

本技术报告结合典型电力数据分发场景，利用区块链、加密等技术，研究不影响数据可用性的高效数据分发技术，建立数据分发溯源机制，提升电力数据分发管控能力，支撑非授权电力数据泄露的高效可信溯源问责。