现代计算机网络

P2P技术新进展: BitCoin和区块链

为什么要学习区块链?

- 比特币和区块链概念火爆
- P2P网络之上能否实现更加复杂的应用?
- P2P应用目前主要特点是分治、复制,还没有实现共同维护一个 复杂数据结构。这个复杂数据结构不是指的P2P本身数据结构, 而是应用的结果。
- 也就是P2P这种去中心化的网络是否能够帮助人们实现更加复杂的协作?

P2P技术新进展:BitCoin和区块链

例如"拜占庭将军问题":

- □ 拜占庭将军问题(Byzantine Generals Problem),是由莱斯利·兰波特 (Leslie Lamport)在其论文《分布式系统一致性问题(Distributed Consensus)》中提出的分布式对等网络通信容错问题。
- □ 拜占庭帝国想要进攻一个城市,为此派出了10个将军率领10支军队,这个城市足以抵御5支常规拜占庭军队的同时袭击。这10支军队不能集合在一起单点突破,必须在分开的包围状态下同时攻击,至少6支军队同时袭击才能攻下敌国。10支军队分散在敌国的四周,依靠通信兵相互通信来协商进攻进攻时间。
- □ 这里的问题是,将军中可能会有叛徒,忠诚的将军希望达成命令的一致(约 定某个时间一起进攻),但背叛的将军会给不同的将军发送不同的进攻时间 阻挠忠诚的将军达成命令上的一致。

P2P技术新进展:BitCoin和区块链

Leslie Lamport在自己的论文中提到两种解决办法:

- □ 一种是口头消息,超过三分之一将军是叛徒就无法达成一致
- □ 一种是书面协议,超过一半就无法达成一致
- □ 这两种办法都需要消耗大量的通信带宽,将军数量一多就无法扩展。

Bitcoin Whitepaper – 2008.10.31

Bitcoin: A Peer-to-Peer Electronic Cash System

Satoshi Nakamoto satoshin@gmx.com www.bitcoin.org

Abstract. A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network.

The network timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work. The longest chain not only serves as proof of the sequence of events witnessed, but proof that it came from the largest pool of CPU power. As long as a majority of CPU power is controlled by nodes that are not cooperating to attack the network, they'll generate the longest chain and outpace attackers. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest

P2P技术新进展: BitCoin

- 中本聪发明的一种P2P形式的虚拟货币。点对点的传输意味着一个去中心化的发行系统。
- Bitcoin: A Peer-to-Peer Electronic Cash System
- "We propose a solution to the double-spending problem using a peer-to-peer network"

P2P技术新进展: BitCoin

- 比特币不依靠特定货币机构发行,它通过特定算法的大量 计算产生,比特币经济使用整个P2P网络中众多节点构成 的分布式数据库来确认并记录所有的交易行为。这个数据 库称为区块链。
- 比特币与其他虚拟货币另一个不同:总数量将被永久限制 在2100万个之内。



虚拟货币历史

在Satoshi Nakamoto之前,有许多人也曾试图创造虚拟货币,但由于技术、监管以及意识形态等方面的原因,均以失败告终。

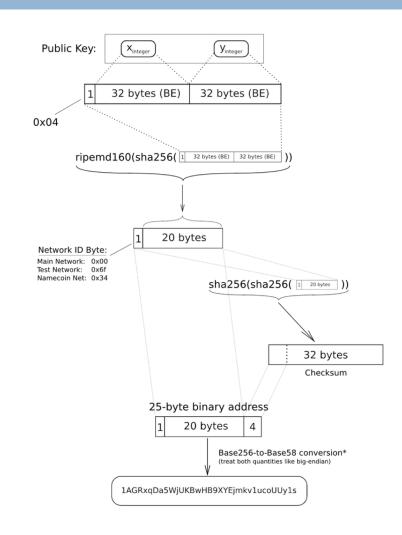
- 1996 诞生的E-gold由一位名叫Douglas Jackson的肿瘤学家出资打造 ,至2009年已有超过500万的注册用户。E-gold将一堆黄金放在金库 中,并发行对应价值的数字现金
- 2006 年诞生的Liberty Reserve是人类在创建集中匿名制的转账服务 上的一次失败的尝试。用户在平台注册之后,可以不经任何认证就给 任意用户转账。

这些虚拟货币问题:为了解决将同一个虚拟货币多次同时使用问题,都需要到一个服务器开设账户,中心数据库记录哪个账户拥有多少货币,交易的时候服务器负责记录,服务器作为可信第三方保障虚拟货币不被双花。但是服务器容易被监管和下线。

直到Satoshi Nakamoto加入了P2P技术,不再依赖第三方,才得到初步成功

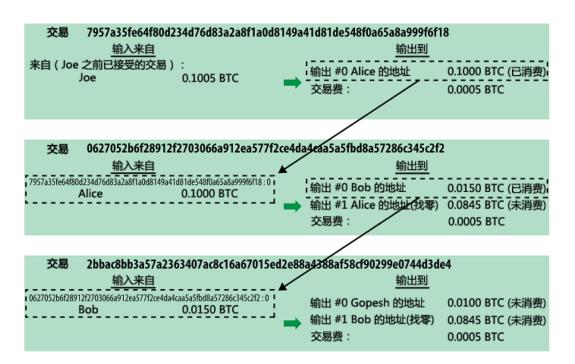
非对称加密:

- ◆ 私钥和公钥,公钥加密只有私钥可以解 密,反之也成立。
- ◆ Bitcoin使用椭圆曲线算法,公钥与私钥是相对应的,一把私钥可以推出唯一的公钥,但公钥却无法推导出私钥。
- ◆ Bitcoin的一个账户就是一个私钥256bit ,拥有私钥就拥有与之绑定的比特币。
- ◆ Bitcoin地址:因为公钥太长了(130字符 或66字符),而且为了安全,地址从公钥推导,长度为25字节。地址最前面加了字符"1",地址末尾添加了4个字节校验位。
- ◆ 交易的时候向Bitcoin地址发送比特币, 只有拥有私钥的人可以领取。
- ◆ 私钥随机产生,无法暴力破解。



比特币交易:

- ◆ 交易是将钱从交易输入移至输出。输入是指钱币的来源,通常是之前 一笔交易的输出。交易的输出则是通过关联一个密钥的方式将钱赋予 一个新的所有者。
- ◆ 所以比特币所有的交易都可以连接起来,查看每一笔资金的流向,换 句话说,每一个比特币都清晰的记录了来源。



比特币交易的一个例子:

- vin即input中对应一个以前 没有花出去的输出,简称为 UTXO(Unspent Transaction Output)。UTXO的txid指明 上次交易的ID, sequence表 明该UTXO的序号。Scriptsig 是解锁脚本。
- 因为一个交易可能有多个out ,vout表示是哪一个
- 比特币的钱包中没有币,只 有私钥和私钥对应的UTX0

```
"version": 1.
  "locktime": 0,
  "vin": [
      "txid":
"7957a35fe64f80d234d76d83a2a8f1a0d8149a41d81de548f0a65a8a999f6f18".
      "vout": 0.
      "scriptSig" :
"3045022100884d142d86652a3f47ba4746ec719bbfbd040a570b1deccbb6498c75c4ae24cb02204
b9f039ff08df09cbe9f6addac960298cad530a863ea8f53982c09db8f6e3813[ALL]
0484ecc0d46f1918b30928fa0e4ed99f16a0fb4fde0735e7ade8416ab9fe423cc5412336376789d1
72787ec3457eee41c04f4938de5cc17b4a10fa336a8d752adf".
      "sequence": 4294967295
 "vout": [
       "value": 0.01500000.
       "scriptPubKey": "OP_DUP OP_HASH160
ab68025513c3dbd2f7b92a94e0581f5d50f654e7 OP_EQUALVERIFY OP_CHECKSIG"
       "value": 0.08450000.
       "scriptPubKey": "OP_DUP OP_HASH160
 7f9b1a7fb68d60c536c2fd8aeaa53a8f3cc025a8 OP_EQUALVERIFY OP_CHECKSIG",
```

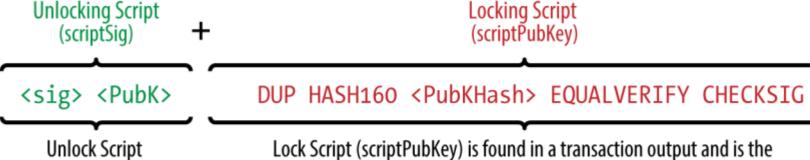
比特币交易的一个例子:

- vout即输出,是把vin中 的比特币发送到多个地址 中
- 右图vout的每个元素对应 一个输出地址,value表 示数量,scriptPubKey表 示加锁脚本,加锁脚本验 证返回True才能使用对应 的BTC
- 一个输出就是一个UTX0
- 比特币的钱包中没有币 , 只有私钥和私钥对应的 UTX0

```
"version": 1,
  "locktime": 0.
  "vin": [
"7957a35fe64f80d234d76d83a2a8f1a0d8149a41d81de548f0a65a8a999f6f18",
      "vout": 0,
      "scriptSig" :
"3045022100884d142d86652a3f47ba4746ec719bbfbd040a570b1deccbb6498c75c4ae24cb02204
b9f039ff08df09cbe9f6addac960298cad530a863ea8f53982c09db8f6e3813[ALL]
0484ecc0d46f1918b30928fa0e4ed99f16a0fb4fde0735e7ade8416ab9fe423cc5412336376789d1
72787ec3457eee41c04f4938de5cc17b4a10fa336a8d752adf",
      "sequence": 4294967295
  "vout": [
       "value": 0.01500000,
       "scriptPubKey": "OP DUP OP HASH160
 ab68025513c3dbd2f7b92a94e0581f5d50f654e7 OP EQUALVERIFY OP CHECKSIG"
     },
       "value": 0.08450000.
       "scriptPubKey": "OP DUP OP HASH160
 7f9b1a7fb68d60c536c2fd8aeaa53a8f3cc025a8 OP_EQUALVERIFY OP_CHECKSIG",
```

比特币交易:

● 解锁脚本和上个交易vout的加锁脚本合并在一起,如果能够通过验证 ,就可以使用上一个交易的vout,即可以使用锁定的比特币



Unlock Script (scriptSig) is provided by the user to resolve the encumbrance ock Script (scriptPubKey) is found in a transaction output and is the encumbrance that must be fulfilled to spend the output

- 解锁脚本中用私钥对交易进行签名<sig>,并提供自己的公钥<PubK>
- 加锁脚本中<EQUALVERIFY>验证<PubK>经过Hash后等于<PubKHash>(BTC地址); CHECKSIG检查私钥签名的<sig>可以用<PubK>验证
- 通过这两步,可以确定一定是私钥拥有者使用了对应地址中的比特币

比特币交易中既然交易的加锁和解锁是两个脚本,那么可以衍生出 非常灵活的变化:

- Time Lock可以约定一段时间后才能使用某些比特币
- Scripts with Flow Control (Conditional Clauses)符合某些条件才能使用,否则不能使用
- Multi-signatures,多个人签名交易才能使用。 2-of-3Multi-signatures是3个人中任意两个人签名才能使用
- 继续发展成为:可编程的货币,智能合约

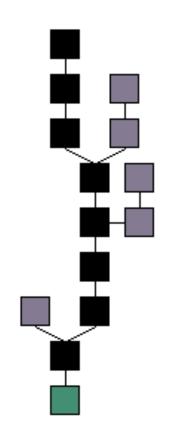
比特币区块:

- 多个比特币交易会打包在一个区块中,每个区块大小是1Mbyte
- 每个交易都会有一个唯一的ID,在区块的tx数组中
- 每个交易vin中包含的比特币数量减去vout中的比特币数量就是交易费

```
"size" : 43560.
   "version" : 2,
   "previousblockhash" :
       "00000000000000027e7ba6fe7bad39faf3b5a83daed765f05f7d1b71a1632249".
   "merkleroot" :
       "5e049f4030e0ab2debb92378f53c0a6e09548aea083f3ab25e1d94ea1155e29d".
   "time" : 1388185038.
   "difficulty": 1180923195.25802612.
   "nonce": 4215469401.
   "tx" : [
       "257e7497fb8bc68421eb2c7b699dbab234831600e7352f0d9e6522c7cf3f6c77".
#[... many more transactions omitted ...]
       "05cfd38f6ae6aa83674cc99e4d75a1458c165b7ab84725eda41d018a09176634"
```

比特币区块链:

- 比特币的区块链接为一个单链表
- 每个区块中保存了前一个区块的hash值,因此任何 人想要篡改链中的一个区块,会导致破坏整个区块 链,被P2P网络的其他节点发现。
- 比特币网络中的矿工负责打包区块,任何人都可以 成为矿工。
- 如果不同的矿工如右图在不同链上打包区块怎么办 ? 全网承认的是最长链(工作量最大)



比特币区块链:

- 矿工都可以竞争打包过程,关键是看谁先计算出来一个合法的nonce,让当前区块hash 值前面有规定个数的0(比如5个0)
- 0的个数由当前全网hash能力决定,每隔一段时间调整,大概10分钟能够计算成功。
- 打包成功的矿工,交易费归他所有

```
"size" : 43560,
   "version" : 2.
   "previousblockhash" :
       "000000000000000<mark>2</mark>7e7ba6fe7bad39faf3b5a83daed765f05f7d1b71a1632249".
   "merkleroot" :
       "5e049f4030e0ab2debb92378f53c0a6e09548aea083f3ab25e1d94ea1155e29d",
   "time" : 1388185038.
   "difficulty": 1180923195.25802612,
   "nonce": 4215469401.
   "tx" : [
       "257e7497fb8bc68421eb2c7b699dbab234831600e7352f0d9e6522c7cf3f6c77",
#[... many more transactions omitted ...]
       "05cfd38f6ae6aa83674cc99e4d75a1458c165b7ab84725eda41d018a09176634"
```

比特币交易问题: 区块奖励, 每个区块有新产生的比特币

- ◆ 为了鼓励挖矿,即哪个节点记录了交易的区块信息,哪个节点就获得了新产生的比特币。
- ◆ 挖矿挖到的比特币是一种特殊的交易(coinbase transaction,), coinbase域对应的为 一个任意字符串, 矿工(1MxTkeEP2PmHSMze5tUZ1hAV3YTKu2Gh1N)可以随便定义这个字符串(由于nonce只有32bit, 所以挖矿需要修改这个字符串达到要求):

```
"hex" :
f03443b0403858402062f503253482fffffffff0110c08d9500000000232102aa970c592640d19de03ff
6f329d6fd2eecb023263b9ba5d1b81c29b523da8b21ac000000000".
   "txid": "d5ada064c6417ca25c4308bd158c34b77e1c0eca2a73cda16c737e7424afba2f",
   "version" : 1.
   "locktime" : 0.
   "vin" : [
           "coinbase" : "03443b0403858402062f503253482f".
           "sequence": 4294967295
    "vout" : [
           "value" : 25.09094928.
           "n" : 0.
           "scriptPubKey" : {
              "asm" :
"02aa970c592640d19de03ff6f329d6fd2eecb023263b9ba5d1b81c29b523da8b210P CHECKSIG".
"2102aa970c592640d19de03ff6f329d6fd2eecb023263b9ba5d1b81c29b523da8b21ac",
              "reqSigs" : 1,
              "type" : "pubkey",
              "addresses" : [
                  "1MxTkeEP2PmHSMze5tUZ1hAV3YTKu2Gh1N'
   1
```

比特币区块链问题:区块串成一个链条,那么链条头部最开始的区块来自哪儿?来自中本聪手工在程序中定义的创世块:

- ◆ 创世块中只有一个交易, ID是 4a5e1e4baab89f3a32518a88c31bc87f618f76673e2cc77ab2127b7afdeda33b
- ◆ 这个块是中本聪挖到的,其CoinBase是一句有深刻含义的话:

CoinBase

04ffff001d0104455468652054696d65732030332f4a616e2f32303039204368616e63656c6c6f72206f6e206272696e6b206f66207365636f6e64206261696c6f757420666f722062616e6b73 (decoded)

Output Scripts

PUSHDATA(65)[04678afdb0fe5548271967f1a67130b7105cd6a828e03909a67962e0ea1f61deb649f6bc3f4cef38c4f35504e51ec112de5c384df7ba0b8d578a4c702b6bf11d5f] CHECKSIG

区块奖励:

- 矿工通过创造一个新区块得到的比特币数量大约每四年(或准确 说是每210,000个块)减少一半。
- 开始时为2009年1月每个区块奖励50个比特币,然后到2012年11月 减半为每个区块奖励25个比特币,之后在2016年的某个时刻再次 减半为每个新区块奖励12.5个比特币。
- 基于这个公式,比特币挖矿奖励以指数方式递减,直到2140年。届时所有的比特币(20,999,999.98)全部发行完毕。换句话说在2140年之后,不会再有新的比特币产生。
- 中本聪一个人挖到了大概98万个比特币

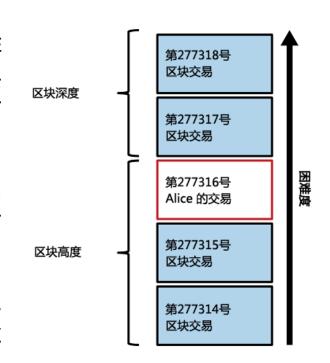
区块奖励:

● 一个区块的奖励是非常惊人的: (大于1M的区块使用了Segwit技术)

高度	播报方	数量	Stripped Size(B)	大小(B)	Weight	平均交易费	块收益
501,386	(i) 58COIN	1,671	972,907	1,074,036	3,992,757	0.00090966	12.5 + 3.63206418 BTC
501,385	BTC.TOP	2,207	894,821	1,308,328	3,992,791	0.00131796	12.5 + 5.26234572 BTC
501,384	⊚ ViaBTC	2,362	948,815	1,146,705	3,993,150	0.00199972	12.5 + 7.98519784 BTC
501,383		1,202	949,607	1,144,383	3,993,204	0.00074979	12.5 + 2.99404692 BTC
501,382	BTC.TOP	2,184	940,133	1,172,803	3,993,202	0.00151977	12.5 + 6.06876758 BTC
501,381	⊚ ViaBTC	531	985,737	1,035,555	3,992,766	0.00035687	12.5 + 1.42491629 BTC
501,380		1,077	965,138	1,097,739	3,993,153	0.00094641	12.5 + 3.77915655 BTC
501,379		255	989,698	1,023,947	3,993,041	0.00039183	12.5 + 1.56460305 BTC
501,378		261	995,646	1,006,061	3,992,999	0.00038308	12.5 + 1.52963143 BTC
501,377	⊘ ViaBTC	2,268	969,005	1,085,994	3,993,009	0.00075471	12.5 + 3.01356664 BTC
501,376	(a) ViaBTC	2,131	939,668	1,174,217	3,993,221	0.00195020	12.5 + 7.78758510 BTC

区块链总结:

- 区块链是所有比特币的交易的一份按照时间排序的公开记录。区块链在所有比特币用户中共享。它被用来验证比特币地址的有效额度和防止双重支付。
- 平均大约每过10分钟,一个新的包含交易的区块会通过挖矿被追加到区块链里。
- 比特币挖矿就是搜集10分钟内产生的交易,找到hash到特定模式对应的随机数。Hash计算能力越强,产生正确hash模式的速度越快
- 区块链被保存到全球的1万多个比特币P2P网络的节点中(越多的节点保存,就越容易验证)

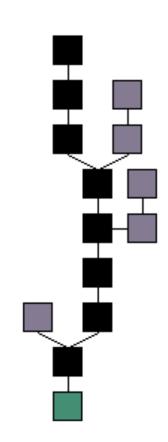


区块链为什么会防止双花交易?

- 区块链是一个单链表。
- 每个交易的输入来自以前交易的输出,如果有人想一个币使用两次,那 么会出现同一个交易输出,被打包到两个新的交易中
- 这个非常容易检测到的,因为单链表,所以UTXO集合在每个区块后一定 对全网所有节点是一致的,那么非常容易检测出一个UTXO是不是被使用 过

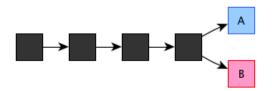
区块链为什么会防止双花交易?

- 同样,想要新产生一个比特币也是不可能的, 因为coinbase交易是固定的。
- 有没有可能一个矿工在一条链上打包一个交易使用了一个UTXO,然后转而到另外一个链上再次打包另外一个交易,再次使用同一个UTXO?

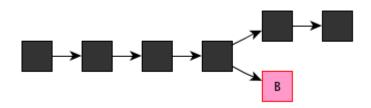


BitCoin原理-区块链竞争

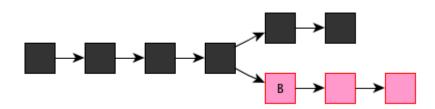
各个矿工是并行工作的,因此完全可能出现这样的情况,有两个合法的区块



此时如果有一个新的区块是基于A的,那么这个主干就延续下去

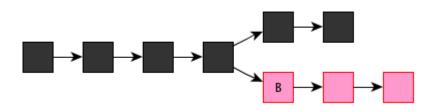


如果另外一个分支增长更快,那么这就是新的区块链



BitCoin原理-区块链竞争

- 攻击的方法: 先在黑色的区块链上创建一个交易, 使用一个UTX0上的比特币, 然后放弃掉这个区块链, 从红色的区块链开始追赶。
- 要求红色的区块链超过原来区块长度,然后在红色链上的区块再打包一个交易,将原来的UTX0再次交易一次,即双花。由于红色链是更长的区块链,所以最终全网承认的是第二次交易。第一次交易对象被欺骗,可能已经支付了货物。
- 伪造区块是一个消耗巨大计算力的过程,攻击者需要掌握超过BitCoin整个P2P网络51%计算能力才能成功。



所以在比特币的世界,算力就是记账的权力。但是如果有这样强大的计算能力,为什么不挖矿赚钱呢?

Segwit技术(隔离验证, BIP 141, soft fork)

我们前面看到交易签名sig是在vin,即input中,存在两个问题:

- 1) 签名跟交易数据混合在一起,不清晰
- 2)签名其实是可以变的(椭圆曲线算法可以有多个可验证签名),一旦签名变化、交易也变化、txid也会变化,导致交易可塑性。

Segwit就是把签名转移到区块的其他数据结构中

Unlocking Script (scriptSig) + Locking Script (scriptPubKey)

<sig> <PubK>

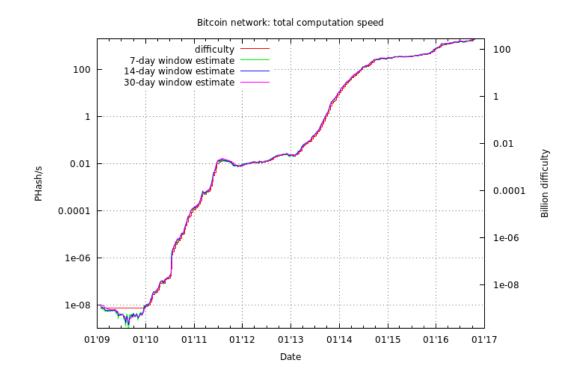
DUP HASH160 < PubKHash> EQUALVERIFY CHECKSIG

Unlock Script (scriptSig) is provided by the user to resolve the encumbrance Lock Script (scriptPubKey) is found in a transaction output and is the encumbrance that must be fulfilled to spend the output

- 每隔十分钟产生一个块,产生的块的链有可能不是最长链(工作量最大),一般6个区块后才能确认
- 那么一个交易要1个小时后才被认为是很难撤销的
- 一个区块大小为1M, 那么整个BTC网络每秒可以承担7笔交易。Segwit后 一般认为每秒14笔交易
- 速度和交易量都受到严重限制,如何将BTC用于现实的支付中?
- 闪电网络: lightning network

P2P技术新进展: BitCoin

2016年4月统计:比特币全网的计算能力相当于全球最快的500台超级计算机计算能力加总之和的两倍。



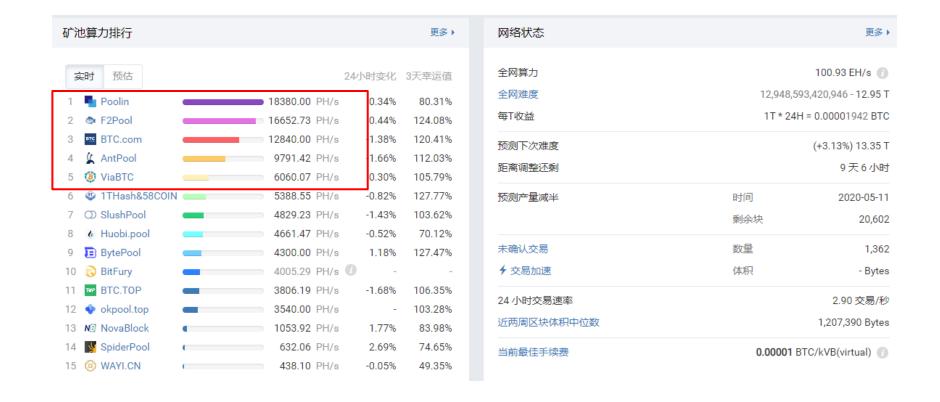
BitCoin原理-区块链竞争

• 还在持续增长:



BitCoin原理-区块链竞争

• 矿工集合起来形成矿池(大数定理),目前算力100E,一台流行的阿瓦隆1066 矿机,8500元,50T算力,即至少有100E*1000P*1000T/50T=200万台在挖矿



BitCoin原理-P2P网络作用

Bitcoin的所有节点构成一个第二代无结构的P2P网络: Gossip protocol

- 每个节点向其他邻居节点获得新的节点信息,并随机连接部分节点
- 向邻居节点采用flooding方式广播交易
- 没有中心或者超级节点, 所有节点地位平等
- 节点可以承担较为复杂的任务

BitCoin原理-P2P网络作用

论文中描述P2P网络的作用: (没有考虑矿工)

- New transactions are broadcast to all nodes.
- Each node collects new transactions into a block.
- Each node works on finding a difficult proof-of-work for its block.
- When a node finds a proof-of-work, it broadcasts the block to all nodes.
- Nodes accept the block only if all transactions in it are valid and not already spent.
- Nodes express their acceptance of the block by working on creating the next

P2P技术新进展:BitCoin和区块链

拜占庭将军问题, Bitcoin的解决办法:

- 每个将军给其他将军发送进攻时间、附加同意等任何消息,必须先解决一道 难题,将答案包含在消息中。
- 这种难题有很多个,每个将军都知道题目,而且一道难题要他的军队一起计算才能在10分钟后得到结果。
- 其他将军收到进攻时间消息后会首先验证难题是否解决正确,然后接着如果同意进攻时间,在后面附加自己的同意消息,如果不同意就提出新的进攻时间,这些消息发出同样需要和难题答案绑定。
- 如果发现有个进攻时间附加的同意消息大于自己之前同意的消息,那么也将 自己的同意附加到这个消息上。
- 那么在同一个进攻时间消息中附加的同意会越来越多(最长链),一旦所有将军们发现在一个消息中,有足够的将军(>=6)同意在一个时间点进攻,就可以开始攻击了。(信息不丢失,信息丢失拜占庭问题无解)
- 如果有叛徒要修改进攻时间,那么他需要跟其他将军竞争解题,叛徒解出难题的数量要超过其他忠诚将军的和,才能破坏忠诚将军们达成一致。

P2P技术新进展:BitCoin和区块链

Bitcoin解决办法方法的核心: Proof of Work

- 这个概念来自Adam Back的一篇论文: Hashcash A Denial of Service Counter-Measure
- 论文中Hashcash用于垃圾邮件过滤。假设所有的邮件服务器遵循一个规则,即所有想发送电子邮件的人,都需要在发送邮件的时候附加一个邮件的Hash值,前N个Bit为0(类似挖矿)。即使一次计算只需要几秒钟,对于海量发送垃圾邮件的系统来说都是致命的,因为多出的CPU时间对它们来说代价是非常大的
- Proof of Work不仅仅可以作为反拒绝服务的手段 ,还可以用来在分布式系统中达成共识



主要参考文献

- Peer-to-Peer Computing Dejan S.Milojicic, Vana Kalogeraki, Rajan Lu Kiran Nagaraja, Jim Pruyne, Bruno Richard, Sami Rolllions, Zhichen Xu, HP Laboratiories Palo Alto HPL-2002-57 March 8th, 2002
- 对等网络:结构、应用与设计;陈贵海等,清华大学出版社, 2007.9
- Chord paper
- Kademlia protocol
- Bitcoin: A Peer-to-Peer Electronic Cash System
- Mastering Bitcoin 2nd

参考书

