浙江大学
ZHEJIANG UNIVERSITY

区块链在数据安全领域的研究进展

学号：22151217

汇报人：王鑫

摘　要

1 摘要

大数据时代,数据已成为驱动社会发展的重要的资产.但是数据在其全生命周期均面临不同种类、不同层次的安全威胁,极大降低了用户进行数据共享的意愿.区块链具有**去中心化、去信任化和防篡改**的安全特性,为降低信息系统**单点化的风险**提供了重要的解决思路,能够应用于数据安全领域.该文从数据安全的核心特性入手,介绍区块链在增强**数据机密性、数据完整性和数据可用性**三个方向的最新研究成果,对各研究方向存在的缺陷进行分析.该文认为,区块链技术的合理应用能够增强分布式环境下的数据安全,有着广阔的前景.

研究背景

- 数据安全研究现状

2.1数据安全研究现状

数据安全是云计算大数据应用背景下的关键问题,也是难点所在.数据安全和数据全生命周期紧密关联,在采集生产、存储流转以及使用过程均面临一系列的安全问题.数据安全通常认为是数据生命周期中的**机密性、完整性和可用性**.

数据机密性缺失会直接导致**数据泄露**,而在分布式环境下,数据共享会带来更深层次的**隐私挖掘**.

数据完整性的需求存在于**数据采集传输、存储和使用**的多个阶段,其目的在于**识别损坏数据**的行为.

数据可用性是一个**系统性**的问题.李建中等人将数据可用性定义为**一致性、精确性、完整性、时效性和实体同一性**.在实际环境中,数据可用性主要面临的威胁来自于DDoS.

区块链增强数据机密性

- 数据加密
- 身份认证
- 访问控制
- 可执行环境
- 隐蔽通道

3 区块链增强数据机密性

应用方向	研究内容	解决的问题
数据加密	区块链应用于可搜索加密、代理重加密、安全多方计算	为密码协议建立可信第三方，为密码协议提供可靠的激励机制
身份认证	去中心化的 PKI 技术和身份管理	解决认证单点化，证书透明度和认证中心渎职的问题
访问控制	去中心化的访问控制模型	解决访问控制单点化问题，实现动态灵活的访问控制
	区块链增强属性加密机制	提高属性密码授权机构的可信度
可信执行	区块链增强可信执行 TEE 的安全性	实现可信的远程状态管理，提高 TEE 的可用性
隐蔽信道	区块链作为隐蔽信道的信息载体	解决隐蔽通信易被干扰篡改，信道单一，隐私性差等问题

区块链增强数据机密性的研究偏重于**学术层面**,能够对数据机密性保护方法存在的**缺陷**进行有效的补充

3.1 数据加密

区块链功能	应用方向	核心思想
分布式可信第三方	可搜索加密	代替中心服务器执行搜索[18]
	代理重加密	代替中心服务器执行重加密[19]
	安全多方计算	提供可信证据公告板，见证加密协议的公平性[34]
可靠的激励机制	可搜索加密	激励各参与方正确执行，惩罚作恶行为[18]
	代理重加密	激励重加密节点正确执行，惩罚作恶行为③
	安全多方计算	保证计算参与者的计算参与度，激励各方完成计算任务，惩罚作恶行为[24-32]

一是用区块链实现去中心化的服务,以建立**分布式的可信第三方**,以对抗**单点化风险**.

二是提供**可靠的激励机制**,从而降低攻击者的**作恶动机**,提高**作恶难度**

3.2 身份认证

(1)基于区块链构建**去中心化的公钥基础设施**,并基于分布式PKI为各类应用系统提供身份认证支撑

(2)基于区块链实现**去中心化的身份管理**,实现类电子身份证系统.

应用方向	研究内容	核心思想
去中心化的PKI系统	基于区块链实现的PKI体系	基于区块链交易实现证书全生命周期管理[36-44]
	改造现有的PKI系统	公开的证书审计和透明的证书撤销[45-47]
去中心化的身份管理	实现类电子身份证的管理	区块进行身份管理,实现多方跨域的身份认证[49-53]

3.3 访问控制

应用方向	研究内容	核心思想
去中心化的访问控制	面向通用场景的访问控制模型	区块链与访问控制模型结合，充当可信实体，实现权限策略不可篡改[54-56]和跨域访问控制[57]
		区块链充当访问控制实体，用交易或合约进行访问控制[58-63]
	面向物联网场景的访问控制	将访问控制权限作为资产，由区块链提供可信的存储，并且进行分布式权限交易管理[62-68]
区块链增强属性密码	密钥审计	记录所有操作增强密钥审计[73]
	提高授权机构的可信度	建立分布式属性判决和密钥管理中心，解决分布式环境下的多授权机构的互信和串谋问题[74-76]

一是实现去**中心化的访问控制**模型,解决信息系统尤其是物联网场景下中心化访问控制的安全和效率问题.

二是对基于密码学的**访问控制属性密码**进行**安全性的增强**,实现去中心化的授权中心

3.4 可信执行环境

区块链和可信执行环境(Trusted ExecutionEnvironment，TEE)结合有两个研究方向.一是用TEE解决区块链系统面临的问题:(1)面向拜占庭容错场景,基于TEE技术实现更加安全高效健壮的**共识算法**; (2)基于TEE构建智能合约的**安全执行环境**.二是用区块链技术弥补TEE无法解决的**安全问题**,主要研究包括Ekiden模型和ELI协议

TEE技术是解决安全计算的重要手段,能够为数据安全保护提供强大的安全模型.但是恶意主机能够对TEE产生很多影响:(1)阻断或篡改网络通信,**限制TEE和外部世界的通信联系**; (2)篡改非易失性数据,将旧的计算状态重新载入TEE的Enclave环境,实施**重放攻击**.即便TEE设计生产过程完全可靠,也会受到这两个问题的影响.

3.4 可信执行环境

针对TEE的可用性问题,目前有两种解决方案:

(1)在TEE的设计中,增加防篡改的非易失性存储器设计,但这类方案不仅会增加成本,而且不适用于分布式的计算环境;

(2)将Enclave的状态管理任务委托给远程的可信第三方,但是这种方法只是将信任根转移到了不同的物理位置,而远程服务器面临着同样的问题.

Ekiden和ELI协议都针对这一问题提出了解决方案,基本原理相似,本文以Ekiden为例进行介绍.在Ekiden中有三种实体：客户端是智能合约的用户;计算节点提供TEE服务,包括合约TEE和密钥管理TEE,并且所有支持TEE的平台都可以加入成为计算节点;共识节点维护不可篡改的区块链账本.Ekiden的流程是合约创建和合约执行.

3.5 隐蔽通道

隐蔽信道是一种**违反通信限制**规则,无法被监测的**隐蔽通信**手段,常应用于军事等特殊领域.隐蔽通信需要可靠的信息载体,如网络数据包、网络协议字段和时间特征等.传统的隐蔽传输通常采用单一信道的定向发送的模式,不仅**容易被检测**，并且传输受到网络环境的影响,**可靠性很差**.另外，身份隐私性也是隐蔽信道需要解决的难题.相关研究者提出用区块链技术实现隐蔽信道,展开了部分探索性工作并证明区块链适用于构建隐蔽信道,但是这个研究领域尚处于初期阶段.仍需要研究者进行大量的研究工作,实现实用化的隐蔽通信.

区块链增强数据完整性

- 数据完整性保护
- 云环境下数据可信管理

4.1 数据完整性保护

应用方向	研究内容	核心思想
数据确权与溯源	实现数据流转和溯源的管理	将数据流转记录和数据完整性证据写入区块链系统，保证数据在流转过程中不被篡改[98-106]
可信日志审计	建立基于区块链的日志审计系统	将日志和日志完整性证据写入区块链系统，实现日志数据无法被删除篡改，且能够恢复[107-113]
区块链＋	区块链应用于各行业	将数据确权和追溯需要的数据写入区块链系统[114-121]

数据确权与溯源，实现数据流转过程的可追溯记录;**可信日志审计**,为信息系统实现更加可信的日志审计系统;**区块链与各行业应用的结合**,为行业应用提供数据完整性保护功能.

4.2 云环境下数据可信管理

云数据审计,云数据可信删除和云虚拟机可信管理.其核心思想都是用**区块链作为不可篡改的存证**,确保**云数据管理服务的可信**.

应用方向	研究内容	核心思想
云数据审计	增强云存储和审计服务的可信度	审计结果记录到区块链网络,确保审计服务的可追溯性[120]
云数据可信删除	可公开验证的云数据删除	将数据删除命令和删除完成的证据存入区块链[121-122]
云虚拟机可信管理	云虚拟机和安全组件度量值的可信管理	用区块链管理虚拟机[123]或TCB[124]的度量值,确保可信度量值完整可信

区块链增强数据可用性

- 拜占庭环境下的一致性算法
- 基于区块链的分布式存储系统

5.1 拜占庭环境下的一致性算法

分布式技术的大规模应用要求数据服务具有高可用性,副本复制技术是提高可用性的关键技术,其核心问题是通过分布式共识算法实现副本之间的一致性.共识算法分为两大类,一种是非拜占庭容错的一致性算法,也叫**崩溃容错**(CFT),另一种是**拜占庭容错**(BFT).核心区别在于,CFT假设不存在恶意篡改和伪造数据的拜占庭节点,而BFT假设存在恶意节点.因此,BFT算法在实现中需要更高的**复杂度**,例如CFT类算法RAFT的复杂度为O(n),而BFT类算法 PBFT复杂度为$O(n^2)$.

5.1 拜占庭环境下的一致性算法

多项式级的通信复杂度导致BFT算法无法应用到复杂网络环境,业内普遍认为100个节点是BFT算法的上限.在实际构建分布式系统,尤其是大规模系统时,通常采用CFT类算法.随着网络安全威胁的日益增加,BFT的安全假设和实际应用场景更加吻合,但是在区块链技术出现之前,没有一个共识算法能够支撑大规模分布式环境下的拜占庭容错.区块链实现了拜占庭环境下的一致性算法,增强了分布式环境下数据和系统服务的可用性,可以看做区块链在数据安全领域的应用.

5.2 基于区块链的分布式存储系统

应用方向	研究方向	研究内容
区块链分布式存储	区块链本身作为分布式存储	用区块链存储数据，增强数据可用性，对抗 DDoS 攻击[134-137]
		区块链替换传统数据库的优化研究[138-142]
区块链应用于分布式存储	单信任主体的分布式存储	通过存储数据标签[134]或数据库操作日志[135]的方式，提高云存储的监管和审计能力
	去中心化的分布式存储	将区块链作为去中心化分布式存储的激励层，确保存储空间和检索服务充足稳定(Filecoin，Sia，Storj)[136,139-141]
		安全[138]和性能[137]的优化研究

(1)把区块链作为**分布式存储的一种实现形式**，讨论其实现、优化和应用场景;

(2)用区块链技术**解决现有分布式存储系统面临的问题**,实现更加健壮的分布式存储系统.

研究挑战

6 研究挑战

1 机密性

（1）数据加密：区块链是一个公开账本，恶意节点虽然无法作恶，但能够获取所有账本数据，进而能够进行**更加深入的数据分析**，因此需要进一步研究如何解决区块链的**隐私保护问题**；区块链需要拜占庭容错协议，但其**效率不高**，给区块链与数据加密结合的方案实用性带来较大影响

（2）身份认证：基于区块链的身份认证技术仍处于初级阶段,应用落地尚需时日.在去中心的尤其是大规模网络中,**密钥恢复、证书撤销、用户隐私保护、跨域身份认证**以及**认证效率**问题都是亟待解决的难点.另外,基于区块链的身份认证机制需要与现有的认证系统进行**融合**,才能够更容易得到推广.例如,引入无需可信管理者的**动态密码累加器**实现更加高效安全的认证;基于**秘密共享技术**实现分布式环境下的密钥恢复;使用**零知识证明**对身份认证过程进行隐私保护.

6 研究挑战

（3）访问控制

去中心化访问控制：

在面向通用场景的基于区块链的分布式访问控制方面,访问控制策略如何进行**更新或撤销**,如何承载复杂的**大数据量的交易**,链上的策略和权限进行**隐私性设计**,如何提高区块链访问控制的**响应速度**，如何实现**跨域的访问控制机制**都是需要研究的内容.

在面向物联网环境的基于区块链的分布式访问控制方面,针对智能生活的应用场景,需要设计**上下文感知的细粒度访问控制模型**,对不同用户采用不同的授权策略;实际应用中跨平台跨信任模型需要根据参与者的动态业务需求覆盖任意场景,需要研究在协作物联网中如何建立**跨平台信任模型**;区块链上的策略是动态变化的,尤其物联网环境中的设备和用户数量巨大,需要研究如何进行**高效快速的策略管理**;区块链部署需要较多的资源,需要研究**轻量级区块链系统**,其加密算法,密码协议和存储结构都需要特殊的轻量化设计.

6 研究挑战

区块链增强属性加密：目前使用区块链技术增强属性加密的基本思路相似,能够一定程度上解决控制**中心单点化问题**、多提高**授权机构互信**以及对抗多用户之间的**串谋攻击**.但是目前方案存在的问题是,多授权机构各自拥有属性私钥，容易产生**密钥泄露风险**.可以采用**秘密共享**等技术,实现分布式的密钥管理,进一步降低中心化的风险.另外,**属性密码滥用和责任认定**是该研究方向的痛点，而区块链有防抵赖的特性,两者的结合是一个有价值的研究方向.

（4）隐蔽信道：区块链应用于隐蔽信道构建是一个比较新的研究方向,当前的研究仅证明了该研究思路的可行性，但仍有较多问题需要解决.比如,如何在完全公开透明的场景下**隐藏数据传输**、如何建立**账号匿名机制**以对抗针对账本数据的**关联分析**、如何提高信息**恢复效率**、如何建立群组之间的**隐蔽信道**等.另外，从正反博弈的角度考虑,针对区块链隐蔽信道的检测技术同样需要进行研究,实现对恶意行为和非法通信的**追踪与溯源**.

6 研究挑战

2 完整性

（1）未能实现完整的数据安全闭环：根本原因在于目前的区块链本身是一个被动的账本，只能够相信起始数据的真实性,无法从语义和内涵上判断链上的记录是否可靠.例如,数据拥有者声称自己拥有数据，并将元数据和证据上链,此时区块链**只能相信而无法主动验证**.无论是数据确权与溯源，日志审计还是应用于云环境下实施数据可信删除,均面临这个问题.因此,需要进一步研究如何基于区块链技术实现一个**完整的数据确权溯源的闭环**,实现数据全生命周期的管控.

（2）用于大数据环境时的效率问题：为了实现数据完整性的保护,区块链平台承载的数据量和应用场景是紧密相关的.大数据应用场景具有**数据量大、数据动态变化、数据交易频繁**的特点,这就要求区块链具有良好的**吞吐量和延迟**.

6 研究挑战

3 可用性

（1）区块链实现分布式存储：基于区块链实现分布式存储时,需要对标传统分布式数据库的技术和功能.由于区块链数据库采用的是拜占庭容错的算法,本身存在性能上的差距，并且不具备完善的数据库功能.因此需要针对应用需求进行大规模的优化.

①将区块链**计算和共识解耦**,降低单节点的计算负担,并且设计更加适合查询与分析的区块链数据结构;②引入可信硬件(TPM，SGX),降低拜占庭容错协议的网络复杂度,**提高共识性能**;③改变共识协议的流程结构,针对具体的应用场景和安全假设,**简化共识协议的设计**,如Hotstuff算法;④采用分片技术(Sharding)提高事务处理速度,**降低成本**;⑤支持声明式语言(如SQL)的智能合约,以实现更**复杂的逻辑**满足分布式数据库系统的要求.

6 研究挑战

（2）区块链应用于分布式存储：区块链为去中心化的分布式存储提供了激励机制和审计机制,有助于激励各方积极参与,从而实现更大规模的分布式存储系统.但是,仍有一些问题亟待解决:1目前的应用模式本质上是去中心化的“网盘”,**不支持复杂的查询检索**;2**安全设计不足**,以IPFS为例,只要得到数据内容地址,就能够拿到完整的数据.;3现有的方案通常将数据加密后分片存储,但是密钥仍旧是集中式的管理,需要研究**去中心的密钥托管**方案，可以采用门限签名、多方安全计算等手段实现,KeyShard项目进行了相关探索.

THANKS