**和王小云教授学习区块链的密码学基础**

韩锋

比较透彻的了解比特币的密码学基础，还是通过清华大学的王小云教授。

王小云教授是国际密码学界的风云人物，2004年、2005年先后破解了比特币密码系统的前身MD5和SHA-1，轰动了国际密码学界，促进了美国整个加密系统进化到SHA-256，也就是现在比特币使用的密码学系统。

在二零一三年十二月一个风和日丽的下午，我到清华大学高研院（杨振宁中心），听王小云教授详细讲解了SHA-256密码学机制。

SHA，是[Secure Hash Algorithm](http://en.wikipedia.org/wiki/Secure_Hash_Algorithm)的缩写，意思是基于哈希（hash）函数算法的加密系统。

王老师首先给我介绍了一下什么是哈希(hash)函数，这是一种加密算法，一般写为：h=hash(m),h代表哈希值，m代表对应这个哈希值的解（message）。哈希函数有这样几个特点：一是已知m，很容易通过h=hash（m）验证出它对应的哈希值h，但反过来很难，就是已知哈希值要求出对应的解m很难。比如大素数分解就是这样一种算法，如果已知一个一百位以上的大数，能分解成几个素数，我们很容易通过把这几个素数乘在一起，就能验证其是不是这个大数的分解。但反过来，如果我们预先只知道这个一百多位（甚至更多位）的大数，要想算出它的素数分解可需要相当大的计算量，甚至这是现在量子计算机所期待完成的任务。正是由于哈希函数有这样运算的不对称性，或者说不可逆性，所以它特别适合为密码学所用，比如哈希值就适合当加密的“公钥”，可以完全公开，但是人们即使得到了公钥，也几乎不可能一下子算出它的“私钥”，也就是哈希函数的解m。但是反过来，如果我们已知私钥m，却很容易验证它对应的公钥就是哈希值h，这就是所谓不对称加密算法。

哈希函数第二个宝贵特性是：如果“解”（也就是私钥）稍有不同，那么它对应的哈希值就会有很大不同，这叫雪崩效应（avalanche effect），比如：

SHA256的一个解m是：("The quick brown fox jumps over the lazy dog")

其对应的哈希值为：

0x d7a8fbb307d7809469ca9abcb0082e4f8d5651e46d3cdb762d02d0bf37c9e592

SHA256的另一个解m’是（注意到与上面一个解差别有多小吗？）

("The quick brown fox jumps over the lazy dog.")

其对应的哈希值就“雪崩”般的变为：

0x ef537f25c895bfa782526529a9b63d97aa631564d5d789c2b765448c8635fb6c

看起来两个哈希值完全没有共同点吧？

哈希函数所具有的这个优秀特性，保证了我们不同的私钥（哪怕只有一点点差别），都不会和对应的公钥搞混！

王老师还介绍说：SHA-256是基于十六进位制的加密系统，也就是每一位上允许有十六个比特的不同信息，一般用十个阿拉伯数字和前六个英文字母表示：0,1,2,3,4,5,6,7,8,9,a,b,c,d,e,f。所以你要是看到这样一个十六进位的哈希值：Hash: 00000000000000004cf3aa249551432fa84da4de05e9cfc3e6d95a5ce8bed5f7

（这是比特币世界2014-02-08 03:06:30美东时间，刚挖出的一个比特币block对应的哈希值），不要觉得奇怪哦！

之所以叫SHA-256，就是因为其哈希值有64位，每一位上有十六比特也就是二的四次方种选择，这样总的哈希值就可以有2的256次方个比特：



真是天文数字哈，所以中本聪把SHA-256加密算法选为比特币的挖矿算法，因为哈希值前面每增加一个零，寻找其解m的难度就会增加二的四次方倍。因为SHA-256还没有像王小云教授之于SHA-1那样被破解，所以任何一个六十四位的哈希值如ef537f25c895bfa782526529a9b63d97aa631564d5d789c2b765448c8635fb6c，找到其解m，都是没有固定算法的，只能靠计算机随机的hash碰撞，所以一个挖矿机每秒钟能做多少次hash碰撞，就是其“算力”的代表，单位写成hash/s,这就是所谓工作量证明机制POW(Proof Of Work)

正是基于SHA-256这种十六进位制的加密算法，所以中本聪在TA最原始的比特币论文中写到：

The proof-of-work involves scanning for a value that when hashed, such as with SHA-256, the hash begins with a number of zero bits. The average work required is exponential in the numberof zero bits required and can be verified by executing a single hash.

意思就是：工作量证明过程包括扫描SHA-256的哈希数由多少个0开头，每增加一个0，平均工作量都会有指数级的增加，就是二的四次方，增加了多少个零就是多少个二的四次方乘在一起倍数的工作量增加，这些将在解一个哈希数（也就是挖一个比特币的block过程）中得到证明。

明白了这一点，我们从blockexplorer上查到00000000000000004cf3aa249551432fa84da4de05e9cfc3e6d95a5ce8bed5f7

这是比特币世界2014-02-08 03:06:30美东时间，刚挖出的一个比特币block对应的哈希值，前面有十六个零，也就是需要每秒工作量（每秒算力，也就是每秒哈希撞击次数）才能挖到矿，这应该也是比特币当前挖矿的全网算力的数量级。

我们再查一下blockchain.info上的全网算力24,107,483.60 GH/s，意思是现在全网算力是每秒哈希撞击次数达到了24P的数量级。

比较一下理论值和实际算力的ln值：



两者基本吻合。比特币系统就是靠对于挖矿的哈希值前面加零，来控制挖币的总量，不管全网算力如果增加，都能够通过在哈希值前面加零来保证每个block目前都只能挖出25个币，好在现在还只加到二的六十次方，离二的二百五十六次方尚远。

如果有一天，比特币世界挖矿算力真是达到了SHA-256加密系统的极限，也就是二的二百五十六次方，那全网算力的成本会达到大约240万亿美元再乘十的五十次方倍，现在来看是人类历史达不到的一个天文数字！