摩石观察

密码是保障网络安全的核心技术和基础支撑，在维护国家安全、促进经济社会发展、保护人民群众利益中发挥着不可替代的重要作用。云物移大智的蓬勃发展，5G、智慧城市、互联网＋政务服务的全力推进，离不开用密码技术来保障网络安全、保护数据安全、保证网上诚信，需要密码学与其它学科深入合作，需要密码产业与其它产业的深度融合，需要产学研管用的真诚协作，需要全社会共同传播密码知识与政策、研究密码应用技术、推进密码应用方案。

为激浊扬清，构建以密码技术为核心、多种技术相互融合的新网络安全体系，推进密码技术科学规范应用，长期深耕于信息安全一线的卫士通公司凝聚了一批国内顶尖的密码专家于2016年建立了摩石实验室。依托密码基础理论，探索密码创新实践，解决密码应用难题，培养密码专业人才，摩石实验室致力于让密码技术更好地服务于网络强国、数字中国和智慧社会。

本着共同的愿景，《信息安全与通信保密》杂志社与摩石实验室精诚合作，专门开辟《摩石观察》栏目。立足于密码本真，反思密码实践，《摩石观察》将以密码人的细致严谨叩问密码创新的真理之门，为广大读者了解、认识、掌握、使用密码技术提供准确规范的参考依据；同时我们期望以密码会友，与理想作伴，热忱邀请有志之士共同探索密码理论与应用的最佳实践，为推动金融等重要领域密码应用与创新发展而奋斗。

公有链上数据如何保护？

贾音
（摩石实验室技术专家）

1 引言

2017年对于区块链技术来说是不平凡的一年。人们通过ICO对区块链项目募集了数百亿资金，极大地壮大了数字货币的规模。随着它被广泛应用，区块链技术在金融、医疗等诸多应用场景下都显示出了它的价值和优势，尤其是公有链。然而，区块链技术尚不成熟，公有链在应用时还有很多缺陷和问题，如可扩展性、数据的隐私性、链上数据来源的可靠性、智能合约的安全性等，这些问题极大地阻碍了公有链技术的进一步发展。本文主要对公有链上的数据如何保护的问题进行探讨，参考学术界

提出的尚未落地的解决思路，结合工业界目前已得到应用的解决方案，对该问题进行总体的介绍。

在公有链技术中，每个节点都可以按照自己的意愿随时地加入、退出区块链网络，网络中的全节点都可以获得一份完整的账本数据。这就意味着，公有链上的所有数据，如普通的转账交易、部署的智能合约、合约的内部存储等都是公开的。无论是企业还是个人，他们都不会乐意将自己的重要数据，如电子医疗记录、交易记录、业务往来信息等公开到区块链上。因此，如何对公有链上的数据进行保护，成为公有链技术发展所面临的一个重要问题。

根据公有链上保存的数据类型，我们把需要保护的数据分为以下两类：无格式数据以及需要进行逻辑运算的数据。下面我们将分别介绍针对这两类数据的保护方案。

2 无格式数据及其保护

无格式数据是指保存在区块链上的，对数据格式无要求、无需进行有效性验证的数据。这类数据在公有链上仅仅是作为交易数据中的一个额外字段而存在，如 Zcash 中的 memo 域以及 Neo 中的交易属性字段。这些字段可以保存一些附加的信息，如交易的备注、发送方想要告知接收方并将其保存在区块链上以作存证的信息等。由于这类信息无具体的格式要求、有效性要求，并且面向的群体比较有限，因此可以直接加密，并通过在群体之间共享密钥的方法来保证消息的可用性。由于这类方案的思路大致相同，我们以 Zcash 为例来详细地论证方案的实施过程。

Zcash 中的货币以票据 Note 的形式存在，$Note=(a_{pk},v,\rho,r)$，其中 v 表示金额，ρ ,r 是两个随机数。交易的发送方 A 在向接收者 B 转账的时候，需要为接收者 B 生成新的随机数 ρ',r'，同时需要将这两个随机数告知 B，并且不能被其他无关人员得到。因此，发送方 A 会根据以下方案与接收方 B 共享密钥，并使用密钥来对随机数 ρ',r'，以及备注信息 memo 进行加密，其具体过程如下：

（1）加密

A 随机生成一对密钥协商方案的密钥（epk,esk），然后根据以下公式生成一个共享秘密，其中 pk_B 为 B 的公钥：

$sharedSecret=KA.Agree(esk,pk_B)$

接着，根据共享的秘密，B 的公钥和一些其他信息，利用密钥提取函数 KDF 生成加密密钥 K，最后使用对称加密方案来对 P 进行加密。

$K=KDF(sharedSecret,pk_B,t)$

$C=Sym.Encrypt_K(P)$

最终传输的密文即为 (epk,C)。

（2）解密

B 收到密文 (epk,C) 之后，首先计算出共享的秘密，其中 sk_B 为 B 的私钥：

$sharedSecret=KA.Agree(epk,sk_B)$

然后，利用相同的方法计算解密密钥 K，最后使用解密算法对密文 C 解密：

$K=KDF(sharedSecret,pk_B,t)$

$P=Sym.Decrypt_K(C)$

由上述过程可以看出，对于链上保存的无格式数据，使用基本的密钥协商、加密方案就

可以满足数据保护的要求。接下来我们主要介绍需要进行逻辑运算的数据在公有链上如何保密。

3 逻辑运算的数据及其保护

保存在公有链上的，需要进行某种逻辑运算的数据包括：普通的交易数据、智能合约的输入、存储以及合约本身。对于这类数据，需要网络上的节点根据预先定义的规则对数据进行验证或运算，如对交易数据的有效性进行验证，对智能合约进行验证、对发送给智能合约的数据进行运算，并对存储进行更新等。对于这类数据而言，不能通过简单的加密来达到保护的目的。依照区块链技术的不同发展阶段，我们将这部分数据分为两类，第一类是以比特币等数字货币为代表的交易数据，第二类是以以太坊等平台为代表的智能合约的相关数据，接下来分别介绍这两类数据的保护方式。

4 比特币等数字货币的基本交易数据

区块链技术发展初期，主要应用集中在以比特币等为代表的数字货币。对于这类区块链应用，链上数据主要是一些基本的交易数据。以比特币的 UTXO 模型为例，其基本数据格式如图 1 所示：

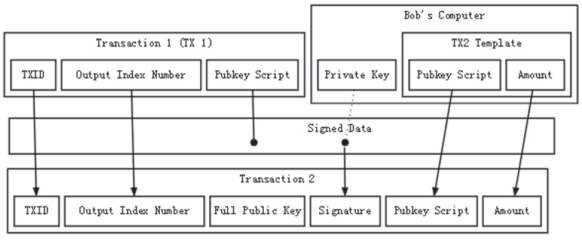

图 1 比特币的数据格式

可以看出，一条基本的交易数据包含了上一次未花费输出的交易号、序号，以及发送者的完整公钥，发送金额，为接收者设置的锁定脚本，以及对整笔交易的签名。由于网络上的其他节点在收到这笔交易后需要对交易进行有效性验证，即验证交易的输入是否合法，签名是否有效，输入输出是否平衡等，简单地对交易数据加密后会造成交易有效性无法得到验证。因此，需要一些其他的方法，从不同角度、不同程度上对数据进行保护。

目前的保护方法可以分为两类：一类是从工程角度出发，采用一些原理简单、便于实现和应用的方法，如比特币的 Coinjoin 机制，达世币的匿名发送机制，以太坊的状态通道机制等；另外一类是基于一些比较高级的密码学工具，如环签名、零知识证明、同态加密，以及一些困难问题如椭圆曲线离散对数问题等从本质上对数据保护方式进行改进，同时又不影响交易的可验证性，但这类方案实现成本较高，同时也带来了效率方面的损失。

4.1 工程方面的解决思路

工程方面的解决思路着重于考虑方案的易用性和应用效率，因此往往会在其他方面有一些牺牲，如保护强度、去中心化等。

比特币的 Coinjoin 机制的基本思想是，如果你想要发起一笔支付，那么就去找一些跟你一样的人，然后做一笔联合支付。

如图 2 所示，联合交易的输入来自三个人，输出也由这三个人组成，同时他们先后对于这笔联合交易进行签名，保证了没有人能够改变交易的输入输出。并且，根据这笔交易，任何

无关方无法推断出哪个地址属于哪个参与方，保证了资金的不可追踪性，并为交易带来了一定的匿名性。

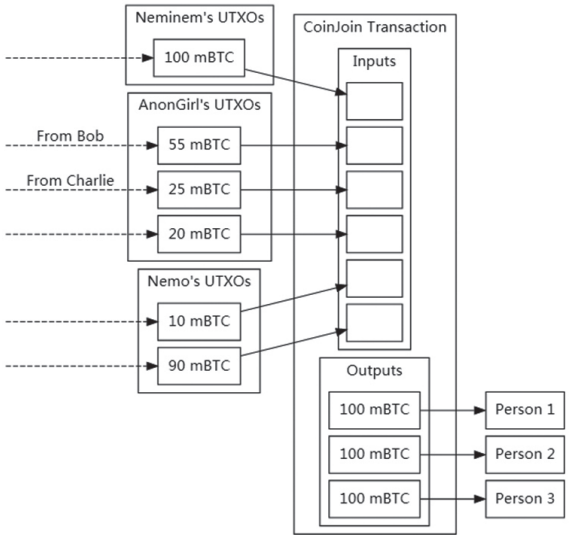

图 2 比特币的 Coinjoin 机制

达世币的匿名发送是基于比特币的 Coinjoin 机制构造的，只是增加了一层主节点帮助需要参与方进行联合交易。同时引入了“链式混淆”，即允许经过最多 8 个主节点来对资金进行多次混淆，降低单一主节点作恶对交易匿名性的影响。

以太坊的状态通道技术是为了解决以太坊的网络处理能力而存在，隐私保护只是它的一个附加功能。它的主要思想是把大量高频、小额的交易放到链外处理，只有初始创建通道和最终结算的交易需要在链上进行。这样就保证了只有初始状态和最终状态为无关方所知，而中间状态只能被交易参与方所掌握。

4.2 密码学工具的运用

与工程方面的解决思路不同，利用密码学工具来实现数据保护的方案则力求对数据实现本质的保护，因此往往能够提供较好的保护强度，但对方案的效率、易用性影响较大。

CryptoNote 协议引入可链接性环签名和一次性地址技术来力求增强比特币的匿名性，后来一些数字货币如 ByteCoin，门罗币等都是基于该协议构建的。

环签名方案最早由 Rivest 等人提出，他们将可能的签名者集合定义为一个环。环签名的好处在于，签名者只要拿到所有环成员的公钥及自己的私钥即可产生环签名，而无需其他成员主动配合。同时验证者拿到该签名，只能判断该签名是由这个环产生的，但不能定位到具体的签名者。可链接性环签名则是指如果一个人先后用自己的同一个公私钥进行了两次环签名，那么这两次签名就可以被链接起来，指向同一个人。在 CryptoNote 协议中，可链接环签名既可以将交易的发送地址隐藏在一个环里，让人无法根据区块链上的交易数据判断交易的发送方，同时也通过可链接性来防止双花。

一次性地址技术则利用了椭圆曲线上离散对数的困难问题。在交易时，发送者 A 可以为接收者 B 随机生成一个随机数 r，并根据 B 的地址计算出一个一次性的交易地址，同时将 R=rG（椭圆曲线上的点）随交易信息一起发送给 B。这时，B 收到该笔信息，可以利用自己的私钥以及 R 的信息将一次性地址对应的私钥计算出来，从而隐藏了真实的接收地址。

零知识证明系统即为知识复杂度为零的证明系统。简单来说，这种证明系统所具有的功能有两个，一是证明者能够向验证者证明某个断言（如果它为真），二是在整个证明过程中，

证明者没有向验证者泄露任何额外的信息。因此，从验证者的角度出发，该系统想要实现的功能是，验证者所接收到的零知识证明的效果等同于一个可信第三方直接告诉验证者该断言为真，如图 3 所示。

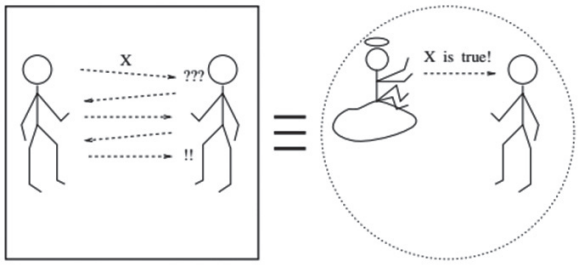

图 3 零知识证明的一个图解

早些的零知识证明系统都是基于交互式证明系统的模型，即证明者和验证者之间必须经过至少一轮的交互。然而随着研究的深入，在 1988 年，Blum 等人指出，通过在证明者和验证者之间共享一个公共字符串，就可以在无交互的情况下实现计算性的零知识证明系统。Zcash 所使用的 zk-snark 即是基于这种公共字符串模型，它允许在证明者和验证者无交互的条件下，通过由第三方预先生成的系统参数（包括证明密钥、验证密钥等信息），来实现零知识证明。

在 Zcash 中，保护交易的数据格式有所变化，一笔资金使用票据 Note=(a_{pk},v, ρ ,r) 来表示。在花费时，发送方只是公布由 Note 单向计算出来的否决值，以及为接收方新生成的票据的承诺值。同时，他还需要提供一个证据，来证明所花费的资金之前已被正确构造，同时他拥有相应的花费权限，并且他为接收者生成的新票据也是按照规则合法构造的。这个证据由 zk-SNARK 方案产生，可以帮助网络上的其他节点在不知道详细交易信息的情况下验证交易的有效性。

同态加密是指，对明文进行某种运算后的结果，等同于对相应密文进行某种运算后再解密的结果。因此，同态加密方案允许对交易金额进行加密，同时通过在密文上进行运算来判断交易的输入输出是否平衡，也即输入输出的金额是否相等。这种同态性思想在门罗币的环形机密交易中得到应用，能够较好地保护交易的金额。

5 以太坊等智能合约平台的相关数据

随着区块链技术的不断发展，比特币基于脚本语言所提供的可编程性给其上应用程序的开发带来了限制，以太坊是第一个图灵完备的分布式智能合约系统。基于此系统，企业和爱好者可以构造各种各样的智能合约应用，如市场预测、供应链等。此时由于区块链上的数据不仅包括普通的交易数据，还包括智能合约相关的输入、存储以及合约本身，因此隐私保护就变得更加困难。

Hawk 率先从理论层面对这个问题进行了解决，它仿照 Zerocash 协议，使用零知识证明来进行构造。该系统由三部分组成，用户、程序员以及代理人。程序员无需具有专门的密码学知识，Hawk 编译器会将程序员编写的程序编译成用户与区块链之间交互的密码协议。代理人不是一个可信第三方，他只需要被信任不会泄露用户的隐私数据，并且当他有恶意行为时，系统会对其进行惩罚并对相关用户进行补偿。

Hawk 在 Zerocash 基本的铸币、消费过程的基础上，加入了合约的计算过程。当用户需要

执行某一计算时，会将加密后的输入数据、金额，以及一个相关证明发给代理人。代理人解密后，按照合约的定义计算输出，并将输出再次加密，同时提供一个该过程被正确计算的证明，最终记录在区块链上，实现资金、存储状态的记录。

6 总结

本文针对区块链上数据如何保护的问题，对现有的解决方案进行了梳理，并对方案中所涉及的密码学知识进行了简单的介绍。区块链技术由于它自身提供的不可篡改、不可伪造、去中心化等优势，近几年吸引了社会各行各界的广泛关注，在金融、能源、供应链等领域也逐渐有了应用案例。然而，区块链技术在发展过程中也暴露出了一些问题，如可扩展性、合约的形式化验证、隐私保护等，这些问题在一定程度上制约着区块链应用的蓬勃发展。同时，业内学者和工程师们也在不断摸索着前进，相信未来这些问题得到进一步的解决之后，区块链技术一定能得到更好的应用和发展。

作者简介

贾音，摩石实验室技术专家，主要研究方向为密码实现和区块链技术应用。

致歉声明

本刊 2018 年第八期刊登的郑美老师所作 《美国网络安全产业人才培养的标准化趋势——< 网络安全人力框架 > 解析》一文中，引用了位华老师和王星老师所作《美国等网络强国网安人才建设举措及启示》一文中部分内容，由于编辑部工作失误，误将文中的引用注释删除，没有刊登出来，因此给双方带来了困扰，对此我们向几位老师致以最诚挚的歉意。

特此声明！

《信息安全与通信保密》编辑部

2018 年