

## 65 | 区块链技术细节：加密和挖矿

2018-5-15 陈皓

前面一篇文章中提到的技术解决了交易信息不能被篡改的问题。但还有一个比较重要的问题，那就是，我们每个人只能发起和自己有关的交易，也就是能发起自己对别人付钱的交易，我们不能发起别人对我付钱，或是别人向别人付钱的交易。

那么，在比特币中是怎么解决这个问题的？让我们先看一些基础的加密技术。

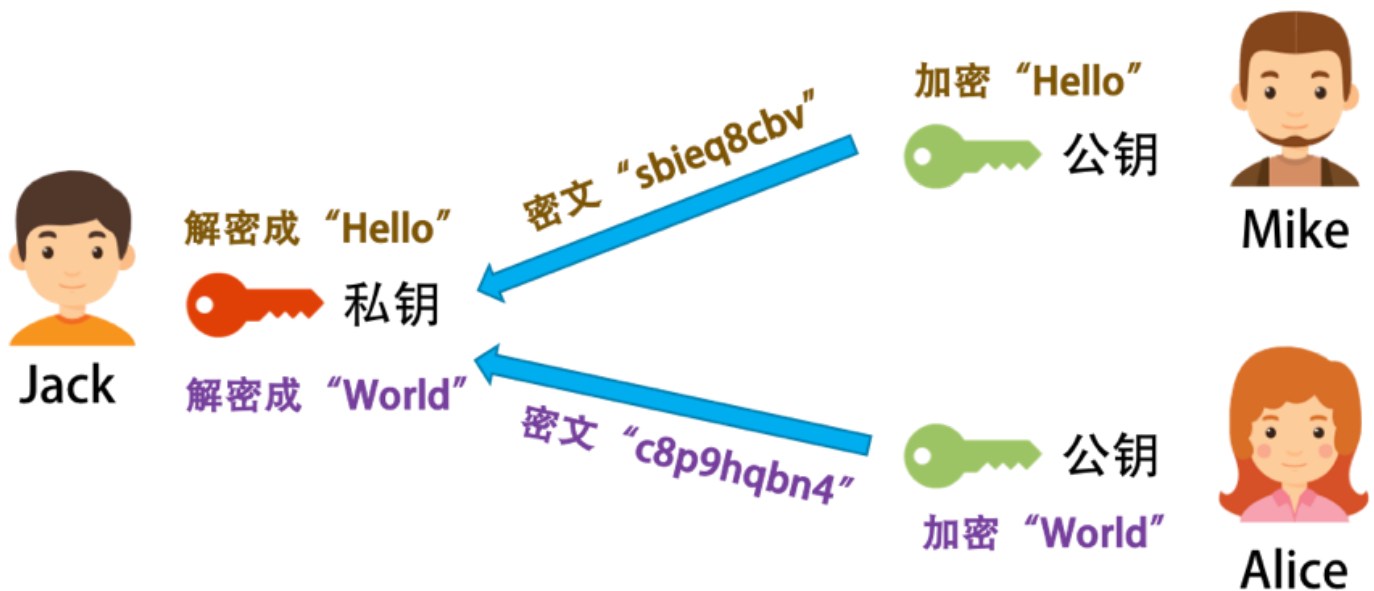
### 比特币的加密方法

#### 密钥对/签名/证书

所谓密钥对，也就是一种非对称加密技术。这种技术，在对信息进行加密和解密时，使用两个不同的密钥。这样一来，我们就可以把其中一个密钥公布出去，称之为公钥，另一个密钥私密地保管好，称之为私钥。

现实社会中，有人使用公钥加密，私钥解密，也有反过来用私钥加密，公钥解密，这得看具体的场景。（比特币使用了非对称加密的技术，其使用了 [ECDSA](#) 密钥对比技术。）

比如，我把我加密的密钥发布给所有人，然后大家都用这个公钥加密信息，但其他人没有私钥，所以他们解不了密文，只有我能解密文，也只有我能看得懂别人用我的公钥加密后发给我的密文。如下图所示。



但是，这会有个问题，那就是每个人都有我的公钥，别人可以截获Mike发给我的信息，然后自己用我的公钥加密一个别的信息，伪装成Mike发给我，这样我就被黑了。于是，我们需要对Mike的身份进行验证，此时就需要用到“数字签名”的概念了。

Mike也有一对密钥对，一个公钥给了我，私钥自己保留。

Mike发自己想要的信息，做个SHA或MD5的hash，得到一个hash串，又叫Digest。

Mike用自己的私钥，把Digest加密，得到一段Digest的密文。我们把这个事叫数字签名，Signature。

然后，Mike把他想发给我的信息用我的公钥加密后，连同他的数字签名一同发给我。

我用我的私钥解密Mike发给我的密文，然后用Mike的公钥解密其数字签名得到Digest。然后，我用SHA或MD5对解开的密文做Hash。如果结果和Digest一致，就说明，这个信息是Mike发给我的，没有人更改过。

这个过程如下图所示。



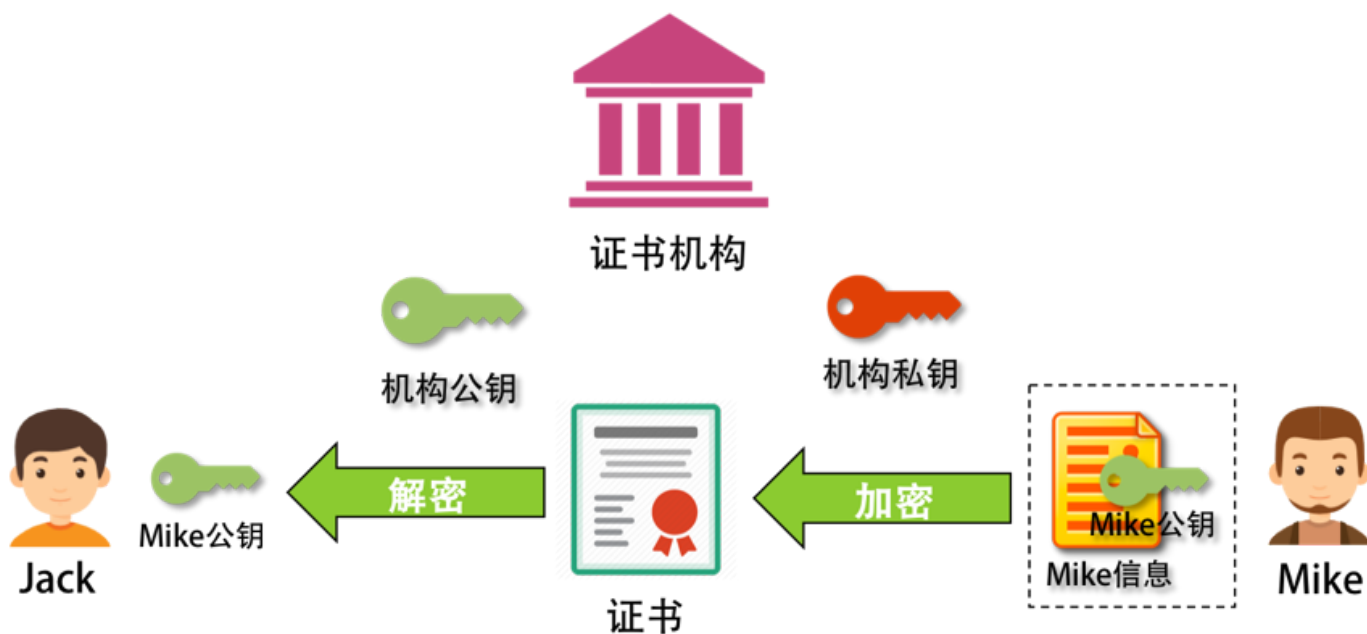
但是问题还没完。假设有个黑客偷偷地把Jack电脑上的Mike的公钥给换了，换成自己的，然后截获Mike发出来的信息，用自己的密钥加密一段自己的信息，以及自己的数字签名。

于是，对于Jack来看，因为他用了黑客的公钥，而不是Mike的，那么对他来说，他就以为信息来自Mike，于是黑客可以用自己的私钥伪装成Mike给Jack通信。反之亦然，于是黑客就可以在中间伪装成Jack或Mike来通信，这就是中间人攻击。如下图所示。



这个时候就比较麻烦了。Mike看到有人在伪造他的公钥，想了想，他只能和Jack找了个大家都相信的永不作恶的权威的机构来认证他的公钥。这个权威机构，用自己的私钥把Mike的公钥和其相关信息一起加密，生成一个证书。

此时，Jack就可以放心地使用这个权威机构的证书了。Mike只需要在发布其信息的时候放上这个权威机构发的数字证书，然后Jack用这个权威机构的公钥解密这个证书，得到Mike的公钥，再用Mike的公钥来验证Mike的数字签名。



上面就是整个密钥对、签名和证书的全部基础细节。比特币也用了这样的基础技术来认证用户的身份的。下面，我们来看看比特币的一些细节。

## 比特币的加密

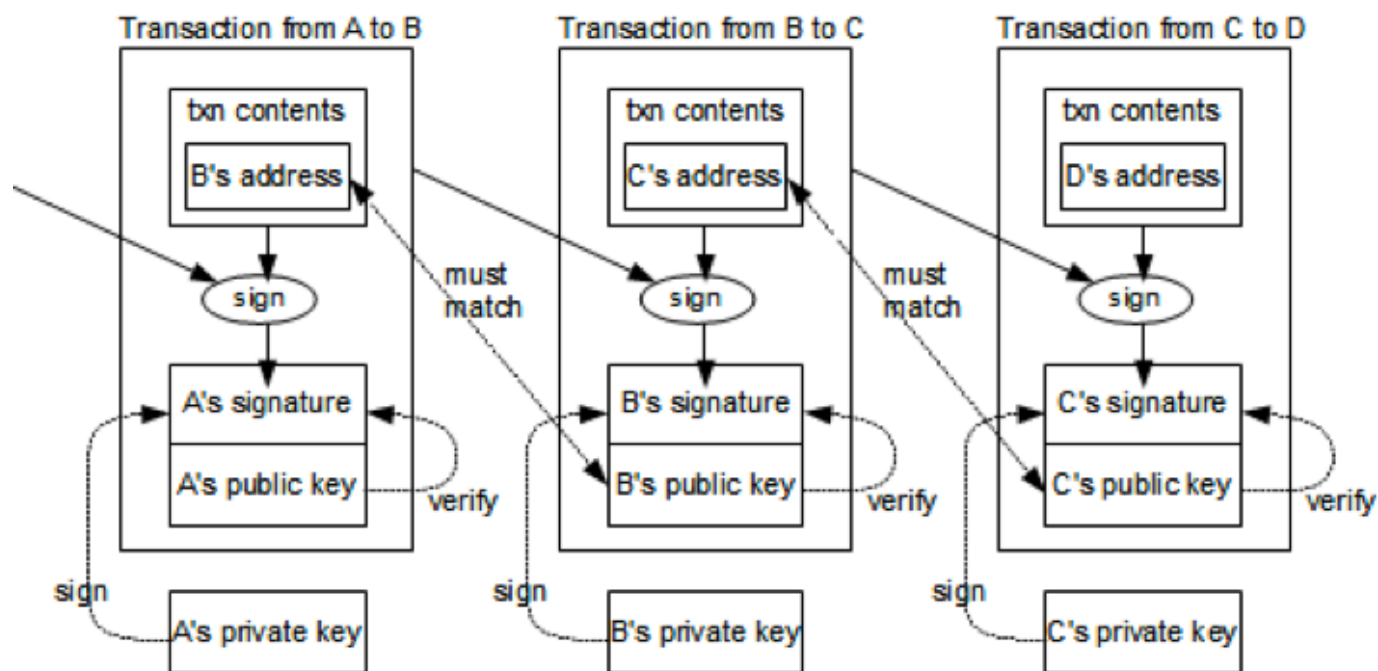
在比特币的世界里，每一笔交易的From和To都是每个用户的公钥（Public Key）。也就是说，使用用户的公钥来做交易的账户。于是，这个过程很简单。

交易的发起方只能是支付方，支付方需要用自己的私钥来加密交易信息并制作相关的交易签名。

网络上其他人会用你的公钥（也就是交易的支出方）来做解密来验证。

为什么不需要那个证书机构呢？不怕中间人攻击吗？这是因为，如果黑客想要伪造一笔别人的交易，那么他需要换掉半数以上结点上的被攻击者的公钥，这不太现实。与其这样做，还不如去偷被攻击者的私钥，可能还简单一些。

下面是一个交易链的图示。这个交易链的钱从A -> B -> C -> D，一共3笔交易。



图片来源：[Ken Shirriff Blog](#)

**发起交易。**我们从第一笔交易可以看到，A用自己的私钥为交易信息和自己的地址生成了交易的签名，然后把交易信息、自己的地址、交易签名和自己的公钥放出去，这样方便别人来验证的确是A发起的。

**验证交易。**在验证时，使用A的公钥解密交易签名，得到交易的hash值。把交易信息和自己的地址做hash，看看是不是和签名解密后的hash值一致。

这里需要注意一个细节，比特币的地址是由我们的公钥生成的，生成规则比较复杂，可以参看Bitcoin的Wiki页 - [Technical background of version 1 Bitcoin addresses](#)。

# 比特币的挖矿

前面说到，在比特币的区块hash算法中，要确保下面这个公式成立：

$$\text{SHA-256}(\text{SHA-256}(\text{Block Header})) < \text{Target}$$

而在区块头中，可以完全自由修改的只有一个字段，就是Nonce，其他的Timestamp可以在特定范围内修改，Merkle Root和你需要记录的交易信息有关系（所有的矿工可以自由地从待确认交易列表中挑选自己想要的交易打包）。

所以，基本上来说，你要找到某个数字，让整个hash值小于Target。这个Target是一个数，**其决定了，我们计算出来的hash值的字符串最前面有几个零**。我们知道，hash值本身就是一串相对比较随机的字符串。但是要让这个随机的字符串有规律，是一件很困难的事，除了使用暴力破解，没有其他办法。在计算机世界里，我们把这个事叫"哈希碰撞"(hash collision)，碰撞前几个位都是0的哈希值。

下面是一个示例。我想找到一个数，其和"ChenHao"加起来被hash后的值前面有5个零。

测试程序如下：

```
import hashlib

data="ChenHao"

n=1
while n < 2**32:
    str = data + `n`
    hash = hashlib.sha256(str).hexdigest()
    hash = hashlib.sha256(hash).hexdigest()
    if hash.startswith('00000'):
        print str, hash
        break
    n = n + 1
```

这是一个暴力破解的算法。这个程序在我的MacBook Pro上基本要10秒钟才跑得出来结果。

找到1192481时，找到了第一个解，如下所示：

ChenHao1192481

00000669e0eeb33ee5dbb672d3bd2deb0c32ef9879ef260f0debbdcb80121160

那么，控制前面有多个0的那个Target又是怎么来的呢？是由Bits这个字段控制的，也就是难度系数，前面需要的0越多，难度也就越大。其中的算法你可以看一下Bitcoin的Wiki上的[Difficulty](#)词条，这里我就不多说了。

**这个难度系数，会在每出2016个区块后就调整一次。现在，这个难度是要在前面找到有18个零。如下所示(一个真实的区块链的Hash值)：**

[000000000000000000424118cc80622cb26c07b69fbe2bdafe57fea7d5f59d68](#)

**\*\***一个SHA-256算法算出来的哈希值有  $2^{256}$  种可能性，而前面有18个零意味着前面有72个bits是零。于是，满足条件的哈希值是有  $2^{184}$  种可能性，概率是  $\frac{1}{2^{72}}$  **\*\***。

是的，很有可能你穷举完Nonce后还找不到，那就只能调整Timestamp和Merkle Root（调整不同的记账交易）了。

所以，一般的挖矿流程如下。

从网络上取得之前的区块信息。

从"待记账区"中获取一组交易数据（有优先级，比如成长时间、矿工小费等）。

形成区块头（计算Merkle Root并设计记账时间Timestamp等）。

开始穷举Nonce，来计算区块头的hash值。如果前面有18个零（小于Target），那么记账成功。如果没有，则从第一步重新开始。

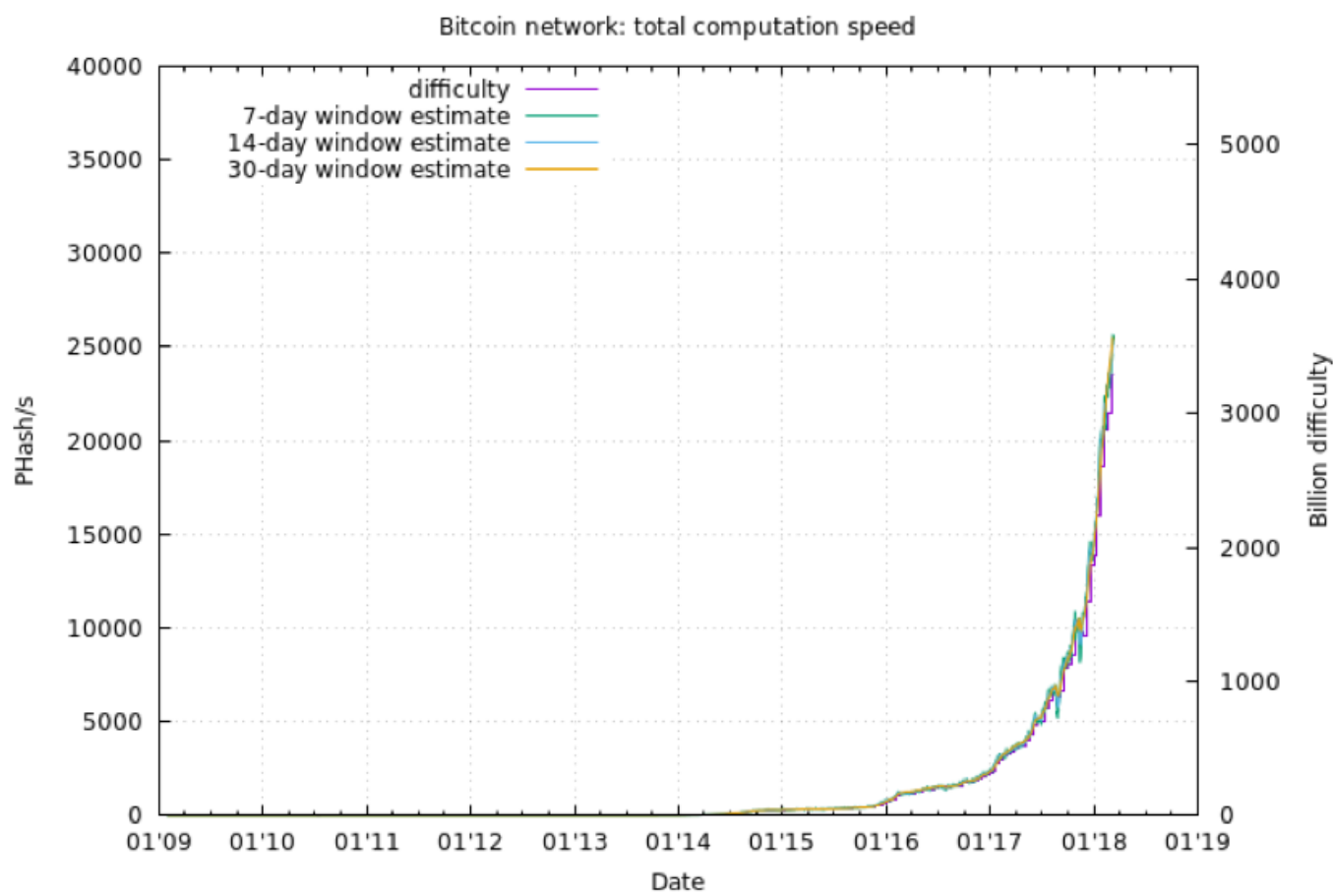
一旦某矿工成功打包一个区块，他就会告诉其他矿工。收到消息的矿工会停下手上的工作，开始验证，验证通过后，广播给其他矿工。

所以，满足条件的这个难度系数成为了挖矿的关键。设置这个难度系数就是为了让全网产生的区域名平均在10分钟一块。而根据比特币无中心服务器的架构，也就是其挖矿的机器数量

是想来就来想走就走的，算力可能会不一样。因此，为了保证每10分钟产生一个区块，当算力不足的时候，难度下降，当算力充足的时候，难度提高。

今天的这18个零，基本上来说，一般的电脑和服务器就不用想了，必须要算力非常非常高的机器才能搞定。所以，在今天，挖矿这个事，已经不是一般老百姓能玩的了。

下图展示了整个比特币的难度历史。



( 图片来源：<http://bitcoin.sipa.be> )

上面这个图只是算力的表现，可能并不直观。我们还是用其耗电量来说可能会更好一些。根据"Bitcoin Energy Consumption Index"统计，截至 2017年11 月 20 日，比特币过去一年挖矿的电力总消耗已累计达 29.51 TWh ( 1TWh =  $10^{12}$  Wh )，约占全球总电力消耗的 0.13%。该数字甚至已经超过近 160 个国家或地区一年的电力消耗，包含冰岛和尼日利亚。若全球的比特币矿工自成一国，该国的电力消耗排名可排到全球第 61 名。

看到这里，你一定要问，为什么要挖矿呢，不就是记个账呗。为了系统地说明这个问题，我们下面来看看去中心化的共识机制。



文末给出了《区块链技术》系列文章的目录，希望这一系列内容对你有启发，有帮助。

[区块链的革命性及技术概要](#)

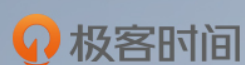
[区块链技术细节：哈希算法](#)

[区块链技术细节：加密和挖矿](#)

[去中心化的共识机制](#)

[智能合约](#)

[传统金融和虚拟货币](#)



# 左耳朵耗子

全年独家专栏《左耳听风》

20000 名程序员的练级攻略

陈皓

资深技术专家  
骨灰级程序员



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

## 精选留言 21



陈小喵~

1522904525

耗子叔，非常佩服您的知识广度和深度。想问下这么多知识您是怎么记录下来的呢？简单来说您是否有一套自己记笔记或收集存储知识的方法...包括知识的分类，整理，定期的回顾，新知识的添加，老知识的内容更新等等...以及使用的工具什么的...这个是否能分享下呢？



作者回复 后面我会写提高学习能力的文章，敬请关注！

---



**龚极客**

1523148171

关于数字签名的图片左侧应该用公钥对数字签名解密，然后把解密结果和md5('hello')对比吧？

作者回复 是的。我的图可能没画好，但是文字描述如你所说。

---



**Ray**

1522938529

关于证书机构颁发公私钥那段，如何做到防止中间人攻击？黑客不是照样可能通过木马方式获得通信双方的公私钥嘛？

作者回复 当然可以，其除了需要伪装成发送和接收方，还要伪装成证书机构，才能做到神不知鬼不觉，证书中有证书的源信息，比如，公钥的服务器域名，你还要伪造DNS服务器.....另外，如果黑客种了木马，那不需要这么复杂了，这意味着私钥都被黑了，这才是灾难——你的身份就是你的私匙！别人拿到了你的私匙，你就已经不是你了。所以，如果你的比特币的私匙被盗，相当于你的钱被盗了，区块链不像银行，其只认私匙，不认人.....

---



**i**

1528016853

难度系数是自动调整的吧？这个调整机制是怎样的？

---



**杨洪林**

1524303633

不太明白下面挖矿的代码为什么计算两次哈希值？一次不就可以验证有没有挖到矿了吗？

```
hash = hashlib.sha256(str).hexdigest()
hash = hashlib.sha256(hash).hexdigest()
```

---



**edisonhuang**



1564535307

区块链中的安全性通过每个人的一对密钥，公钥和私钥，以及对自己数据做数字证书的方式来保证。挖矿则是通过计算暴力穷举一个合法字符串的过程，挖矿难度与算力相关动态调整，全网算力越强挖矿难度越大

---



**zjg**

1555915713

“然后 Jack 用这个权威机构的公钥解密这个证书，得到 Mike的公钥，再用 Mike 的公钥来验证 Mike 的数字签名”，我觉得文中这句话有错误。

证书的数字签名是由CA的私钥生成的，所以数字签名的验证需要用CA的公钥来验证，而不是Mike的公钥。

---



**沙漠之鹰**

1549776998

比特币是怎么保证有限的

---



**李海洋**

1549545711

比特币的总量是怎么控制的，算法上怎么做的，10分钟产生一个区块又是怎么定的

---



**尾巴的爸爸**

1539919358

尊敬的陈老师，您好！

我在阅读专栏时，顺便学习代码，发现了 一个问题，期望您的解答。

在测试程序部分：

```
str = data + `n`
```

这行代码是否存在问题？这样的话，这个传入的字符串始终是“ChenHaon”，程序就陷入死循环了。

我修改了一下，将变量名改名为input\_str:

```
input_str = data + str(n)
```

程序能正常运行，得到结果，耗时约1.72秒（台式机）。

请问这样改是否合理？

多谢

---



**够钝**

1537362988

那我们还是去银行把，，，既然黑客这么□□

---



**anbien**

1529844104

“用这个权威机构的公钥解密这个证书，得到 Mike 的公钥，再用 Mike 的公钥来验证 Mike 的数字签名”，Mike的公钥本身应该没用CA私钥加密吧？加密的应该只是Mike证书内容的摘要

---



**neohope**

1529664403

其实感觉Fabric的整体架构和经典的安全架构更靠近一些。挖矿的化，是不是增加一下分叉的相关知识，以及BITCOIN和ETH如何奖励矿工稍微好一些？

---



**Cong Chen**

1528561769

如果可以使用Mike的公钥来验证签名，而Mike的公钥是全网发布的，那么Mike所发的信息不就变成全网可截获了吗？所以这种信息不确保私密性？

作者回复 这种方式只是为了验证消息是不是Mike发出来的。不是为了私密。

---



**Wilson\_qqs**

1524098537

比特币的挖矿中那个公式用的是<target而不是=.意思是只要找到一个符合条件的hash就可以了？

---



**Wilson\_qqs**

1524096512

同样的疑问被一个读者提问了。就是权威机构颁发证书来防止中间人攻击。其实，黑客同样可以把jack电脑上的机构公钥换成自己的，然后截取证书，用机构公钥解密证书获取各种信息再伪装成自己的发给jack.只是这种方法对黑客来说成本有些高？这种方式并不能绝对防止中间人攻击对吧，耗子哥？

作者回复 是可以的。但为了成为中间人，攻击者不但要能同时和服务器，客户端通信，还要嵌入到服务器和客户端的通信链路之中，将服务器的数据转发给客户端，将客户端的数据转发给服务器。实现这样目的的手段有多种，比较常见的有DNS劫持和局域网ARP欺骗。

---



**一般无**

1523170274

请问所谓的小费是怎么回事？

为什么比特币会是有限的？

有没有可能一笔交易额太小了没有挖矿机愿意为他记账？

作者回复 小费是：我转你10元，但给你12元，其中2元是给矿工的小费。手续费：比特币不鼓励小额交易，对于小额要收手续费。比特币如果是无限的，就会导致通货膨胀。第三个问题，一笔交易会有未确认时间，未确认时间越长被记账的优先级就越高。

---



**李连杰**

1523133159

交易费用太高了，算力或电力消耗就是维护区块链交易制度所需的交易费用，肯定是行不通的，不符合经济学原理。人是追求利益最大化的，所以大家有动力一起建立和遵守某个规则，这个规则一定使大部分人的个体收益和整体收益增加。感觉区块链很“脑残”啊，少了最重要的一根弦。

---



**逆行**

1522980718

“挖矿”是有奖励机制的，即奖励比特币，而比特币又是有限的，等比特币到达上限后，谁又来给“挖矿”买单呢？或者没有了奖励机制，谁又来打包交易生成区块呢？

作者回复 交易的手续费

---



**shooter**

1522918271

<https://www.jianshu.com/p/954e143e97d2>

哇 大大也在关注区块链

恰巧看到有人写的 btc如何生成地址 比较的详细

---



**horizon**

1522910981

资源浪费啊！