区块链链上数据隐私保护的探索与研究

作者：朱烨东 袁波 出版时间：2018 年 08 月

摘要：

区块链是一种新的构建信任的技术，它被认为是人类历史上的第四次技术革命，将有可能彻底改变人类社会价值传递的方式。信任的基础是隐私数据能够得到有效的保护，本报告探索与研究区块链链上数据隐私保护的相关技术，从区块链的架构、交易、传输、存储及三方认证五个方面讲述了链上数据隐私保护的方法和策略，最后综合分析了隐私保护的两面性及对未来的展望：区块链链上数据隐私保护的各种方法在实现上既有互补又有依赖，它们之间只有通过紧密的配合才能更好地解决链上数据整体的隐私安全问题。同时，伴随着传统密码学及量子密码学技术的快速发展，区块链的应用市场也在日趋成熟，相信未来链上数据的隐私保护也将有新的、更大的突破。

关键词： 隐私保护区块链数据安全身份认证三方认证

Abstract:

Blockchain is a new technology that establish trust. It is considered as the fourth industrial revolution in human history，and will potentially change the way values are transmitted in society. As the basis of trust is the effective protection of data and privacy. In this article，we explore and research on the technology to enhance privacy protection from the perspective of data protection on blockchain. We present the analysis of methods and strategies of data protection on blockchain，in respect of blockchain architecture，storage，transmission，transaction，and tripartite authentication. Finally，we analyze the pros and cons of data protection on the blockchain comprehensively and discuss the future outlook.

Keywords： Privacy ProtectionBlockchainData SecurityIdentity AuthenticationTripartite AuthenticationDigital Signature

一 引言

区块链技术以人类社会第四次技术革命的态势，开启了重新定义人类社会价值传递方式的新篇章。美国、英国、欧盟、加拿大、俄罗斯、德国、日本、澳大利亚、中国等已经开始探索和制定区块链的相关标准与规范。

作为一种可靠的账本系统，区块链不仅让链上数据的每一次变化都能真实明确地记录在案，还能做到隐私保护和数据共享。但是随着大数据时代的到来，个人隐私数据（比如：身份、位置、银行账号、交易流水、社交等内容）正在以各种方式被泄露、滥用甚至形成黑色贩卖隐私数据的产业链。区块链技术借助去中心化、不可篡改、免信任、时间戳、分布式数据存储、智能合约、共识机制、密码学等方式集体维护的可靠账本系统，有望解决这个数据隐私保护难题。

目前金融科技、监管科技、数据存证、产权保护、商品溯源、物联网、供应链、公益慈善、共享经济等领域正在逐步开展区块链试点应用。本报告主要从区块链的架构、交易、传输、存储及三方认证五个方面对链上数据隐私保护的方法和策略进行探索和研究，目的是防止区块链上的敏感信息被人为或网络爬虫等非法跟踪、检索，从而遭遇骚扰、诈骗等情况。

最近中科金财研究院上线的区块链系统采用本报告所述的多种隐私保护混合技术，实际测试表明它能够有效地隐藏链上数据的交易细节，可以对交易主体的身份、内容及交易过程进行很好的保护。

二 架构上的隐私保护

如果区块链系统仍然采用一链通天下的设计方案，就很难解决链上数据隐私保护的问题，尤其是区块链各个节点间全数据拷贝的模式为不法分子获取全量数据提供了便利。

本节提出的链中链架构（包含一重多链和二重多链）是一种创新的、高效的区块链多链技术解决方案，特别为链上数据隐私保护进行了架构上的创新设计。通过短链、中链和长链的相互配合可满足大部分私有链及联盟链的业务需求，同时解决了传统区块链系统吞吐量低、交易慢、隐私保护能力弱的弊病，在区块链的扩展性、管理性、伸缩性、维护性等方面有了很大发展。

（一）一重多链架构

一重多链架构分为系统链和业务链，此架构为构建私有链的多链业务需求提供了优秀的解决方案。一重多链架构见图 1。

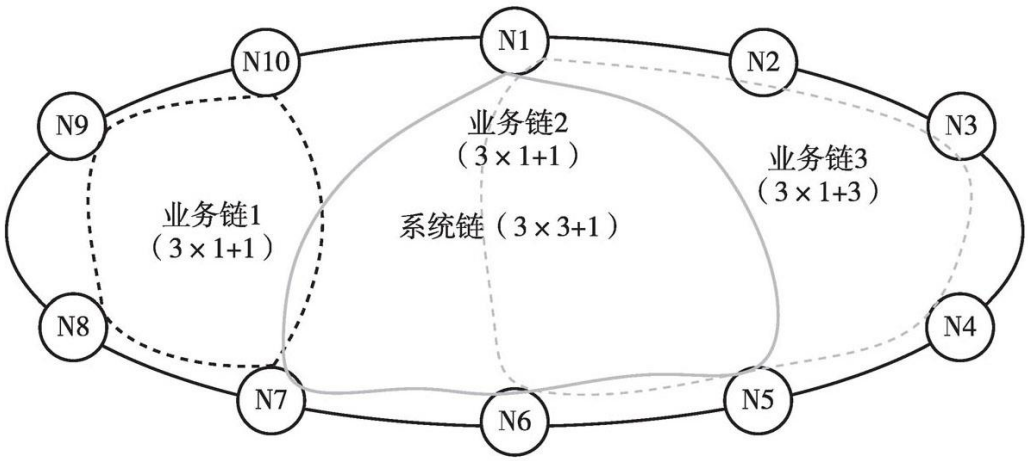

图 1 一重多链架构

一重多链架构以高性能、数据隐私保护、数据逻辑隔离等为主要技术特点。

典型应用是链上 CA（Certificate Authority）系统，CA 系统主要涉及身份登记管理服务和证书注册服务。传统证书颁发 CA 服务的结构为树形结构，存在单点故障、负载不均衡等诸多问题，而链上 CA 系统的根 CA（RootServer CA）和多个中间 CA（Intermediate CA）均配置成区块链系统中的逻辑节点，通过链上 CA 提供的服务可以实现数据共享、负载均衡并可以消除单点故障等传统 CA 问题。

（二）二重多链架构

二重多链架构是为跨企业、跨行业、跨区域等多场景互通的联盟链提供架构方案。二重多链架构见图 2。

二重多链架构以数据隐私保护、数据物理隔离、交互共享为主要技术特点。典型应用是联盟链间的数据共享，举个例子：企业 A 需要把部分交易数据 Pd1 和全量数据 Hash1 贡献给联盟链 M，企业 B 也需要把部分交易数据 Pd2 和全量数据 Hash2 贡献给联盟链 M，而联盟链 M 只会拥有 Pd=Pd1+Pd2 的交易数据的并集以及 A 和 B 全量交易数据的 Hash=Hash1+Hash2，不会拥有 A 和 B 的全量交易数据，这就为企业级区块链内部数据的隐私保护提供了很好的可控性。

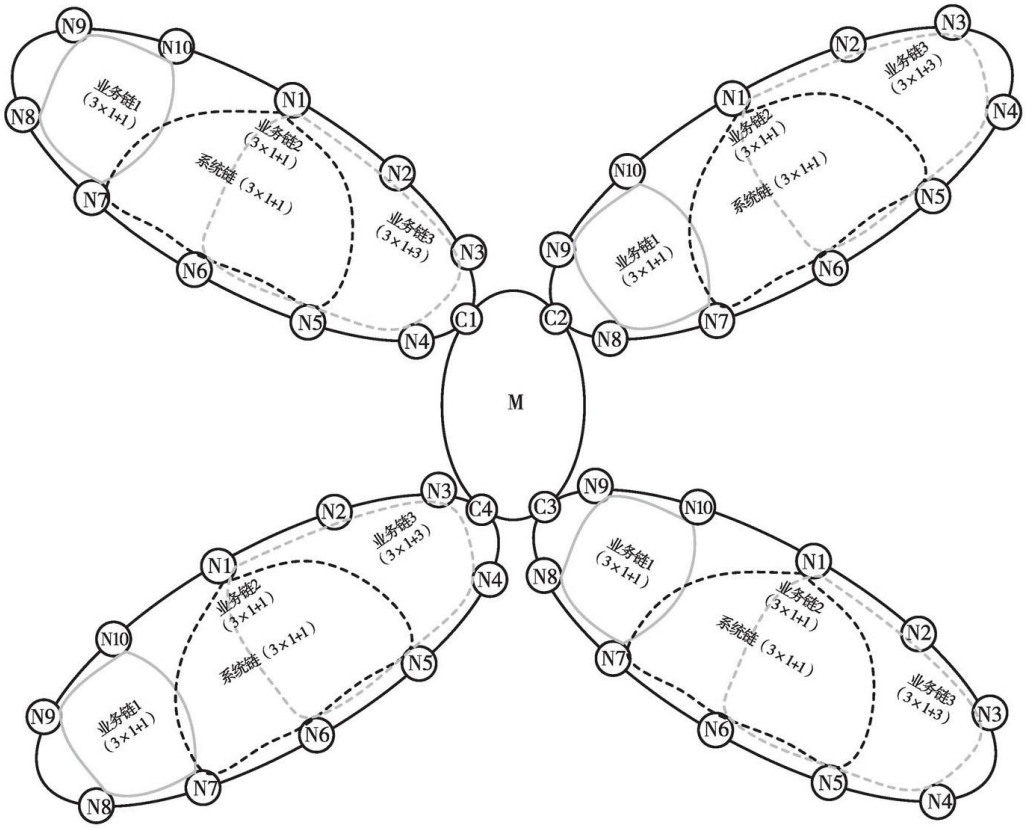

图 2 二重多链架构

三 交易过程中的隐私保护

在基于账户模型的区块链系统中，一笔交易数据的产生通常会涉及一方或多方账户信息的变动，如图 3 所示。

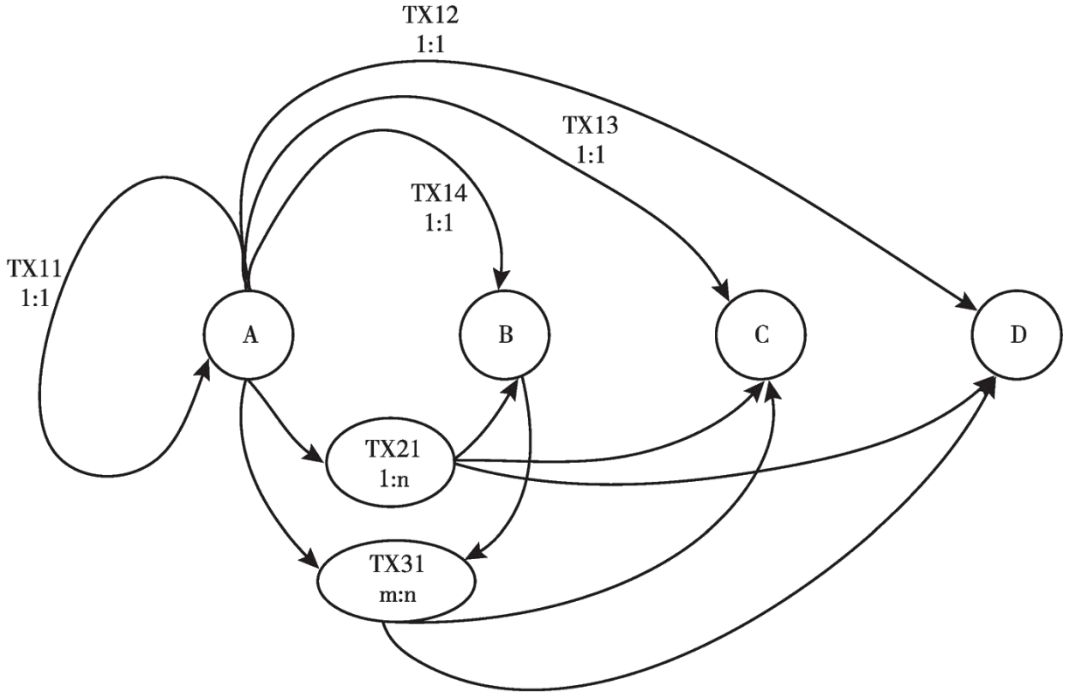

图 3 交易关联

一笔提交至区块链上的交易数据可能会涉及以下四种账户处理类型之一。

1.A 提交的交易数据只涉及 A 自己，主要用于 A 账户信息的更新、修改等操作。此类交易属于 A 账户单方地址参与的 1：1 型交易，如图 3 中的 TX11 所示。

2.A 提交的交易数据涉及 B 或 C 或 D，主要用于转账交易等简单操作。交易发起地址是 A，接收地址是 B 或 C 或 D，如图 3 中的 TX12、TX13、TX14 所示，此类交易属于 1：1 型。

3.A 提交的交易数据涉及 B、C 和 D，主要用于复合类型交易的操作。交易发起地址是 A，接收地址是 B、C 和 D，如图 3 中的 TX21 所示，此类交易需要进行分账处理，属于 1：n 型。

4.A 和 B 将组合交易数据包发送给 C 和 D，此类主要用于多账户间的组合交易。交易发起地址是 A 和 B，接收地址是 C 和 D，如图 3 中的 TX31 所示，此类交易需要进行先合账后分账处理，属于 m：n 型。

区块链链上数据在交易过程中的所有参与方的隐私问题同样重要，不仅要对交易接收方进行隐私保护，同时也应该对交易发起方进行隐私保护，只有把所有参与方的隐私数据都进行了保护，才是真正对区块链交易过程的全方位保护。

隐身地址通过临时密钥和地址来保护接收方的隐蔽性，而环签名是保护发送方的匿名性。在实践过程中我们需要把隐身地址和环签名一起结合起来从而才能达到更好的隐私保护的目的。

（一）地址保护

一般情况下通过分析区块链上的关联交易并结合社会工程学、大数据分析等技术可以跟踪到交易地址背后的真实身份。在区块链技术领域里出现了隐身地址的相关技术，目的就是隐匿交易痕迹以避免被跟踪。隐身地址是一种用于保护交易接收方隐私安全的技术，隐身地址的实现方式是发送方在创建交易的时候会为接收方随机生成一个交易接收地址，发送方同时需要确定接收方有打开该随机地址的钥匙。这样通过每次随机更换接收方的接收地址，他人就无法关联分析到谁是接收方，从而有效地防止跟踪，保护了接收方的隐私安全。

基础隐身地址协议（BSAP）是 2011 年提出的，2013 年有了改进的隐身地址协议（ISAP），2014 年又出现了一种双重密钥隐身地址协议（DKSAP）。

目前双重密钥隐身地址协议是一种高效的隐身地址解决方案，已经在很多交易系统中得到实施，该协议包括两对密钥，即“扫描密钥”对和“支付密钥”对，接收方可以共享“扫描私钥”和“支付公钥”给中间代理或扫描服务器，但是，这些中间代理或扫描服务器是无法计算出用于支付付款的临时私钥，它们仅可以代表接收方扫描这些区块链交易。

（二）签名保护

签名保护可以用来隐藏交易中的发送地址，比如使用环形签名可以使交易拥有多个发送方，但是这么多发送方中只有一个是真的，黑客通过查看环签名无法分辨出哪个地址是真正交易的发起方。环签名的重要性不言而喻，但是在介绍环签名之前，我们有必要先来了解一下盲签名和群签名。

1.盲签名

盲签名（Blind Signature）与其他签名算法最大的不同点在于需要将待签名的内容事先盲化。对于一般数字签名，签名者知道消息的原始内容，而盲签名需要掩盖原始内容，从而达到保护隐私的目的。

盲签名的最终输出应该与正常签名的输出相同，但是盲签名具有盲性和不可追踪的特点，因此在区块链系统中，引入盲签名机制可以将区块链和隐私保护技术有效结合，保障区块链用户的原始待签名数据的隐蔽性。在现有的区块链支付系统内，用户的支付行为是可追溯的，支付的完整情况、交易双方的地址都能够追踪，而在一些应用场景，例如匿名投票、匿名拍卖、电子现金等系统，我们希望能保护用户的隐私数据，保持一定的匿名性。

以匿名投票为例，参与投票的选民第一步需要把选票内容做盲化处理并附上自己的身份信息，投票中心在核验完选民的身份信息以后开始对盲化的选票内容进行签名，统计中心对该签名去盲化后就构造了安全有效的选票，该选票被保存到区块链上既能利用区块链的不可篡改的特性，又可以通过盲签名保护选民的隐私。

2.群签名

群签名（Group Signature）是一种可撤销的匿名签名算法。在群签名算法中，一个群内的所有成员均可以代表整个群体进行签名，验证者只需要使用单个群公钥即可验证该签名是否来自本群成员，但不能确定具体是哪个群成员，因此群签名具有一定的匿名性，形成对群成员的隐私保护。由于群管理员具有超级权限，可以打开签名，从而获取签名者的具体身份，因此该算法也有一定的不安全性。

在区块链应用场景中，经常会有同时满足匿名性和可追溯性的需求，而群签名便是能够满足该需求的一项密码学技术，特别是在金融、经济、军事、管理等领域。群签名和盲签名的结合形成的群盲签名系统已经在电子现金系统中使用了，中央银行担任该系统的管理员，多个银行可以在该系统中参与电子货币的发行，同时该系统可以结合区块链账本，有效利用区块链不可篡改、群签名匿名和可追溯的特性，形成一个安全可靠的电子现金系统。

3.环签名

环签名（Ring Signature）是基于群签名演变出来的，环签名和群签名较大的不同有两点：一是环签名简化了群签名，撤掉了管理员，这样环签名的匿名性是不可撤销的；二是任何一组用户都可以作为环签名的一个小组，可以看出环签名是一种更加去中心化的签名者模糊方案，因而对于区块链场景更加契合，也更容易被接受。在现有区块链系统中匿名性是一个非常重要的研究方向，在很多新型区块链系统的设计理念中，隐私保护和匿名性是其关键的卖点。

（三）零知识证明

零知识证明（Zero-Knowledge Proofs，ZKPs）是指在不泄露任何信息的情况下可以证明自己是某种权益的合法拥有者，零知识证明协议的内容大致如下。

1.协议的每一步必须串行有序执行。

2.协议至少有证明者和验证者参与。

3.协议执行完以后标示某项任务的完成。

零知识证明主要用于验证，在区块链交易过程中有大量需要验证的环节，所以可以将零知识证明大面积应用在区块链交易的验证过程中。一个最大的好处是它既能做到信息验证又能保护中间数据不被泄露，这种信息验证的方式对隐私保护是非常有效的。

四 传输中的隐私保护

在传输过程中窃听和篡改网络数据包是常有的事情，区块链交易数据包在节点间传递过程中可能也会被黑客截获，这对含有密码、敏感信息、个人资料的区块链交易数据包构成了极大的安全威胁。下面两种策略可以帮助解决传输过程中数据隐私保护的问题：一是链路加密；二是端到端加密。

1.链路加密

链路加密需要在链路节点上增加为所有数据包提供加解密服务的密码装置，在链路的中间节点上会有暴露明文数据包的风险，如果起始点和终止点之间跨越很多中间链路节点，那么整条传输链路的安全状况是由最薄弱的中间节点决定的。

2.端到端加密

端到端加密不像链路加密那样需要增加额外的密码装置，数据包在发送端使用软件或硬件加密，在接收方相应地使用软件或硬件解密即可。

区块链系统的各个节点之间的通信一般采用的是 P2P 协议，从方便易用的角度来说一般建议采用端到端加密。如果区块链节点间采用 HTTP 通信的话，那么建议考虑将 SSL 或 TLS 等技术整合使用以实现传输中的数据保护。

五 存储上的隐私保护

存储层加密的交易数据，一般只有交易参与方才能解密。通常存储在区块链中的交易数据特别是某些关键字段的内容（比如：交易额度、交易参与方等敏感信息）可以用对称加密、非对称加密也可以用同态加密。

对称加密代表算法有 DES、3DES、AES、SM4，非对称加密代表算法有 RSA、ECC、SM2，而同态加密则要相对复杂一些。

同态加密一个与众不同的特点就是能够对密文进行运算，这样就可以帮助我们解决存储加密与数据验证之间的矛盾。其意义在于可以从根本上避免将数据从区块链平台中取出时导致信息泄密。同态加密可以把链上数据的处理权与所有权进行分离，从而有效地保护用户的隐私数据不被泄露。

例如，企业的财务数据记录在链上，这部分数据并不适宜公开，但是会有相关审计部门需要对这些财务数据进行查询、审计等，这个时候利用同态加密机制，审计人员就可以在不解密链上数据的情况下进行查询、检索、比较等操作。在整个处理过程中无须对区块链上的数据进行解密，这种链上数据不外流的操作方式是一种非常好的数据隐私保护方式。

六 三方认证的隐私保护

（一）CA 认证

CA（Certificate Authority）是一个采用非对称密码体制实现的中心化的第三方证书授权中心，它主要负责签发、认证、管理已颁布的数字证书。区块链系统是一种去中心化或多中心化的分布式系统，如果直接把 CA 放在区块链系统中使用，就可能会出现“四不像”的结果，但是区块链系统需要学习和借鉴 CA 系统在身份认证、证书管理、安全管理等方面成熟的技术与经验。

在区块链多链架构下可以通过构建“链上 CA”系统来解决 CA 中心化的问题。利用“链上 CA”的分布式特性可以为区块链的节点和用户进行证书的签发和管理，不仅避免了 CA 系统的中心化问题，也可以利用 CA 保护链上交易数据的隐私安全。

（二）Kerberos 认证

Kerberos 是一种采用对称密码体制在非安全的网络环境中，为用户以安全的方式进行身份认证与授权的三方协议。Kerberos 的密钥分发中心（KDC）一般可以通过搭建一个主从配置来防止单点故障的问题，只是仍然不能解决类似拒绝服务攻击的威胁。

如果可以利用区块链的多链架构构建“链上 KDC”，那么就能解决 KDC 过度中心化的问题。利用“链上 KDC”的分布式特性为区块链节点的进出提供授权与认证服务，可以保证新增或退役的区块链节点在认证过程中涉及隐私安全。

七 隐私保护的两面性

在区块链网络中，特别当涉及企业级区块链应用的时候，链上数据的隐私保护和透明监管就成了矛盾的双方。隐私保护是为了保护个人敏感数据不被他人非法盗用，但完全的匿名又有可能为各类犯罪行为提供温床，出现没法监管的局面，所以隐私保护像一把双刃剑，如果完全做到匿名就可能会导致监管不到很多网络犯罪行为，从而无法追究匿名用户的责任，这也就会为一些违法犯罪行为提供不可追查的保护伞，会造成一系列社会问题。

如何既能满足监管，又能不侵害个人的数据隐私，这是区块链隐私保护技术需要反复博弈和平衡处理的一个重要问题。虽然目前还没有一个完美的解决方案，但在实际企业级区块链项目实施过程中，还是要混合使用多种隐私保护的技术，力求做到隐私与监管之间的平衡兼顾。

八 总结与展望

本报告从区块链的架构、交易、传输、存储及三方认证这五个角度对区块链链上数据隐私保护的方法和策略进行介绍，并且探讨了各种隐私保护的实现原理、特点及彼此的联系。总之，区块链链上数据隐私保护的各种方法在实现上既有互补又有依赖，它们之间只有通过紧密的配合才能更好地解决链上数据整体的隐私安全问题。同时，伴随着传统密码学及量子密码学技术的快速发展，区块链的应用市场也在日趋成熟，相信未来链上数据的隐私保护也将有新的、更大的突破。