深入浅出区块链技术入门 - 区块链是什么

在区块链网络上,数据以区块的形式存储,想象一下有很多存储着数据的**区块**,它们被**链接**在一起,这些数据一旦被链接就对链上的任何人都可见,并且再也无法改变了。这是一项具有非凡革新意义的技术,可以用来记录我们能想到的几乎所有数据(如:产权、身份、余额、病历等等),**不用担心被篡改**。我们以比特币为例,来看看区块链是什么样子的。

比特币区块链是现存历史最悠久的区块链,它只存储比特币的交易数据,就像一个庞大的交易记录库,可追溯至第一笔比特币交易。

假设有三个存储着数据的区块,如下图:

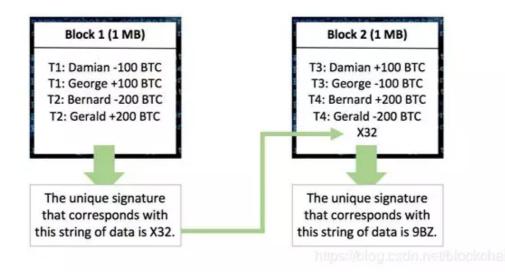


这三个区块内都存储了一些交易数据,就像三个独立的excel表格一样,记录了交易的内容。

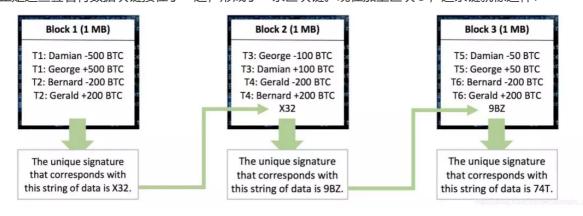
区块1按照时间顺序从第一笔交易开始记录,直到存满(1MB),之后的交易记录会继续依次存到区块2、区块3,以此类推.

每个数据块会根据内部数据生成一个特殊的独一无二的数字签名。

将区块1的数字签名添加到区块2中,就把区块1的数据和区块2的数据关联起来了,区块1的签名与区块2的其他数据一样成为了区块2签名的基础。如下图:



正是这些签名将数据块链接在了一起,形成了一条区块链。现在加上区块3,这条链就像这样:



为什么说它无法更改

先看看签名是怎么生成的。

将数据块中数据串,代入一个哈希函数得到一个独一无二的64位的值就是签名。

这看上去好像只需要逐个生成新的签名就可以神不知鬼不觉的更改数据了。怎么避免这种情况的发生呢?

我们再定义一下挖矿

区块链协议会对签名有一些要求,只有合格的签名才可以作为签名用,比如比特币区块链会根据前面一定数量的区块签名的算力,给出目前签名的难度,比如只有不少于连续10个零开头的签名才算是合格的签名,对应的区块才可以上链。

不要小看这"连续10个零开头"的规定,就是它让篡改数据的人头疼。

因为并不知道代入什么值才可以得到连续10个零开头的64位的值,只能不断的改变代入哈希函数的值计算,直到得到合格的签名为止,这个过程中不知道需要代入多少次,纯粹看运气。

数据块中的交易记录、上一个块的签名,是不可变数据,如果只将它们代入哈希函数会得出一个固定的值。

数据块中除了那些不可变数据,还另外添加了一段特定长度的、可以改动的数据,这就是nonce。nonce不是预先确定的数据,而是根据实际需要而找出的一串完全随机的数字。

也就是不断改变nonce,直到找到一个合格的签名,确定下nonce的值。这种反复更改nonce,对区块数据进行哈希运算以寻找合格签名的过程叫做挖矿[1],也就是矿工做的事。

最长链规则

在区块链上的所有矿工都必须在区块链协议[2]下工作,这是中本聪共识机制的一部分,始终以最长链为主链(有效链)的原则,所谓的"最长链规则"。(不是所有区块链都采用了中本聪共识机制)。

现在,我们再次假设有一个攻击者修改了某个数据块的数据。并且他想为之后所有的数据块生成新的签名。注意,网络上不只有他一个矿工,如今比特币区块链上有数百万矿工,其他的矿工都在夜以继日的挖矿,在不断的生成新的数据块,新的签名。这个攻击者要有多大的算力多么爆棚的运气才可以超过全网的矿工。

挖矿需要投入大量的电力,转化成算力,来找到合格的签名。也就是说找到一个合格的签名相当难,而且是需要大量成本的,而要超过全网的矿工,其成本投入不可估量。

总结

区块链上的矿工数越多, 整条链的安全性就越高。

有一种例外,恶意参与者的算力真的超过全网其他人的总和,从理论上讲是有可能篡改区块链的,这就叫做51%攻击,目前遭受过51%攻击[3]的著名区块链如bitGold、Verge、Ethereum Classic。

本文首发于系统学习区块链[4]技术博客——深入浅出区块链[5] - 打造高质量区块链技术博客,学区块链都来这里,关注知乎[6]、微博[7]。

References

[1] 挖矿: https://learnblockchain.cn/2017/11/04/bitcoin-pow/ [2] 协议: https://learnblockchain.cn/2019/01/09/consensus-security-51/ [4] 系统学习区块链: https://learnblockchain.cn/2018/01/11/guide/ [5] 深入浅出区块链: https://weibo.com/517623789