

硕 士 研 究 生 读 书 报 告

徽标

描述已自动生成

题目 区块链与隐私保护

作者姓名 罗中天

作者学号 22151015

指导教师 程学林

学科专业 软件工程

所在学院 软件学院

提交日期 二○二二年一月

摘要

自从2008年比特币出现后，区块链技术如火如荼地发展着。区块链作为去中心化的分布式账本，对金融和其他领域带来深远的影响，但是由于它的账本是全网公开（对于公链）或者是在一个范围内公开的（对于联盟链），所以也面临着严重的隐私保护方面的问题。本文将介绍隐私保护的相关定义，区块链在隐私保护方面的威胁，进一步分析和介绍用于区块链隐私保护的相关技术，主要包括混币，零知识证明等。

关键词：区块链，隐私保护、混币、零知识证明

1. **区块链隐私**

信息系统中隐私指数据拥有者不愿意被披漏的敏感数据或者数据所表征的特性。

区块链技术为了在各个分散的结点之间达成共识以实现数据同步，必须公开信息，例如在像比特币类似的公链上，就是交易信息。而公开信息和隐私保护是相违背的，这就给加密货币和区块链上的隐私安全带来了很大的威胁。根据区块链技术的特点，区块链隐私可以分为两类，即身份隐私和交易隐私。

* 1. **身份隐私（匿名）**

身份隐私指用户身份信息和区块链地址之间的关联关系，即匿名(anonymity)。

在计算机科学中，使用假名并不是真正意识上的匿名，称为p​seudonymity，匿名还需要包括不可链接性(u​nlinkability)。不可链接性是相对于一个具有特定能力的敌手(adversary)而言的，即一个用户与一个系统交互若干次，从敌手的视角来看无法将这些交互的信息和同一个用户关联起来。不可链接性包括三层含义：同一个用户的不同地址难以被关联；同一个用户的不同交易难以被关联；一笔交易的发起者和接收者难以被关联。第三层含义需要说明一下，它不是指写进区块链的单笔交易的发起者和接收者不能被关联，由于所有交易记录都是公开的，所以这也是办不到的，也不是指一笔资金的最终发起者和最终接收者不能被关联，因为这也是可以在公开的交易记录中通过链接的关系进行逆推的，而是指在一个匿名集(anonymity set)中，敌手无法分辨出某个用户的交易信息是哪个，即使它知道该用户的交易在这个匿名集中。显然匿名集越大，匿名性就越好。

以比特币为例，比特币只是p​seudonymity，而不是真正的匿名。在比特币系统中，用户可以任意地创建地址并且用这些地址进行交易，由于钱包的参与，可能每次交易都会使用不同的地址，这就是一种使用假名的行为。但是在经典的比特币系统中，一个用户的多个地址是可能被有能力的敌手给关联起来的，具体链接的技术和手段会在之后展开来介绍。由于比特币系统的交易数据是公开的，任何人都可以查看，所以一旦多个地址被关联起来，该用户的所有交易记录就会变得没有任何秘密可言，用户的隐私安全会受到严重威胁。

* 1. **交易隐私**

交易隐私指区块链中存储的交易记录和交易记录背后的知识。

在早期的区块链数字货币应用中，交易记录通常是公开的，不需要额外的保护措施。但是随着区块链技术被应用到银行等金融领域，交易记录属于重要的敏感数据，需要采取额外措施限制非授权用户的使用。

1. **区块链隐私威胁**

区块链的去中心化和隐私安全存在着天然的矛盾，图1表示一些加密货币和中心化的传统货币在隐私安全方面的优劣。下面将按区块链系统的体系结构从网络层、交易层、应用层三个方面进行介绍。

日程表

描述已自动生成

图 1 加密货币和传统货币在隐私方面的对比

* 1. **网络层**

网络层的主要威胁是恶意节点可以轻易接入网络，监听网络层的通信数据，可以得到节点的IP地址，节点之间的拓扑关系，网络传输信息，通过以上信息攻击者可以将网络层捕获的交易信息和始发节点的IP地址关联，从而对用户身份隐私造成威胁，这种分析方法被称为交易溯源技术。

网络层的隐私安全问题是由底层的P2P网络引起的，可以使用Tor洋葱网络或者mix net等匿名技术来解决。

* 1. **交易层**

交易层的主要威胁是指交易信息的全局账本都是公开的，任何加入区块链网络的节点都可以获得完整的副本，通过分析全局账本中的交易记录和相关的背景知识，攻击者可能会对用户的身份隐私和交易隐私带来威胁。

有一些方法可以通过分析交易信息将多个地址和用户关联起来。如果一笔交易有多个输入，那么这些输入的地址可以被关联起来，在大多数情况下，可以认为这些地址属于同一个用户。维护一个地址集，如果存在一个地址和目前地址集中的某个地址同属于一个交易的输入，那么就把这个地址加入到地址集中，重复上述过程，最终会得到一个关于某个用户的比较完整的地址集。除了利用交易的输入，输出也可以成为突破口，如果可以辨别出一笔交易的找零地址，那么就可以关联更多的地址。在​Fistful of Bitcoins的研究中，他们利用了钱包的习惯，即每次的找零地址都会是一个新的地址，如果一笔交易输出的地址是一个新的地址，那么就认为其为找零地址，再加入到上述的地址集中，得到了图2的结果。

图表, 雷达图

描述已自动生成

图 2 clustering of address

其中圆圈的大小表示流入该地址集的资金数量。接着通过一些背景知识，和服务提供者进行比特币交易等的方式，他们将这些地址集关联的用户和现实世界的身份对应了起来，得到了图3的结果。

图片包含 伞, 大, 华美, 雨

描述已自动生成

图 3 Labeled clusters

上述的方式基于分析区块链系统中的交易网络图，所以把它们统称为t​ransaction graph analysis。

* 1. **应用层**

应用层的主要角色包括使用区块链技术的用户和提供区块链服务的服务商，这两者在处理区块链业务时都有可能带来隐私泄漏的威胁。很多用户可能会将自己的地址发布到论坛或者其他社交网站上，从而地址和身份信息相关联。区块链服务商也可能由于主观或者客观的原因泄漏用户数据和敏感信息，间接导致攻击者能够将区块链全局账本的匿名地址和真实用户相关联。

1. **区块链隐私保护技术**
   1. **混币(Mixing)**

混币，顾名思义，就是多笔交易的输入方将钱先投入到一个中间机构，这时各个输入方的钱就被混在一起，中间机构再将混在一起的钱按照原各笔交易的金额和输出地址构建新的交易，这样敌手就很难依靠区块链的全局账本追踪一个用户的资金流向了。

* + 1. **钱包**

钱包自身带有一些混币的功能，钱包会在线存储你的资金，并且在你需要的时候返回给你相对应数量的资金，一般来说，返回给你的钱和你投入的钱不会是同一笔钱。

但是，将钱包用来混币会有很多问题。第一，并不是所有的钱包都能保证有混币功能，钱包存在混币功能主要是为了工程上的方便。第二，钱包会在内部记录你的资金流向，将你的存款和提现关联起来，这也是钱包安全性和合法性的保证，但是这从隐私保护方面来说，是存在着泄漏风险的。第三，正规的钱包会要求记录你的身份信息，如果这些信息泄漏，会比不使用钱包导致的结果更差。从这些方面来看，钱包和传统货币中的银行类似。

* + 1. **专门的混币服务**

和钱包相对比，它们承诺不记录你的交易信息，也不要求你的身份信息。但是你需要相信它们的承诺，相信它们不会盗取你的资金。不像钱包，它们不存储资金，用户希望混币的过程尽可能地快，这就意味着混币所使用的资金池会比较小。

* + 1. **混币的原则**
       1. **使用多个混币服务**

使用一系列的混币服务，一个接着一个，这样只要其中一个混币服务能够遵守承诺，删除掉你的记录，就能保证没有人可以将你的这笔交易的输入地址和输出地址关联起来了。

图示, 维恩图

描述已自动生成

图 4 多个混币服务

* + - 1. **统一混币中交易的金额**

如果一次混币中各笔交易的金额不同，那么就会大大缩小匿名的集合，敌手甚至可以通过金额直接将输入和输出关联起来，这大大降低了混币的有效性。同时，统一混币中交易的金额，可以使得使用多个混币服务更加方便，不需要额外的拆开与合并的操作。

事实上，对于所有的用户使用一个标准的 chunk size 是不切实际的，如果取的太大，金额比较小的交易就无法参与混币，如果取的太小，金额比较大的交易需要拆成很多份进行多次混币，降低混币的效率，增加混币的成本。且如果金额大小的标准过多，每个标准的匿名集就会比较小，所以在实际操作中，2-3个金额大小的标准比较合适。

* + - 1. **小费应该全收或者不收**

混币服务是一个商业活动，需要收取小费。收取小费最常用的方式是从每笔交易里都抽取相对应的小费，但是这样会违反上述的统一混币中交易金额的原则。所以混币服务使用的方式是对应一笔交易的输入，要么全收，要么不收，通过概率随机的方式，比如收取0.1%的消费，那么就1000次里吞掉一次交易的全部费用。要通过这种方式收取小费，就需要让用户相信认同它的概率随机方式，不会欺骗用户，有一些密码学的手段可以达到这个目的。

* + 1. **去中心化的混币(Coinjoin)**

Coinjoin，用点对点的协议代替中间机构来实现混币。在协议中，多方参与者将各自的交易的输入输出组合在一起，生成一笔交易。在交易中输入和输出出现的位置是随机的，外界无法得知输入和输出的映射关系。交易构建完毕后，每个参与者确认自己的输出地址在交易中且金额正确，确认完毕后，进行签名。

Coinjoin协议的过程可以拆分为5个步骤：

1. 寻找参与者
2. 交换输入输出：使用匿名的通信协议相互交换输入和输出，使得其他参与者也不知道输入和输出的映射关系，以防敌手作为参与者，这里的匿名通信协议可以使用mix-net。
3. 构造交易：交易由其中的一个参与者进行构造，构造完成后内部进行传递。
4. 每个参与者确认完输出后进行确认
5. 广播交易

这种协议容易发生拒绝服务攻击，只要破坏方不签字即可。防止这种攻击的一种方式是往协议中引入付费的机制。

* 1. **zk-SNARK**

前面提到的混币只能在一定程度上缓解而不能根治去中心化引入带来的固有问题，即无论是混币交易还是权益证明，都是需要公开的。零知识证明要解决的问题是：以不透露一个论断的任何信息为前提，向你证明这个论断是对的。

zk-SNARK是“zero knowledge Succinct Non-interactive Argument of Knowledge”的缩写，是一种无交互的零知识证明：

简明（Succinctly）——独立于计算量，证明是恒定的，小尺寸的

非交互性（Non-interactive）——证明只要一经计算就可以在不直接与 prover 交互的前提下使任意数量的 verifier 确信

可论证的知识 (with Argument of Knowledge) —— 对于陈述是正确的这点有不可忽略的概率，即无法构造假证据；并且 prover 知道正确陈述的对应值（即：证据）

* + 1. **原理**

一个需要证明的论述本质上是在一系列约束范围内的解。

协议证明的媒介是多项式，通过一系列的转化，我们可以将关于在一系列的约束条件下解的问题转化为关于多项式的问题。

文本

描述已自动生成

图 5 zk-SNARK问题转化图

下面通过例子来引出协议。

比如现在我们要证明知道一个数

图片包含 图表

描述已自动生成

可以写出约束条件（这里将数拆成二进制的方式可以显著减少逻辑门的数量）：

图片包含 表格

描述已自动生成

协议能处理的约束为a\*b=c的形式，将上述约束表示为这种形式：

图片包含 表格

描述已自动生成

然后转化为R1CS矩阵，R1CS矩阵及3个矩阵图片包含 文本

描述已自动生成，满足

图片包含 公司名称

描述已自动生成

其实就是将变量与其系数分开来，其中

表格

中度可信度描述已自动生成

三个矩阵分别为

日历

中度可信度描述已自动生成图片包含 日历

描述已自动生成图片包含 日历

描述已自动生成

接着选取随机数root，将R1CS转化为QAP，有多少个约束，就选取多少个root，这里为了方便起见，选取文本

描述已自动生成，构造出多项式的点值形式，再差值出系数形式，以矩阵A为例：

文本

描述已自动生成

然后和相应的变量乘起来再累加，就可以得到操作数多项式

图片包含 表格

描述已自动生成

这个操作数多项式对应着约束，将x=1代进去就是第一个约束，将x=2代进去就是第二个约束，等等，也就是说有

文本

中度可信度描述已自动生成

也就是说

图片包含 文本

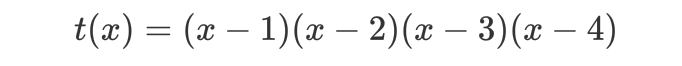
描述已自动生成

有根

文本

描述已自动生成

令



那么就有



根据LPCP的理论，即任意一个d阶的多项式P，都可以通过随机验证多项式在几个点上的取值来确定这个多项式的每一项系数是否满足特定的要求。这里使用的是Schwatz-Zippel 定理，即两个系数不相同的多项式最多只有有限个交点，或者说两个系数不相同的多项式不可能有共享的连续段，证明的方法是将将一个n阶的多项式和一个m阶的多项式连立，根据方程解的性质，最多只有max(n,m)个解。所以多项式在任意点上的计算结果都可以看作其唯一身份的表示。

所以verifier对上述多项式任意点的采样，如果prover都可以用其已知的变量的值代进去满足等式成立，就可以认为prover知道在原约束条件下的解，即prover证明成功。

假设约束的数量为d，变量的数量为n，那么我们可以根据上述的分析生成一个协议：

文本, 信件

描述已自动生成

文本

描述已自动生成

文本, 信件

描述已自动生成

对于上述协议一些细节的解释：

一些文字和图片的手机截图

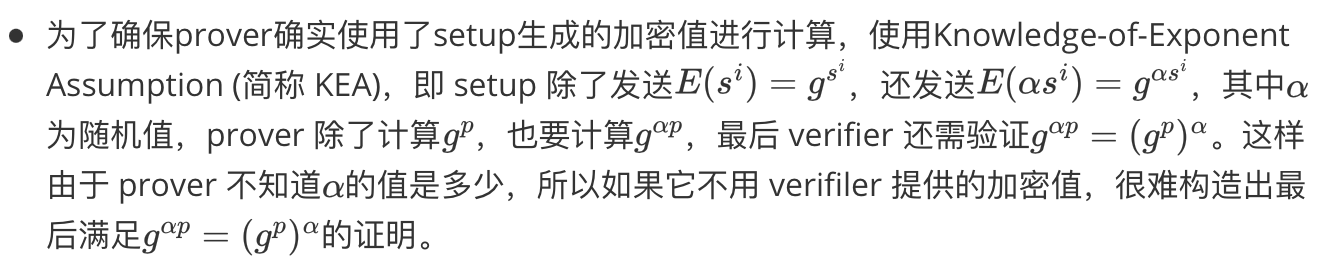
描述已自动生成

文本

描述已自动生成

文本, 信件

描述已自动生成



文本

描述已自动生成

后面对于该版协议的一些问题的改进由于篇幅原因不再介绍，最终生成的zk-SNARK协议为：

**图形用户界面, 文本

描述已自动生成**

**文本

描述已自动生成**

**文本, 信件

描述已自动生成**

1. **参考文献**