

密码学技术何以为信？深究背后的计算困难性理论

原创 李昊轩 微众银行区块链 3月18日

来自专辑

WeDPR隐私保护周三见

第3论 / 隐私保护
周三见

李昊轩

微众银行区块链核心开发者



和我微信交流



































隐私保护为何选用密码学算法？密码学算法背后有哪些神奇的数学理论？3何时比9大？计算可逆性错觉究竟是如何在数学领域被打破？

这里，我们将从密码学信任的理论基础出发，分享在隐私保护技术方案中应用密码学技术的一些思考：如何理解密码学算法的能力边界，如何客观地比较不同密码学算法对于隐私保护方案有效性的影响。

这一切，要从密码学神奇的“不对称性”说起。

神奇的“不对称性”

早在公元前，古埃及、古罗马、古希腊等古文明均已开始使用密码技术来保护信息的机密性，历史上最早的不对称性表现为选用特殊的信息编码方式，如果第三方不知道具体的编码方式，则难以解码对应的信息。

A 	A 	I 	U,W 	B 	P 
F 	M 	N 	R 	H 	H 
KH 	KH,CH 	S,Z 	S 	SH 	K 
K 	G 	T 	TJ 	D 	DJ 
Y 	Y 	U,W 	M 	N 	L 

大约经过4000多年的发展，也就是近代20世纪初，现代密码学正式成型，引入了关于不对称性更为严谨的数学定义。比较有代表性的早期论文包括1929年Lester S. Hill在美国数学月刊上发表的《Cryptography in an Algebraic Alphabet》。

20世纪末，随着因特网的普及，大量敏感数据在网络上进行传输，产生了大量的数据内容保护的需求，密码学技术也因此得到飞速发展。

在现代密码学中，关于不对称性，大家最熟悉的概念莫过于“公钥”和“私钥”。

以加密通信为例，主人公小华要向他的朋友美丽通过加密的方式发送一份电子邮件，可以先找到美丽的公钥，使用公钥对邮件内容进行加密，并将加密后的密文发送给美丽。美丽收到邮件内容的密文之后，通过自己的私钥进行解密，最终得到邮件内容的明文。

以上过程中，密码学算法神奇的不对称性体现在以下问题中：

- 为什么只有美丽可以解密邮件内容？
- 为什么其他人不能通过美丽的公钥反推出她的私钥？

这些问题的答案，都要归结于密码学中的计算困难性理论。

0.2.

计算困难性理论

在隐私保护场景中，计算困难性理论具体表现为，对同一隐私数据主体，通过不同计算路径，获得相同信息的计算难度具有不对称性。不对称性中，相对容易的计算方式被用来构造授权的数据访问，而困难的计算方式被用来避免非授权的数据泄露。

构造这样的不对称性的方式有很多，最经典的方式之一，就是千禧年七大难题之一——P和NP问题。

P问题是确定性图灵机，即通用计算机计算模型，在多项式时间($O(n^k)$)内可以计算获得答案的一类问题。NP问题是确定性图灵机在多项式时间内可以验证答案的正确性，但不一定能计算出答案的一类问题。

关于同一份答案，验证过程比计算过程要容易很多，由此我们可以构造出密码学算法所需要的计算难度不对称性。

NP问题是否能够通过有效的多项式时间算法转化成P问题，由此破解计算难度不对称性？目前学术界尚无定论。

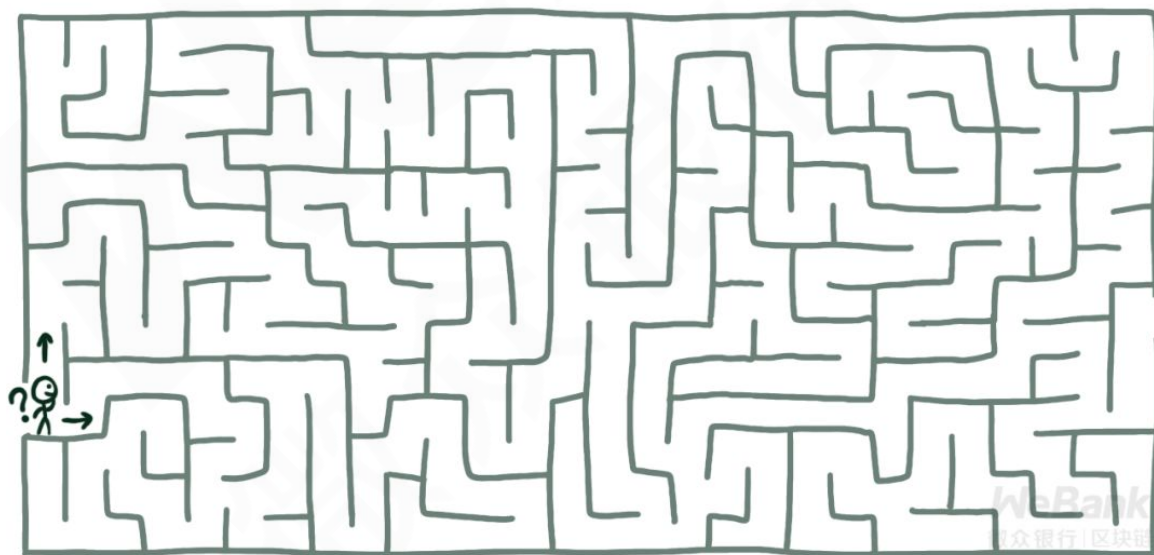
理论研究进一步表明，对于NP问题集合中的核心问题，即NP完全问题，如果能够找一个有效的多项式时间算法来解决任何一个NP完全问题，那其他所有NP问题都可以基于这个算法来构造出有效的多项式时间算法。由此，之前提到的计算难度不对称性将不复存在。

幸运的是，经过将近70年的科学探索，这样的算法并没有被发现。在有限时间内，现代计算机难以求解这些问题的答案，所以现代密码学可以比较安全地基于这些NP完全问题来构造有效的密码学算法。

0.3

神奇的“计算困难问题”

形象地讲，计算困难性理论的核心就是构造一个迷宫，如果不知道捷径，是很难到达出口的。



我们日常所用的各类密码学算法，其有效性都与这一理论息息相关，这里重点以非对称密码学算法为例，介绍其中经典的迷宫构造蓝图，即三大计算困难问题：

- 大数分解困难问题
- 离散对数困难问题
- 椭圆曲线上的离散对数困难问题

大数分解困难问题

给定两个大素数 p 和 q ，计算 $n=p*q$ 是容易的。然而，给定 n ，求解 p 、 q 则是困难的。

整数的素数分解是数论中最著名的问题之一，目前，求解素数分解最有效的方法称为数域筛法，即通过构造代数数域不停地对整数可能的集合进行迭代运算。

目前，大整数分解问题仍不存在更有效的分解方法，因此密码学一些方案利用大数分解困难问题构造相应协议，如RSA系列算法将其困难性规约为大数分解困难问题。如果大整数分解困难问题被破解，使用RSA密码方案保护的隐私数据也会相应遭到破译。

离散对数困难问题

在模为 n ，生成元为 g 的有限域中，给定整数 a ，计算 $g^a = b$ 是容易的。然而，给定 b 和 g 计算 a 则是困难的。

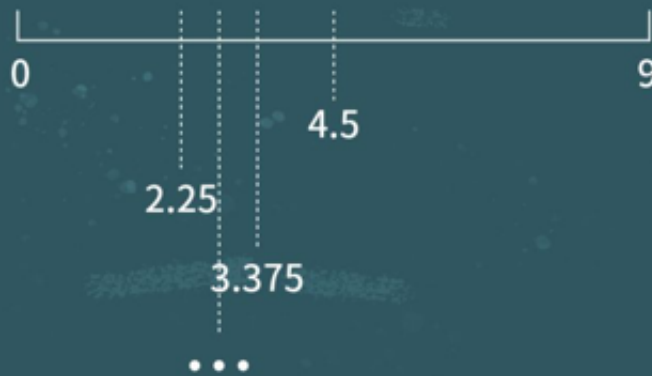
许多新接触密码学的读者都会对离散对数问题产生计算可逆性的错觉，看起来就是进行一次 \log 运算的事情，但真相并非如此。

在实数域，元素有一个非常重要的性质，全序关系，所以很容易比较大小。例如，在实数域中 $9 > 2$ 且 $3 > 2$ ，一定能推出 $9 > 2$ 。

在计算 $\log_2(9)$ 时，计算机会对以元素9为输入的函数结果进行二分查找法，首先计算 $(9/2)^2$ 和9进行比较，再计算 $((0+9/2)/2)^2 \dots$ 。通过不停比较元素大小的性质，从而计算 \log 最终的结果。

计算机的计算过程：二分查找法

$$\log_2 9 = 3.1699$$



$$4.5^2 > 9$$

$$2.25^2 < 9$$

$$3.375^2 > 9$$

...

WeBank
微众银行 区块链

然而，在有限域中，元素之间并不存在全序关系。在模为7的有限域中，可以看到诸如9等于2，3比9大的关系存在。

因此，无法通过有效的算法计算上述过程中的 a 。许多著名的密码协议安全性都是建立在离散对数困难性上的，如Diffie-Hellman密钥交换协议、ElGamal加密、DSA算法等。

椭圆曲线上的离散对数困难问题

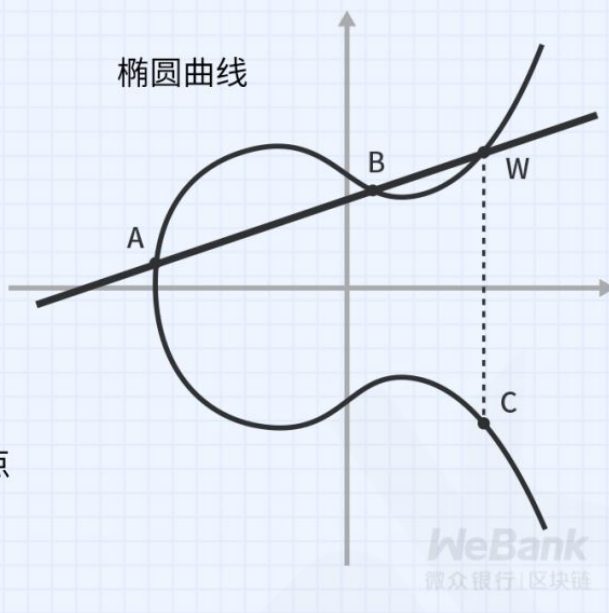
当前，椭圆曲线密码学算法是当前密码应用的主流，每一个隐私数据都能以坐标 (x, y) 的形式，表示为椭圆曲线上的一个点。与一般离散对数困难问题类似，椭圆曲线上的离散对数困难问题可以表示为：

在有限域 F 上的椭圆曲线群，点 P 为曲线上某个点，给定整数 a ，计算 $a \cdot P = Q$ 是容易的。然而，根据 P 和 Q 计算 a 则是困难的。

有别于普通代数运算，椭圆曲线上的点运算定义如下：

$$C=A+B$$

- ① 画一条直线连接A、B两点，该直线与椭圆曲线相交于W点。
- ② 通过W点，画一条垂直于x轴的直线，与椭圆曲线的另一个交点就是C点。



可以看到，椭圆曲线上的点运算和普通实数域上的运算有很大差别，当前尚未存在一种有效的算法对椭圆曲线离散对数问题进行破译。目前，最常用的公钥密码算法体系ECDSA、EdDSA、国密SM2等都是基于这一困难问题。

0.4.

客观比较不同的密码学算法

由于不同的密码学算法构造使用了不同的困难问题，对应地，不同的困难问题也势必会引入不同的安全假设。

理解这些安全假设，是企业进行技术选型，客观地判断基于不同密码学算法构造隐私保护方案孰强孰弱的关键。

这里，我们需要进一步引入“安全参数”的概念。

安全参数是一个衡量密码学算法保护隐私数据强度的数值。对位于“同一等级”的安全参数值来说，不同密码学算法的安全级别基本相同，即面对已知最有效的攻击方式，算法被破解导致隐私数据泄露的概率相同。

一般情况下，安全参数值的大小，直接体现为密钥长度的长短。在同一等级下，安全参数值有大有小，对应的密钥长度也有长有短。

基于不同困难问题的密码学算法密钥最小长度，美国国家标准与技术研究院NIST作如下推荐，其中，每个单元格表示需要使用密钥长度的最小比特数。

安 全 参 数等级	基于大数分解困难问 题的公钥加密算法	基于离散对数困难问题 的公钥加密算法	基于椭圆曲线上的离散对数 困难问题的公钥加密算法
1	1024	1024	160
2	2048	2048	224
3	3072	3072	256
4	7680	7680	384
5	15360	15360	512

通过上表，我们可以看到，即便密钥长度相同，选用不同困难问题获得的安全级别是不同的。一般而言，基于同一困难问题构造的技术方案，密钥长度越长，安全性越高，相应地，系统效率越低，其中往往也伴随其他系统设计上的取舍。

不同场景应按照业务需求选择适合的技术方案和密钥长度，具体有以下几点需要特别注意：

- 隐私保护技术方案的安全性取决于其使用的密码学算法实现中最低的安全参数等级。
- 在未指明安全参数的前提下，进行密码学算法的安全性比较没有实际意义。
- 如果安全参数值很小，一般表现为对应的密钥长度很短时，无论密码学算法设计多么精妙，实际效果可能都是不安全的。
- 由于困难问题选用上的差异，密码学算法的理论强度没有最强，只有在满足特定安全假设下的够强，强行比较基于不同困难问题的密码学算法哪个更强通常没有实际意义。

计算困难问题归根结底还是一个计算问题，随着计算机计算能力的增强，或是算法理论研究进展的推进，这些困难问题的安全性都会发生变化。如RSA加密算法，NIST密钥管理准则认为，2010年后，1024位的密钥不再安全，需要增加到2048位的密钥长度，预计其安全有效性可以保持至2030年。

对于企业而言，这里的启示在于，不能简单地认为，隐私保护技术方案现在有效，就保证了10年后依旧有效。无论什么样的隐私保护技术方案都有其时效性。

企业如果能够根据权威技术组织推荐的安全参数、算法方案及时更新现有的系统，困难性理论就能够有效保障隐私保护技术方案的有效性历久如新。



正是：密码学技术易守难攻，困难性理论当居首功！

作为密码学安全的基石，计算困难问题和相关的安全参数，是企业有效进行密码学算法选型的关键考察点。企业应用落地时，需充分考虑隐私数据保密的有效期，选择合适的密码学算法和密钥长度，对数据安全性和系统效率进行必要衡量。

除了与密码学算法直接相关的计算困难问题，一个完整的隐私保护技术方案通常还需要构造密码学协议，来组合多种密码学算法。密码学协议引入了多方之间的交互，由此也引入其他重要的安全假设。

这些安全假设对评价隐私保护技术方案的整体安全性、有效性、实用性至关重要，具体分析，敬请关注下文分解。

---END---

《隐私保护周三见》

“科技聚焦人性，隐私回归属主”，这是微众银行区块链团队推出《隐私保护周三见》深度栏目的愿景与初衷。每周三晚8点，专家团队将透过栏目和各位一起探寻隐私保护的发展之道。

栏目内容含括以下五大模块：关键概念、法律法规、理论基础、技术剖析和案例分享，如您有好的建议或者想学习的内容，欢迎随时提出。

栏目支持单位：零壹财经、陀螺财经、巴比特、火讯财经、火星财经、价值在线、链客社区

往期集锦

第1论 | [隐私和效用不可兼得？隐私保护开辟商业新境地](#)

第2论 | [隐私合规风险知几何？数据合规商用需过九重关](#)



长按二维码关注

微众银行区块链



白皮书下载 | 订阅干货 | 进群交流 | 合作联系