大数据背景下的密码学综述

（中北大学 物联网工程专业 学号 1807004120　姓名 秦嘉豪）

摘　要：随着当今社会科学技术快速发展，以互联网为基础的众多新技术被广泛运用的情况下，大数据应运而生。大数据通常来自于人们生活的方方面面，每天都会有海量的数据产生。每一个人都是数据的生产者和使用者，但是企业往往为了保护隐私而保证隐私导致数据共享性差，甚至出现了数据窃取等问题。区块链技术是一种在密码学、统计学、经济学和计算机科学与技术等多学科交叉的新技术。

关键词：密码学；加密算法；区块链；密码学应用及发展

1 引言

在当今信息大爆炸的时代，时时刻刻会产生海量的数据通过安全可靠的方式传输，数据安全已经严重影响了人们的生活，如何保证数据的安全已经成为了迫切需要解决的问题；随着当今社会科学技术快速发展，以互联网为基础的众多新技术被广泛运用的情况下，大数据应运而生。大数据通常来自于人们生活的方方面面，每天都会有海量的数据产生。每一个人都是数据的生产者和使用者，但是企业往往为了保护隐私而保证隐私导致数据共享性差，甚至出现了数据窃取等问题。因此，在传输关键的数据时，很有必要采用加密方法来保护其中的信息安全，而数字签名被用作大数据时代的信息交换中成熟的加密算法、确保信息在网络传输过程中保证信息的真实姓。区块链技术是一种在密码学、统计学、经济学和计算机科学与技术等多学科交叉的新技术。它是一个去中心化‘不可篡改、全程留痕可以追溯、集体维护’公开透明的特点。区块链与隐私保护的结合可以降低第三方监管不严的风险，在一定程度保证数据安全，具有更加广阔的应用价值。因此，本文对在大数据背景下的密码学和区块链具有极为重要的意义。

数据是信息的载体，为了使其方便的进行保存、处理和查询，经常存储于数据库中。而信息是含有价值的数据，二者之间相互联系。一方面，数据挖掘技术能够从大量的数据中进行分析，对用户行为进行预测，另一方面，用户将失去对数据的使用权。因此，为了保证数据信息安全，在技术上需要完成完善的安全加密算法。

2 密码学基本概念

2.1 密码学

密码学可分为：古典密码学和现代密码学。在西方语文中，密码学一词源于希腊语 kryptós“隐藏的”，和 gráphein“书写”。古典密码学主要关注信息的保密书写和传递，以及与其相对应的破译方法。而现代密码学不只关注信息保密问题，还同时涉及信息完整性验证（消息验证码）、信息发布的不可抵赖性（数字签名）、以及在分布式计算中产生的来源于内部和外部的攻击的所有信息安全问题。古典密码学与现代密码学的重要区别在于，古典密码学的编码和破译通常依赖于设计者和敌手的创造力与技巧，作为一种实用性艺术存在，并没有对于密码学原件的清晰定义。而现代密码学则起源于 20 世纪末出现的大量相关理论，这些理论使得

现代密码学成为了一种可以系统而严格地学习的科学。

2.2 对称加密算法

对称密钥加密是密码学中的一种加密法，是以转换其中一个数字、字母或仅字符串随机字母，一个秘密密钥会以特定的方式变更消息里面的文字或字母，例如更换字母相对位置（例如 hello 变成 lohel）。只要寄件者与收件者知道秘密密钥，他们可以加密和解密并使用这个数据。对称加密具有计算量小，效率高的特点。作为传统的对称加密的密码、原理比较简单。对于 26 个英文字母用标号 0-25 表示，密钥即为字母移动的位数，通过密钥实现加密解密，若消息窃取者不知道加解密的偏移量，将无法从密文中获取任何有价值的信息。然而因为加密逻辑较为简单，可以通过暴力枚举和频率分析法可以较为简单的获取有效的信息。暴力枚举是密码破解技术中最基本的技术，即对所有情况逐一尝试，直到密文的内容可读，从而推算出密钥的偏移量以便对消息进行持续的窃取；另一方面，由于对称加密算法比较简单，频率分析法也可以实现密文的破解。通过大量实验数据分析，可以较为轻松的破解出传送的明文。

2.3 非对称加密算法

非对称加密算法有两个密钥：公开密钥和私有密钥。公钥与私钥是一对，如果用公钥对数据进行加密，只有用对应的私钥才能解密。因为加密和解密使用的是两个不同的密钥，所以这种算法叫作非对称加密算法。 非对称加密算法实现机密信息交换的基本过程是：甲方生成一对密钥并将公钥公开，需要向甲方发送信息的其他角色(乙方)使用该密钥(甲方的公钥)对机密信息进行加密后再发送给甲方；甲方再用自己私钥对加密后的信息进行解密。甲方想要回复乙方时正好相反，使用乙方的公钥对数据进行加密，同理，乙方使用自己的私钥来进行解密。RSA 算法是最常用的算法。

3 区块链概念

区块链技术具有去中心化存储、隐私保护、防篡改等特点，提供了开放、分散和容错的事务机制，成为新一代匿名在线支付、付款和数字资产交易的核心，被广泛应用于各大交易平台。同时也给金融、监管机构、科技创新、农业以及政策等领域都带来了深刻的变革。据国家信息安全漏洞共享平台统计，自 2016 年到 2020 年的低位高危漏洞分析中，中危漏洞发生的频率较高。

区块链分为公有链、联盟链、私有链和混合链四大类。公有链是网络中任何人都可以随时访问的区块链系统，通常被认为是完全去中心化、匿名性高和数据不可篡改的区块链。联盟链为若干企业或机构共同管理的区块链，参与者要事先进行注册认证，因此相对于公有链来说，联盟链的参与节点较少。数据由认证后的参与者共同纪律和维护，这类节点拥有读取数据的权限。私有链是一种由某个用户控制的区块链，控制参与节点个数则严格，因此交易速度极快，隐私等级更高，不容易遭受攻击，相比于公有链系统由更高的安全性，但去中心化程度极大被削弱。

4 安全与认证

4.1 Hash 函数算法

哈希函数是一种将可变长度数据映射到固定长度摘要的函数,对输入数据的任何更改都

会导致哈希列中发生不可预测的变化。哈希函数具有三大经典特性：单向性（Hiding）、抗碰撞性（Collision resistance）、结果不可预测性（Puzzle friendly）。其中单向性：已知输入可以得到输出，但是已知输出无法逆推出输入。抗碰撞性：由于哈希函数拥有 2256 个输入空间，计算量趋于无限大，想要构造一个输入使其结果为当前值几乎不可能。结果不可预测性：对于已经公布的交易或是随机值，要推测得到特定特征的输出也是不可能的。在区块链的交易过程中为了保证数据的安全，首先对所有产生的交易数据都取哈希值 Hash（M）；其次将得到的哈希值与交易明文拼接得到最终发送内容 S；然后将 S 用发送方自己的私钥进行数字签名，最终将数字签名写入区块。

将任意长度的二进制值串映射为固定长度的二进制值串，这个映射的规则就是哈希算法，而通过原始数据映射之后得到的二进制值串就是哈希值。Hash 函数在整个过程当中扮演了极为重要的地位。Hash 函数是一种单项密码体制，有着不可逆的映射过程，现在常见的 Hash 函数加密算法有 MD5,SHA-1,SHA-256,SHA-2,SHA-512,SHA-3 等。

4.2 MD5 算法

MD5 即 Message-Digest Algorithm 5（信息-摘要算法 5），用于确保信息传输完整一致。是计算机广泛使用的杂凑算法之一（又译摘要算法、哈希算法），主流编程语言普遍已有 MD5 实现。

MD5 算法具有以下特点：

1、压缩性：任意长度的数据，算出的 MD5 值长度都是固定的。

2、容易计算：从原数据计算出 MD5 值很容易。

3、抗修改性：对原数据进行任何改动，哪怕只修改 1 个字节，所得到的 MD5 值都有很大区别。

4、强抗碰撞：已知原数据和其 MD5 值，想找到一个具有相同 MD5 值的数据（即伪造数据）是非常困难的。

MD5 的作用是让大容量信息在用数字签名软件签署私人密钥前被"压缩"成一种保密的格式（就是把一个任意长度的字节串变换成一定长的 16 进制数字串）。

4.3 SHA 算法

安全 Hash 函数（SHA）是使用最广泛的 Hash 函数。由于其他曾被广泛使用的 Hash 函数都被发现存在安全隐患，从 2005 年至今，SHA 或许是仅存的 Hash 算法标准。SHA 由美国标准与技术研究所（NIST）设计并于 1993 年发表，该版本称为 SHA-0，由于很快被发现存在安全隐患，1995 年发布了 SHA-1。2002 年，NIST 分别发布了 SHA-256、SHA-384、SHA-512，这些算法统称 SHA-2。2008 年又新增了 SHA-224。由于 SHA-1 已经不太安全，目前 SHA-2 各版本已成为主流。

	SHA-1	SHA-224	SHA-256	SHA-384	SHA-512
消息摘要长度	160	224	256	384	512
消息长度	$<2^{64}$	$<2^{64}$	$<2^{64}$	$<2^{128}$	$<2^{128}$
分组长度	512	512	512	1024	1024
字长度	32	32	32	64	64
步骤数	80	64	64	80	80

注:所有的长度以位为单位。

图 1 SHA 算法发展史

4.4 区块链

区块链主要采用哈希算法和非对称加密算法，哈希算法可以从固定输入得到唯一输出且计算过程不可逆，适用于构建区块和确认交易的完整性。非对称加密算法利用一组公钥私钥对接收到的交易进行数字签名和验证，确保交易内容和交易双方身份没有被篡改。区块链常用的非对称加密算法有 ECC（椭圆曲线密码学）、RSA（公开密钥密码体制）、ECDSA（椭圆曲线签名算法）等。

区块链加密模式如图 2 所示。用户 1 发起一笔交易给用户 2 的加密流程如下。

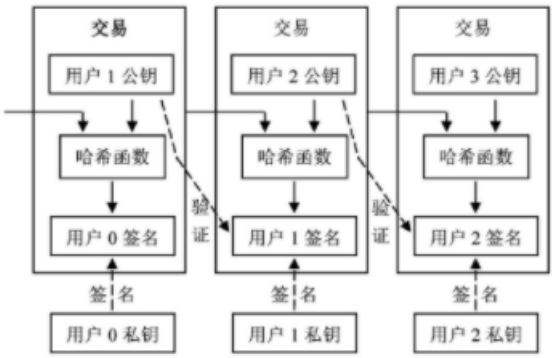

图 2 区块加密模式

1. 用户 1 对交易内容进行哈希计算得到唯一的哈希值，用私钥对哈希值进行签名。
2. 用户 1 将签名后的哈希值、交易内容等信息打包后向全网广播。
3. 用户 2 接收到广播包后，根据用户 1 的公钥对签名进行验证，以证明该交易确实由用户 1 发起；同时对交易内容进行哈希计算，将得到的哈希值与广播包的哈希值进行校验，以验证交易对象是否为用户 2 及交易内容是否合法。
4. 与此同时，区块链全网每个全功能节点都可以对交易内容进行验证。

4.5 智能合约

智能合约（Smart Contract）由 Szabo 在 1994 年首次提出，它是一段可以自动执行的计算机程序，在区块链中应用复杂的可编程语言和工程操作对区块链的执行步骤做约束，当账户触发特定条件时合约自动执行。其目的是为了保证合约双方不能恶意篡改合约内容，并能够在没有中心管理者的监管下确保合同有效实施，其代码具有自动执行和跟踪协议条款和条件的能力，因此，智能合约是自我验证、自我执行和不可逆的，智能合约被认为是区块链革命的第三代产品。

4.6 数字签名

在日常生活中, 一个人用签名来表明自己的身份。在法律上来讲, 一个可执行的签名方案应具备下述三个条件:1) 接收方能验证签名的真伪, 并且签名具备不可仿造的特点;2) 发送方始终不能否认自己的签名;3) 第三方能在双方争执时进行协调。手写签名具有较难的模仿特点, 因而广泛用于正式的合同签订中, 用来避免后续的法律纠纷等问题。然而, 电子信息化成为了时代主旋律, 数字签名技术逐步替代手写签名, 使数据的传输依然具有真实性和有效性。数字签名技术同样应用了私钥和公钥的概念, 其中公钥用于验证签名, 私钥用于产生签名。因为私钥只存在于发送消息者本人的手中, 因此具有极高的可靠性。

简单来讲, 数字签名的原理是发送者将原文通过 Hash 函数算法对原文生成一个信息摘要,

加密后随原文一起进行传输, 接收者用公钥解密, 并用 Hash 函数算法产生信息摘要并与解密后的信息摘要进行对比。如果不匹配则表明在传输途中经过了第三方的篡改, 相反则表明传输途中未被篡改, 消息真实可靠。这为信息网络中数据的传输安全奠定了良好的基础。

5 大数据此信息保护

如今, 在以互联网为基础的时代背景下, 用户的信息和数据都往更加开源的方向发展。但有些数据无论是对个人还是公司都是拥有隐私性或者商业性的, 中国正在处于一个信息共享化的时代, 比如你在网站上私密的查看了一些内容, 在你的购物 APP 中马上就有了相应物品的推荐。当然对于这样基于数据共享下的大数据分析, 在大多数情况下是有利于我们生活的, 但有些数据是属于个人的隐私, 这些数据是大家不想被拿来共享的。下面就法律条款对互联网数据、个人信息的保护做一定阐述。

5.1 数字版权法

数字版权法 (DMCA) , 也叫千禧年数字版权法, 可想该法律条文在千禧年颁布, 对互联网时代网上作品著作权的保护提供了法律依据。这也是保护互联网数据的第一步, 网上信息传播时, 防止他人对公共发布信息的篡改和摘抄。

5.2 合同法

合同法作为适用频率最高的法律之一, 与人们的生活密切相关。说到数据安全防护方面, 在生活中使用互联网 APP、网页等都需要实名注册, 在注册时会有选择是否同意或接受该程序的相应法律条款。仔细阅读大量合法程序的法律条款, 在条款中都有基于法律相关数据保护的说明。

5.3 网络安全法

网络安全法, 该法在原人大决定、工信部规定等基础上, 借鉴吸收其他国家的政策, 将个人数据保护收入《网络安全法》之中, 可谓在个人数据保护行为上的一个重大突破。该法规定, 当发生或者可能发生个人信息泄露、毁损、丢失的情况时, 要求网络运营者及时履行告知义务, 以使用户知晓, 增强用户对相关安全行为的警惕性, 同时报告有关主管部门, 让其第一时间对该行为作出审查。

5.4 刑法

刑法, 基于上述各个法律的相辅相成, 在数据安全环境奠定了一定基础, 但为了满足大数据时代数据安全的新形势和新诉求, 还应最大程度上加强惩治力度, 增强威慑力, 完善我国《刑法》相关体系和规定。传统刑法主要针对的是保护信息系统的完整性, 而在大数据时代中云计算资源保护和密码保护问题日益显现, 我国《刑法》在应对互联网技术变革时要做出适度回应和调整。

6 结语

本文主要介绍了在大数据时代背景下密码学的应用、密码学的发展史、几种主要的加密算法以及区块链技术在密码学的应用，阐述了密码学的基本概念。信息安全是大数据时代重要的研究方向。在技术上需要完善的安全保障算法, 在法律上要有明确的约束。本文介绍了信

息传输中的对称加密和非对称加密算法, 数字签名技术和 SHA-1 算法, 并且结合数字版权法, 合同法, 网络安全法, 个人数据法和刑法对法律与大数据信息保护进行了分析。

参考文献

[1] 文雨，彭程，刘家铭等．大数据时代信息安全保障算法和法律的融合[I]．北京：信息科技，2018，14

[2] 单康康，袁书宏等．区块链技术及应用研究综述[J]．电信快报，2020，（11）

[3] 刘双印，雷墨翳兮，王璐等．区块链关键技术及存在问题研究综述[J]．计算机工程与应用，2021,11

[4] 肖国镇，卢明欣，秦磊，来学嘉．密码学的新领域：DNA 密码．科学通报，2006（51）：1139-1144

[5] 罗婉平. 现代计算机密码学及其发展前景．江西广播电视大学学报，2009（3）：79-80

[6] 曹珍富，薛庆水．密码学的发展方向与最新进展．计算机教育，2005（1）：19-21

[7] 杨波．现代密码学[M]．北京：清华大学出版社，2017

投稿截图：

稿件类型：	中文稿件
稿号：	znj202111032
标题：	大数据背景下的密码学综述
作者：	秦嘉豪[中北大学](通讯作者)
关键词：	密码学;加密算法;区块链;密码学应用及发展
摘要：	随着当今社会科学技术快速发展，以互联网为基础的众多新技术被广泛运用的情况下，大数据应运而生。大数据通常来自于人们生活的方方面面，每天都会有海量的数据产生。每一个人都是数据的生产者和使用者，但是企业往往为了保护隐私而保证隐私导致数据共享性差，甚至出现了数据窃取等问题。区块链技术是一种在密码学、统计学、经济学和计算机科学与技术等多学科交叉的新技术。
上传文件：	原稿全文：文献综述.pdf
计划栏目：	其他
基金类别导航：	无基金
通信邮箱：	q656673477@gmail.com
联系地址：	山西省太原市尖草坪区中北大学
邮编：	000000
手机：	18603428353
著作权授权声明：	全体著作权人同意：论文将提交《中国农业教育》期刊发表，一经录用，本论文数字化复制权、发行权、汇编权及信息网络传播权将转让予《中国农业教育》期刊编辑部。 同意

稿件类型：	中文稿件
稿号：	2021110002
标题：	大数据背景下的密码学综述
作者：	秦嘉豪[中北大学](通讯作者)
关键词：	密码学;加密算法;区块链;密码学应用及发展
摘要：	随着当今社会科学技术快速发展，以互联网为基础的众多新技术被广泛运用的情况下，大数据应运而生。大数据通常来自于人们生活的方方面面，每天都会有海量的数据产生。每一个人都是数据的生产者和使用者，但是企业往往为了保护隐私而保证隐私导致数据共享性差，甚至出现了数据窃取等问题。区块链技术是一种在密码学、统计学、经济学和计算机科学与技术等多学科交叉的新技术。
上传文件：	原稿全文：信息安全技术课程报告.docx
通信邮箱：	q656673477@gmail.com
联系地址：	山西省太原市尖草坪区中北大学
邮编：	000000
手机：	18603428353
著作权授权声明：	全体著作权人同意：论文将提交《电信快报》期刊发表，一经录用，本论文数字化复制权、发行权、汇编权及信息网络传播权将转让予《电信快报》期刊编辑部。 同意