区块链数据安全治理及关键技术解析

作者：朱烨东 袁波 ZhuYedong YuanBo出版时间：2019 年 08 月

摘要：

近年来随着区块链技术的快速发展，很多企业已经或正在将业务数据上链，大量的企业数据上链以后就会面临区块链数据安全治理的问题。本报告主要讨论联盟链和私有链的数据安全治理问题，以联盟组织和企业机构为数据安全治理的基本单位分析了区块链数据安全治理的理论与技术框架，为企业数据安全上链提供标准化、体系化的建设方案并最大限度地释放链上数据的业务价值。在本报告最后我们又对未来区块链数据安全治理的美好前景进行了展望。

关键词： 数据交换数据共享区块链数据数据分类分级数据安全治理

Abstract：

In recent years，with the rapid development of blockchain technology，many enterprises have already or in the process of migrate business data to blockchain，Meanwhile a large number of enterprise data will face the problem of blockchain data security management. This paper mainly discusses the data security governance of the alliance chain and the private chain. The alliance organization and enterprise organization are the basic units of data security governance for the theoretical and technical framework of blockchain data security governance is analyzed，and provide standards for enterprise data security for maximize the business value of blockchain data for enterprise. At the end of the article，we have prospects for the future of blockchain data security governance.

Keywords： Data exchangeData SharingBlockchain DataClassification and Grading for DataData Security Governance

一 引言

近年来随着云计算、大数据、物联网、区块链等技术的迅猛发展，数据安全问题也备受关注。大家都已经认识到在拥有大量数据以后，如果没有合理的数据安全治理方法防止数据泄露或滥用，就会对公民、社会、国家产生灾难性的危害，比如：2016 年 8 月山东女学生徐玉玉因电信“精准诈骗”而不幸离世。下面再列举 2018 年发生的多起严重的企业数据泄露事故：

1.2018 年 3 月 Facebook 有 8700 万名用户数据泄露；

2.2018 年 6 月前程无忧有 195 万条个人求职简历泄露；

3.2018 年 6 月圆通有 10 亿条用户信息数据被出售；

4.2018 年 8 月华住旗下多家连锁酒店有 2.4 亿个入住记录泄露。

虽然这些案例在区块链领域尚未发生，但是随着区块链技术的快速发展，很多企业已经或正在将业务数据上链，大量的企业数据上链以后就会面临区块链数据安全治理的问题，怎么防止企业数据泄露或被滥用也将成为联盟链和私有链需要解决的核心问题之一。

二 区块链数据安全治理现状

在介绍区块链数据安全治理之前我们先介绍一下数据安全治理的现状，并引用 Gartner 和 Microsoft 在数据安全治理上的框架进行分析。

（一）当前数据安全治理框架介绍

Gartner 认为当前数字业务正在为企业创造价值，但不能忽视其不断增长的业务风险和责任。安全和风险管理领导者应该构建适当的数据安全治理框架，以减少数据安全隐患所带来的风险。这些风险包括：

1.隐私保护的合规要求；

2.数据泄露对组织声誉和客户信任度的影响；

3.混合 IT 环境下通用数据安全策略的制定；

4.身份访问管理等安全产品通用策略的共享。

2018 年 4 月，Gartner 提出了数字安全治理框架（Data Security Governance，DSG），试图从组织的高层业务风险分析出发，对组织业务中的各个数据集进行识别、分类和管理，并针对数据集的数据流和数据分析库的机密性、完整性、可用性创建了 8 种安全策略。同时，数据管理与信息安全团队可以针对整合的业务数据生命周期过程进行业务影响分析（BIA），进一步挖掘各种数据隐私和数据保护风险，从而减少整体的业务风险。

不同于 Gartner 的数据安全治理，Microsoft 的数据治理框架（Data Governance for Privacy，Confidentiality and Compliance，DGPC）主要从人员、流程和技术三个角度出发，将框架重点放在数据安全的“树状结构”上，进而识别和管理与特定数据流相关的安全和隐私风险以及需要保护的信息。

在人员领域，DGPC 把数据安全相关组织分为战略层、战术层和操作层三个层次，每一层次都要明确组织中与数据安全相关的角色职责、资源配置和操作指南；在流程领域，DGPC 认为组织应首先检查与数据安全相关的各种法规、标准、政策和程序，明确必须满足的要求，并使其制度化与流程化，以指导数据安全实践；在技术领域，Microsoft 开发了一种工具（数据安全差距分析表）以用来分析与评估数据安全流程控制和技术控制相关的特定风险。

（二）我国数据安全治理的法律法规

《网络安全法》《电信和互联网用户个人信息保护规定》等的发布，是从国家战略高度和法规层面强调数据安全保护。安全和风险管理领导者应当引入适用于企业的数据安全治理框架去减少安全威胁、数据驻留和隐私问题带来的风险，从而避免由此带来的信誉损失和经济损失，树立数据安全治理的组织、制度、审计、管理、规划的管理思想。

由于当前国内外对区块链数据安全治理还处于开始阶段，大家都在讨论和探索怎样安全可控地利用好链上数据，因此本报告在借鉴、参考了大量传统关系型数据库及大数据平

台的数据安全治理方法后抛砖引玉式地提出以下针对联盟链和私有链的数据安全治理的方案，这些方案仅供大家在区块链项目实施过程中参考。

三 区块链数据安全治理研究和实施方案

区块链数据安全治理以链上数据的安全可用、可控为最终目标，围绕这一目标，我们将区块链数据安全治理的模型命名为区块链洋葱安全模型（见图 1）。

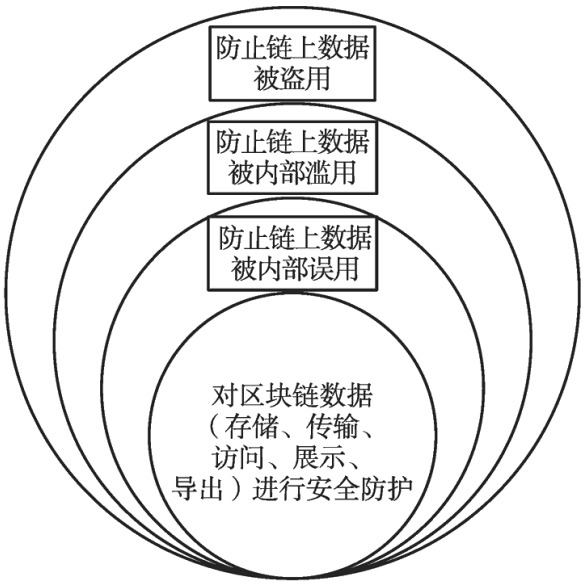

图 1 区块链洋葱安全模型

区块链洋葱安全模型主要包括 4 个层面：

1.防止链上数据被盗用；

2.防止链上数据被内部滥用；

3.防止链上数据被内部误用；

4.对区块链数据（存储、传输、访问、展示、导出）进行安全防护。

（一）防止链上数据被盗用

联盟链和私有链本身是需要授权才能登录和使用的区块链系统，这与传统的分布式应用系统的安全防护有很多共通之处，所以我们可以将防火墙、安全网关、网闸、堡垒机（跳转机）、防病毒检测系统、入侵检测系统等常规安全产品前置在联盟链和私有链网络中以防范非法入侵或非授权登录。这些安全产品对于内/外部恶意或病毒式入侵可以解决一部分安全问题，但并不能解决内部人员违规操作的问题，据有关部门统计，企业数据有一半以上的泄露是内部人员故意为之。在实际工作过程中，我们不仅要时刻提防数据被外部人员非法窃取，还要建立、健全使用联盟链和私有链上数据的授权审批机制和数据安全治理的规章制度，做好内部人员的技能和职业道德培训，防止内部人员违规操作导致区块链上重要数据被泄露。

链上数据被盗用通常是指黑客通过不合法的途径去攻击网络系统获得链上数据，或公司内部员工在不遵守公司规章制度的情况下私自拷贝或下载链上数据。

链上数据被盗用案例 1 某黑客甲精通网络攻击技术，通过对 A 公司某区块链节点的攻击，直接获取该节点服务器内区块链的持久化数据，导致该公司区块链上数据被盗。

链上数据被盗用案例 2 某职员甲在 A 公司任职区块链开发工程师期间，在未经主管同意的情况下，私自下载链上数据到本地，从而造成公司链上业务数据被盗和泄露。

链上数据被盗用的防范需要重点关注外部黑客主动攻击和内部职员故意泄露两个方面。对于黑客攻击，我们可以加强网络安全防护，比如采用防火墙、网络隔离设备、入侵检测、防病毒等防范措施；对于内部职员泄露，我们需要通过专业技能和职业道德的培训，防止内部职员因违规操作而导致区块链上重要数据泄露。

（二）防止链上数据被内部滥用

联盟链和私有链以多中心化节点的方式存储数据，节点拥有业务数据，如果不加以限制，操作人员就可能会在权限范围内随意访问和使用链上数据，这会导致用户或企业数据被滥用。

对客户的业务数据进行分类分级是解决区块链上数据被滥用的方法之一，对于不同级别和分类的数据设置不同的数据加密级别。区块链数据分类是指根据业务数据的属性或特征，将其按照一定的原则和方法进行区分和归类，并建立起一定的分类体系和排列顺序，以便更好地管理和使用区块链上的业务数据。区块链数据分级是指按照一定的分级原则对分类后的区块链上的业务数据进行定级，根据不同的级别采用不同的加密防范措施。

对链上数据进行分类分级和数据加密可以有效地减少操作人员滥用数据的风险。

（三）防止链上数据被内部误用

链上数据被内部误用通常指的是公司的技术人员由于缺乏专业训练或者个人的粗心大意，将链上数据错误地使用在非对应的场景上，从而造成非主观故意的不当后果。

链上数据被内部误用案例 1 员工甲编写了一段链码，该链码的功能为查询指定公司 a 产品在某一时间段内的出货量。员工甲在编写的过程中误将 a 产品的代号写成了 b 产品的代号，结果标题虽然显示为 a 产品，数据却是 b 产品的出货量，从而导致数据被误用。

链上数据被内部误用案例 2 员工乙接到通知需要将某个节点下的非核心数据进行备份和筛选。公司的技术规范中明确表示后缀为.block 的文件不可在未经技术主管确认下进行移动或备份。但员工乙由于未仔细学习操作规范，并且系统文件未针对访问人员设定分级操作权限，在接到通知后，员工乙就擅自将节点下关于区块存储数据的文件夹进行全部备份，从而产生了数据泄露的潜在危险。

链上数据被内部误用案例 3 员工丙在区块链系统中设置了一个每小时拉取日志的定时任务，该员工在未明确命名规则的情况下，将日志文件的命名规则设置为

YY_MM_DD_HH，而这恰好与数据文件的备份命名规则一致。在定时任务运行后，日志文件和数据文件被同时拉取，从而产生数据混乱的危险。

综上所述，预防链上数据被内部误用可以从以下 4 个方面着手：

1.公司方面需要定期对操作人员进行职业道德培训；

2.技能方面需要经常对操作人员进行提升技能的培训；

3.系统方面需要为区块链应用系统增加高级访问控制功能；

4.数据方面需要为链上数据增加分级授权使用功能。

（四）对区块链数据进行安全防护

需要从区块链数据本身的特点出发，进行综合安全防护。以下我们以超级账本（Hyperledger Fabric v1.4）和公正账本（Justledger v2.0）为例，从链上数据的存储、传输、访问、展示、导出五个维度进行区块链数据安全治理的研究（如图 2 所示）。

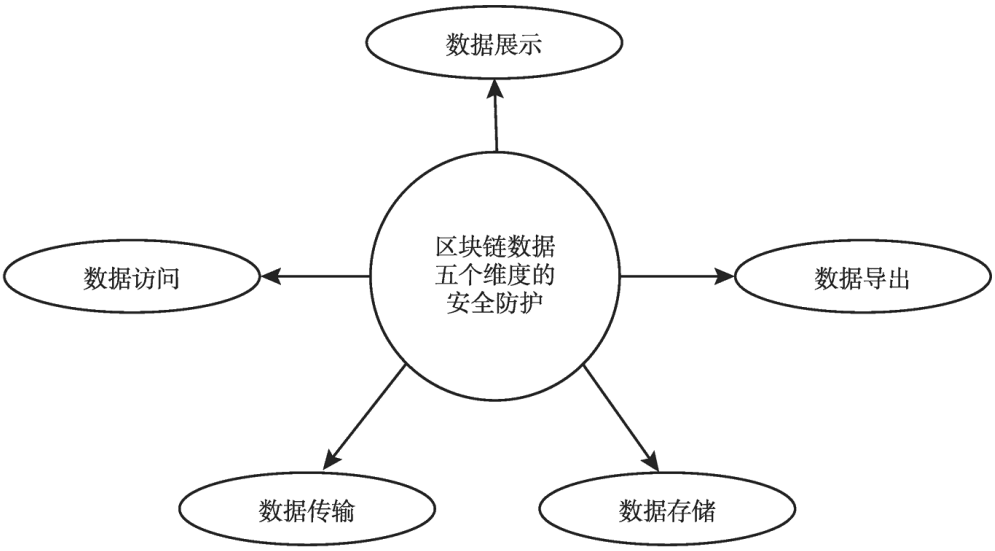

图 2 区块链数据五个维度的安全防护示意

需要说明的是，本报告介绍的区块链数据安全治理的技术和方法适用于超级账本和公正账本的区块链系统，对于其他类型的区块链系统也具有一定的参考价值。

1.区块链数据存储

超级账本和公正账本有三种类型的数据存储方式。

（1）账本数据，也就是区块链数据，是以文件形式存储的不可篡改的交易数据。

（2）索引数据，包括账本索引、区块索引和历史数据索引，主要是为了加速查找。

（3）状态数据，最新的区块链状态数据是链码执行业务逻辑的关键。

如图 3 所示，我们可以根据账本数据和状态数据进行不同类别和密级的分类，对每个通道中的账本进行分类分级并单独制定安全策略，明确通道中账本数据资产的分布和使用状况，只有实现账本数据存储加密才能保证原始账本数据的安全可用。状态数据是账本数据的最新状态值，可以在进行存储加密的同时结合数据访问管控措施来保证状态数据安全。

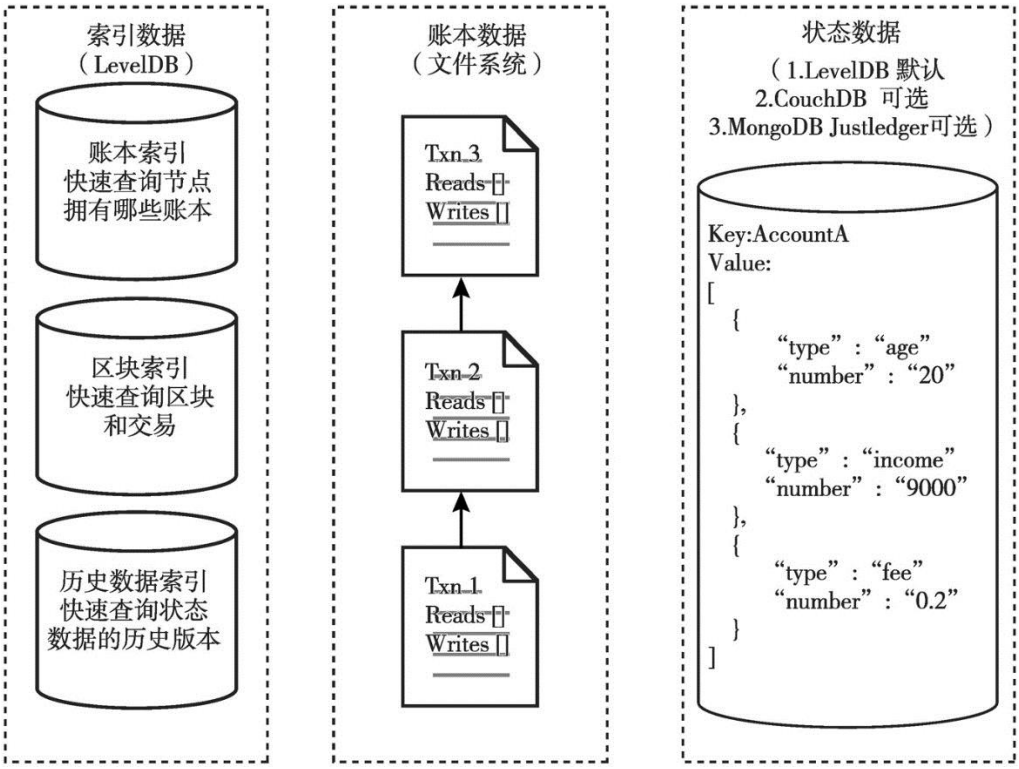

图 3 超级账本和公正账本的链上数据存储模型

2.区块链数据传输

区块链数据在传输过程中可能会面临数据窃听、数据篡改、身份伪装等安全威胁。通常我们可以采用链路加密和端到端加密等方式来减少传输过程中的风险。

链路加密需要为每个节点额外增加密码装置，密码装置主要为数据包提供加解密服务，整条传输链路的安全性遵从木桶原理，即由安全最薄弱的节点决定。

端到端加密不需要为节点增加密码装置，数据包在发送端使用软件或硬件加密，在接收端使用软件或硬件解密即可。目前区块链节点之间一般采用点对点协议进行通信，所以我们建议使用端到端的数据包加密方式，常用的做法是在 P2P 协议中加入 SSL 功能即可极大提升数据包的传输安全性。

3.区块链数据访问

联盟链和私有链一般都需要经过登录授权才能访问和使用链上数据，所以用户需要先发起链上数据访问请求，这个访问请求通常需要经过防火墙、安全网关、防病毒检测系

统、异常行为检测系统等之后才能到达区块链核心系统，然后由核心系统的链上代码负责为该请求返回链上数据。

通常防火墙、安全网关、防病毒检测系统可以借用传统的IT 软硬件基础设施，而异常行为检测系统则需要根据用户访问区块链数据的历史行为进行动态学习和建模，通过模型预测和评估该用户访问请求的合法性并及时拦截和告警。比如：一段时间内高频检索某个区块链账户的余额信息或异地高频登录区块链系统等行为都应该被区块链的异常行为检测系统阻止。

4.区块链数据展示

一般通过大屏进行全局可视化数据展示的目的是助力企业运营决策，通过条形图、饼状图、曲线图等可以更加直观、实时地观察链上数据的变化情况并辅助领导层在第一时间做出关键决策。

区块链数据展示通常会涉及交易数据展示、区块数据展示、统计数据展示，甚至部分敏感数据展示等。通过登录授权机制可以预防一部分数据展示的安全问题，但当涉及敏感数据展示时，就需要采用多种安全防护措施，比如定时敏感信息遮盖、敏感信息自动打码等。

5.区块链数据导出

区块链数据导出的目的包括存储备份、数据分析、数据共享、数据测试等，无论出于何种目的，只要是区块链上的数据脱离了区块链网络，我们就称之为数据下链。

数据下链的常用场景之一是通过 BI 分析工具对链上数据进行分析，目的是挖掘链上数据潜在的业务价值。但是如何从区块链上获得待分析的数据？一般我们有两种方法：一种方法是直接在链上通过 SDK 开通数据分析接口，然后与 BI 分析工具对接，但是由于目前 BI 分析工具的局限性，这种方法的实现难度相对比较大；另一种方法是将链上数据导出，暂存在文件系统或数据库系统中再与 BI 分析工具进行对接，这是目前常用的区块链数据分析方法。

数据下链之后最大的问题是如何保证下链数据的可控、可追溯，这是数据下链管理的关键所在，常用的技术包括数据脱敏和数据水印。

数据脱敏既可以保护业务数据的安全，又不妨碍数据的分析结果。区块链数据下链之前需要经过数据脱敏环节以对链上真实敏感数据进行模糊处理，脱敏后的数据需保留原始数据格式，关键敏感信息依照需求可以全部或部分用特殊符号代替，如身份证号码、交易地址、交易账号、交易名称等。

数据水印是为了保持对区块链下链之后数据的持续追踪，当发生或发现数据泄露之后，可以通过数据水印技术对泄密源头进行追溯。数据水印技术通过增加伪数据、不可见字符、可替换数据、单一分发源标示等在数据泄密之后通过泄密样本追踪到泄密源头，从而能够对泄密者进行定责和处罚。

结语

本报告介绍的区块链数据安全治理的技术和方法适用于联盟链和私有链的组网结构，对公链的数据安全治理也有一定的参考价值。区块链数据安全治理的基本原则不影响链上数据的正常使用，一般联盟链和私有链通过验证用户身份来获得链上数据的访问权限，其数据安全治理本身是一个复杂的系统工程，需要联合采用多种安全治理措施来预防数据被盗用、滥用、误用，同时又需要从存储、传输、访问、展示、导出等多个维度对数据内容本身进行安全防护。本报告仅是我们对当前联盟链和私有链的链上数据安全治理的研究总结，虽然当前国内外对该技术的研究还处于探索阶段，但我们相信未来肯定会出现更多、更好的关于区块链数据安全治理的技术和产品。