# 前言

本周主要是在补充比特币的知识体系，再次通过公开课和查阅资料把共识协议和系统实现整理出来。这周也听了吕老师的讲解，也了解比特币和以太坊的之间的关系，我也在网上搜索了一些以太坊的介绍，也是根据比特币的知识相比较，进而加深比特币的知识，具体的链接我会在附录上展示。这周因为比特币的知识还有好多，我还是没能整理完，这其中的知识量，如果用语言整理，其工作量确实超过了我的想象，下周我会尽可能的把比特币的知识点整理完。

# 1.比特币有哪些组成

## 1.1比特币有哪些组成?

如果说把比特币当成真正的货币，那比特币的组成有哪些成分呢？这些成分有涉及到什么知识呢？《Master Bitcoin》提到了比特币的组成：

比特币包括：

* 一个去中心化的P2P网络 （比特币协议）
* 一个公共交易账本 （区块链）
* 一套规则，用于独立的交易确认和货币发行 （共识规则）
* 一个机制，用于对有效区块链实现全球去中性化共识 （工作量证明算法）

简化来说，1.协议。2.区块链。3.共识规则。4.工作量证明。

其中所涉及到的知识大体包括：**1.密码学基础。2.比特币的数据结构。3.共识协议4.系统实现。5。比特币网络。6.挖矿算法和难度调整。7.比特币脚本。8.软分叉和硬分叉。9.匿名和隐私保护作用**。其中每一部分知识我会在下面慢慢展开。

# 2.比特币的知识

## 2.1密码学基础

由于比特币属于加密货币(crypto-currency)，当然会涉及到密码学的知识，比特币属于一种开源信息的货币，其中的交易记录，交易金额，交易信息都面向全网开放。

比特币中加密涉及到两个内容：1.hash(哈希)。2.签名

1.密码学中用到的哈希函数被称为cryptographic hash function: 它有两个重要的性质:

1. collision(这里指哈希碰撞) resistance
2. hiding 哈希函数的计算过程是单向的，不可逆的。

除了密码学中要求的这两个性质，比特币中用到的哈希函数还有第三个性质：

1. puzzle friendly 指哈希值的预算事先是不可预测的。

比特币中用的哈希函数叫做**SHA-256(secure hash algorithm)**以上三个性质它都是满足的。这里这三个性质其实有详细解释，但是考虑到字数，我会放到附录的链接中。

2.签名又涉及到哪些密码学呢？在比特币系统中创建账户:在本地创立一个公私钥匙对(public key ,private key)，这就是一个账户。其中公私钥匙对是来自于非对称的加密技术(asymmetric encryption algorithm)。非对称密钥是用一对密钥而不是一个，加密用公钥，解密用私钥，加密和解密用的都是接收方的公钥和私钥。公钥是不用保密的，私钥要保密但是私钥只要保存在本地就行，不用传给对方。

通俗来讲，公钥相当于银行账号，别人转账只要知道公钥就行，私钥相当于账户密码，知道私钥可以把账户上钱转走。总体上来讲就是公钥和私钥都是用来签名的。然而这种签名机制就是采用密码学中的非对称加密。

## 2.2比特币的数据结构

比特币所涉及到的数据结构包括：

1. 哈希指针(Hash pointer)
2. 默克尔树(Markle Tree)

1.其实比特币最基本的结构就是区块链，区块链就是一个一个组成的链表，但区块链与普通链表不同的是：区块链采用了哈希指针代替了普通指针。这样的特点是，每个区块都根据自己的区块内容生成自己的哈希值，并且每个区块都要保存前一个区块的哈希值。其实就是有了哈希指针就可以保证该区块的唯一性，并且保证整个区块一旦写入就无法进行篡改。

2.而对于默克尔树Markle Tree，该数据结构类似于我们之前学到的二叉树一样。

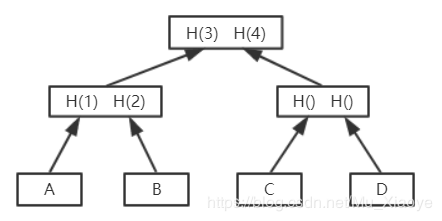


图2-1Markle Tree的示例

这种结构的好处：只要记住哈希值，就能检测出对书中的任何部位的修改。比特币当中各区块之间用哈希指针连接在一起，每个区块所包含的交易组织成一个merkle tree的形式，最下面一行data blocks每个区块实际上是一个交易，每个区块分为两部分，分别是块头和块身(block header ,block body)。块头里面有根哈希值，每个区块所包含的所有交易组成的merkle tree的根哈希值存在于区块的块头里面，但是，块头里没有交易的具体内容，只有一个根哈希值，块身里面是有交易的列表的。利用下面的图可以理解merkle tree的应用场景：

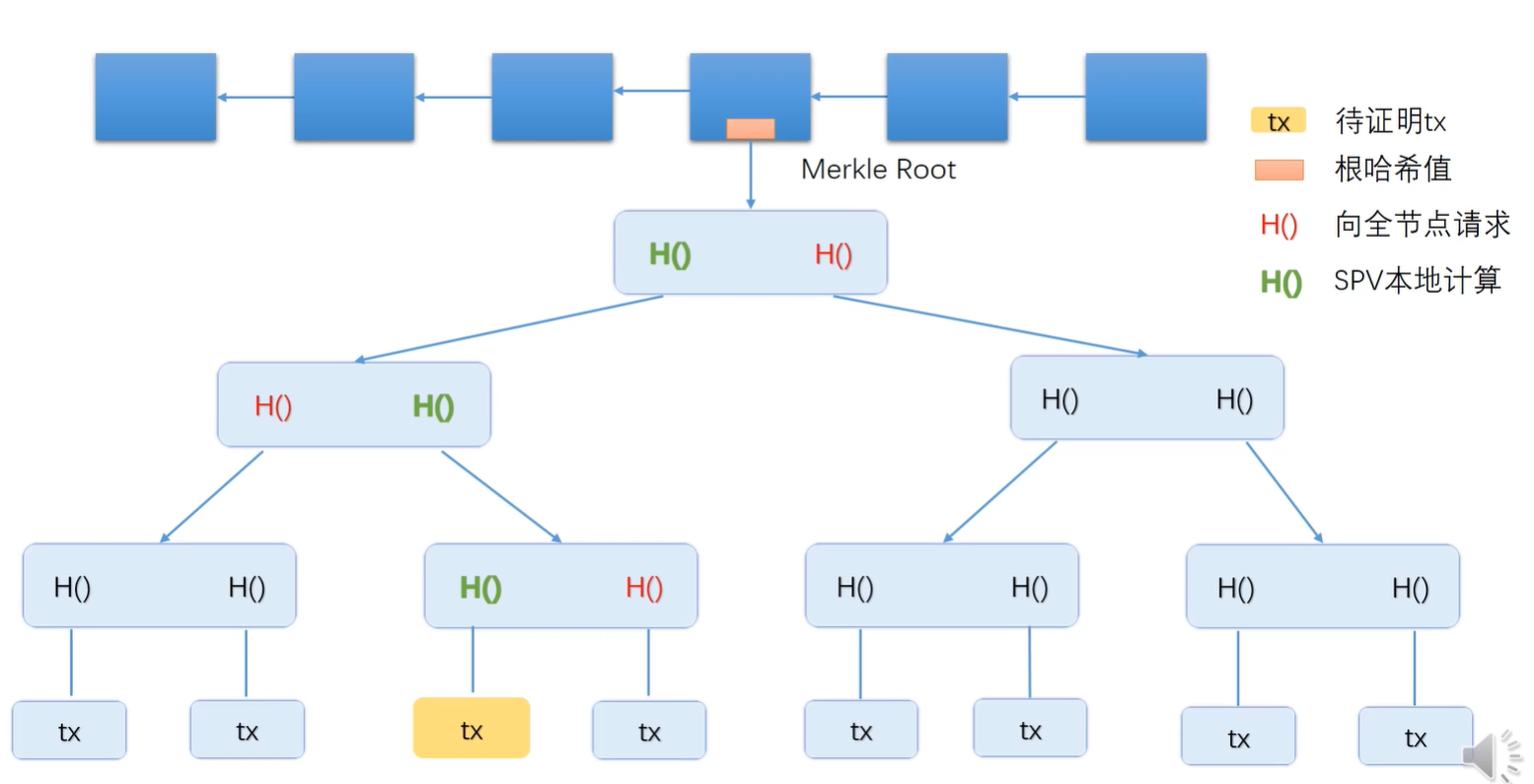


图2-2Markle Tree的应用场景

## 2.3比特币的协议

### 2.3.1比特币的发行和交易

去中心化货币要解决两个问题:①数字货币的发行②怎么验证交易的有效性，防止double spending attack。在传统中心化货币体系中，这些问题我们可以交给第三方机构（如：央行）。当引入去中心化思想后，系统中节点平等，交易不通过第三方，那么货币发行权的分配必然是一个需要解决的问题。而双花攻击，顾名思义就是一个货币花了两次。对用户来说，可以将同一货币花费两次。因为数字货币本身为带有签名的数据文件，是可以进行复制的。

当然作为去中心化的货币，比特币是要克服这些问题的，比特币采取：1.比特币的发行是由挖矿决定的来解决发行问题。2.依靠区块链的数据结构来解决双花问题。

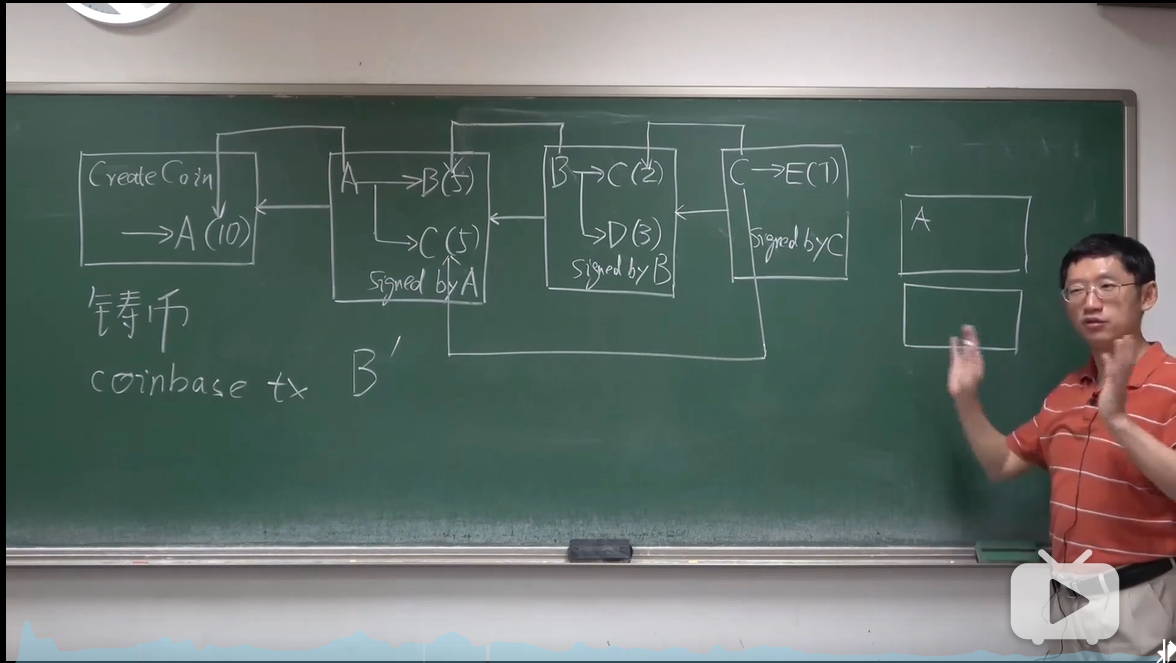


图2-3 Bitcoin的交易情况

从图中我们能够看到初始的币来源于矿工挖块得到的区块奖励，也就是铸币，是由系统分配的，也就是上面我们所提到的：比特币的发行是由挖矿决定的来解决发行问题。然后其中当A和B进行交易时，系统回去寻找A这个币的来源，通过哈希指针去寻找。同时当A与B,C交易后，B，C在进行交易转账时，系统也会根据哈希指针去寻找是谁给了B和C钱，依次类推。总之，当我们在比特币系统中进行交易时，系统会查看我们当前交易之前的记录，才能完成我们现在的交易，来保证我们这些钱是真的存在，而且还没有花出去，系统才能执行我们所发出的交易。这样就保证了防止双花攻击，不能同一份钱花两次。

### 2.3.2比特币共识协议(consensus in BitCoin)

区块链里的内容是如何写到区块链里面的呢:每个节点，每个账户都可以发布交易，交易是广播给所有节点的。有些交易是合法的，有些是非法的。谁来决定哪些交易应该被写入下一个区块中呢?按照什么顺序写呢?如果每个节点自己决定可以吗?如果每个人在本地维护一个区块链，那区块链的统一性得不到保证，而账本的内容是要取得分布式的共识(distributed consensus)。于是就引来了比特币共识协议，在这里我也了解到，一个去中心化的系统必然是分布式的，但分布式系统不一定是一个去中心化的。

概括来讲，比特币中的共识协议需要解决的最大的问题就是，有些节点可能是有恶意的。我们假设系统中大多节点是好的，如何取得共识就很重要了。

我们通常在生活中一般采取的共识都是投票，但是这种方式往往带来很多问题，如果比特币种按照账户的数目进行投票的话，如果每个人可以创建多个用户，也就可以投多张票，这中方法就会显得很危险了,很有可能导致51%攻击，俗称女巫攻击。

比特币系统巧妙的解决了这个问题，虽然仍然是投票，但并非简单的根据账户数目，而是采用**计算力进行投票，**也就是采用POW(Proof Of Work)机制，每个节点都可以在本地组装出一个候选区块，把它认为合法的交易放在里面，然后开始尝试各种nonce值(占4 byte)，看哪一个能满足不等式H(block header)≤target的要求。如果某个节点找到了符合要求的nonce，它就获得了**记账权**。从而可以将区块发布到系统中。其他节点受到区块后，验证区块合法性，如果系统中绝大多数节点验证通过，则接收该区块为最新的区块并加入到区块链中。

所谓的记账权，就是往比特币账本里写入下一个区块的权利。只有找到这个nonce，获得记账权的节点才有权利发布下一个区块。其他节点收到这个区块之后，要验证这个区块的合法性，同时也要验证区块中的交易是否也合法。如果有一项不符合要求，这个区块就是不能被接受的。如果所有条件都符合，也不一定接受。

那所有条件都符合，为什么还不一定接受呢？这就引出了区块分叉的问题，如图

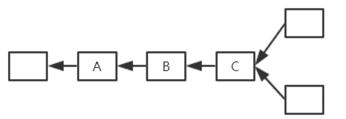


图2-4 区块的分叉

图中的两个区块都符合上述的判断，但由于是同一时间段发布，导致区块链产生了分叉，这种分叉情况一般分两种：一种是确实两个恰好同时挖出区块来，一种就是属于分叉攻击了。前者发生分叉的情况下，会暂时保存分叉情况，在缺省情况下，节点接收最先听到的区块，该节点会沿着该区块继续延续。但随着时间延续，必然有一个链胜出，由此保证了区块链的一致性。（被扔掉的区块称为“孤儿区块”），但区块链只承认最长合法链，随着时间推移，必然存在某一条链变成最长合法链。这样，也就会导致合法区块也会被拒绝。而后者就相对比较严重是属于分叉攻击，是一些非法矿工想要从中获取利益，分叉攻击如图：

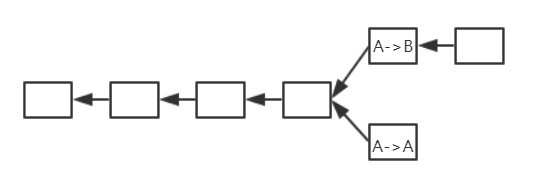


图2-4 分叉攻击

A用户对上面的A转账给B的记录回滚，从而非法获取利益。在两条链上，发现交易都合法。这是一个典型的双花攻击。A给B转账后，用分叉攻击将钱又转回来，覆盖掉原来的记录。

在比特币系统中，这种情况实际上很难发生。因为大多数矿工认可的是最长的合法链，会沿着上面的链继续挖下去。而A这个攻击者要想回退记录，就必须使得下面的链变得比上面的链还长。理论上来说，攻击者需要达到整个系统中51%的计算力，才能使得这种攻击成功。但要达到51%的计算力这种能力去办坏事，显然所涉及到的成本是在太大了。

## 2.4比特币的系统的实现

### 2.4.1UTXO

区块链是去中心化的账本，比特币使用的是基于交易的这种账本模式(transaction【交易】-based ledger【账本】)。系统当中并不会显示每个账户有多少钱。

比特币系统的全节点要维护一个叫UTXO (unspent transaction output)(还没有被花出去的交易的输出)的数据结构。区块链上有很多交易，有些交易的输出可能已经被花掉，有些还没有被花掉。所有没有被花掉的输出的集合就叫做UTXO。

一个交易可能有多个输出。假如A给B 5个比特币，B花掉了。A也给了C 3个比特币，C没有花掉。这时5个比特币就不算UTXO，而3个比特币算。UTXO集合当中的每个元素要给出产生输出的交易的哈希值，以及它在这个交易里是第几个输出。这两个信息就可以定位到UTXO中的输出。

那要UTXO集合有什么作用？为了检测上文的double spending(双花)。即检测新发布的交易是否合法。因此全节点要在内存中维护UTXO这样一个数据结构，以便快速检测double spending。

每个交易要消耗掉一部分输出，也会产生新的输出。还看上面的例子，B花掉的5个比特币虽然不在UTXO里面，但如果他转账给D，而D没有花掉，那么这5个比特币又要保存在UTXO里面。如果D始终不花，那么这个信息要永久保存在UTXO里面。有可能是不想花，也有可能是把密钥丢了。每个交易可以有多个输入，也可以有多个输出，所有输入金额之和要等于输出金额之和。即total inputs=total outputs。因此一个交易可能来自多个地址，可能有多个签名。有些交易total inputs略微大于total outputs。假如输入1比特币，输出0.99比特币，另外0.01比特币作为交易费给获得记账权发布区块的节点。

### 2.4.2比特币的激励机制

可能我们看到上面，每个节点需要提供能算力和电力成本，要投入大量的资金去挖矿，就是为了竞争记账权，为什么呢？

比特币系统设计之初便考虑到了这个问题，那就是引入激励机制。比特币通过设置**出块奖励**来解决该问题，一个获得合法区块的节点，可以在区块中加入一个特殊交易（铸币交易）。事实上，这种方式也是唯一一个产生新比特币的途径。比特币系统设计规定，起初每个区块可以获得50个比特币，但之后每隔21万个区块，奖励减半。

区块奖励也不能完全作为挖矿的奖励，发布区块的节点为什么一定要把你的交易打包在区块呢?他们还要验证你的交易的合法性，如果交易较多占用的带宽会比较大，网络传播速度也会更慢。所以这时只有区块奖励是不够的。

因此比特币系统设计了第二个激励机制:交易费(transaction fee)。也就是你把我的交易打包在区块里，我给你一些小费。交易费一般很小，也有一些简单的交易没有交易费。

# 3.本周总结

这周主要还是看比特币的知识，同时也开始接触了以太坊方面。而已也在7月24号听了吕老师对区块链方面的讲解，确实发现自己要学的东西还有很多。今天所写的读书报告也没有完全写完，因为当我真的去写这个比特币的知识点的时候，我发现这其中的知识点要比我想象的多的多，上面确实有一部分是从网上借鉴过来了，但大多都转换成我自己的语言进行组织起来，我也尽可能的把里面的内容写的清楚一点，但总感觉写不完。其实我现在已经把比特币的大体知识都过了一遍，但是要想完全掌握其中的细节，我感觉还是需要很大努力，毕竟人家中本聪花了好几年的成本，我会慢慢把比特币的知识点补充清楚的，下周我就打算再看一遍公开课，争取把这个知识点贯通，以达到了解区块链的行情。

# 附录

**1.北京大学肖臻老师《区块链技术与应用》公开课(来自于B站):** <https://www.bilibili.com/video/BV1Vt411X7JF?p=1>

**2. 北京大学肖臻老师《区块链技术与应用》公开课系列笔记4:**

<https://blog.csdn.net/Mu_Xiaoye/article/details/104334253>

**3. 北京大学肖臻老师《区块链技术与应用》公开课系列笔记5:**

**:** <https://blog.csdn.net/Mu_Xiaoye/article/details/104342513>

**4. 知乎的以太坊介绍:** <https://zhuanlan.zhihu.com/p/30922425>

**5. 以太坊学习资料汇总整理  ：**

<https://www.chainnode.com/post/377756>

**6. 个人的github的开源库：** <https://github.com/Hello-Afa/Graduate_career.git>