密码学入门系列（一）

马宇峰 溪塔科技 2019-12-13



在区块链的整个体系中大量使用了密码学算法，比如用于 PoW 的哈希算法，用于完整性验证的 Merkle Tree，用于交易签名与验证的数字签名算法，用于隐私保护的零知识证明等等。

可以说密码学是保证区块链安全的基石，而区块链的广泛应用也推进了密码学的发展。在区块链内核 CITA 的 v0.18 中，新增了「基于 Rust 语言的国密算法库」新特性。这次更新，使用户在尊重版权的前提下，即可自由调用 Rust 实现的国密算法库，来匹配业务场景所需的国密签名算法，大幅降低企业用户及开发者获得高性能区块链底层设计服务的成本，方便用户打造最贴近业务需求的区块链。

本文作者是秘猿科技研究院密码学研究员马宇峰，阅读本文将带领你走进密码学的世界，初探这个神秘之境的风景。

如果您对技术感兴趣，欢迎关注我们开源代码库（点个 Star 呗）：
GitHub：https://github.com/citahub

什么是密码学

密码学的英语单词是 Cryptograghy，是由希腊单词 Kryptos（隐藏）和 Graphin（写）派生出来的，最初代表的意思是用来隐秘的传递信息。隐藏和写就是隐写，在古典密码学的发展中就有一门称为隐写术的技术，比如说藏头诗就是一种隐写术。在《巨人的陨落》中，艾瑟尔和

弟弟比利就是通过每隔两个单词就会加一个单词来作为加密后的密文，这也是隐写术的一个例子。隐写术发展到今天演变为数字水印技术，一般在文件中加一个标识信息（即数字水印），可以起到追踪溯源，防伪和版权保护的作用。

密码学一开始的功能是在有恶意攻击者存在的环境下，保护双方通信安全，现在是用来保护信息安全的核心技术。

现代信息安全的基本要求：

- 信息的保密性 Confidentiality：防止信息泄漏给未经授权的人（加密解密技术）
- 信息的完整性 Integrity：防止信息被未经授权的篡改（消息认证码，数字签名）
- 认证性 Authentication：保证信息来自正确的发送者（消息认证码，数字签名）
- 不可否认性 Non-repudiation：保证发送者不能否认他们已发送的消息（数字签名）

古典密码学

以时间划分，1976 年以前的密码算法都属于古典密码学，基本使用在军事机密和外交领域，它的特点就是加解密过程简单，一般用手工或机械就可以完成。

古典密码学现在已经很少采用了，然而，研究古典密码的原理对于理解构造和分析现代密码都是十分有益的。尤其是对称加密技术，它就是从古典密码学中演化进来的。

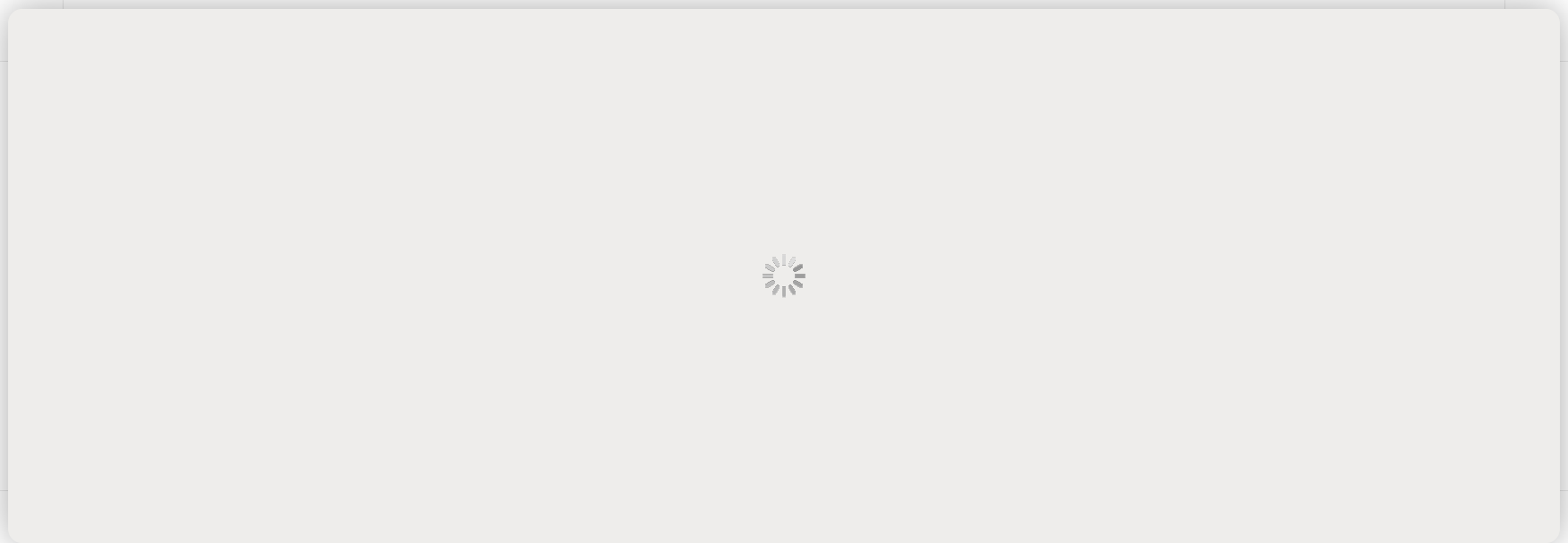
古典密码学中最经典的两种算法：

- 置换密码：又称换位密码，加密过程中明文的字母保持相同，但是顺序被打乱。只要把位置恢复，就能得到明文。
- 代换密码：明文中的每一个字符被替换成密文中的另一个字符。接收者对密文做反向替换就可以恢复明文。

一个典型置换密码的例子：公元前 500 年的古希腊斯巴达邦城，存在一种叫做「棍子加密」的加密方法。找一个腰带，将信息横着写在腰带上，但是这个信息是完全打乱的，需要一个可

以解密的棍子，将腰带缠绕在棍子上，就可以恢复出明文。这也是最简单的置换密码的方式。

后来随着抽象代数的出现，有了矩阵之后，就可以做一个复杂的置换加密，对运算做一个置换，用置换矩阵来解密，恢复出明文。代换密码中最简单的是凯撒密码，它是一种单表代换密码，加密方式就是通过对字母的位移进行加密，比如把字母表右移三位，上面是明文表，下面是对应的密文表。如下图所示：



每一个都有相应代表的位置，像 A 代表着 D，B 代表这 E。单表代换密码就很容易被破解，只要用频率分析表就可以破解，这是根据人类自然语言中，字母出现的频率不同来进行破解的。比如英文 E 是使用最频繁的，其次是 T/R/N/I/O/A/S 等；有些字母使用的很少，例如 Z/J/K/Q/X 等，这样就可以获得英文字母使用频率分布表，这个表是根据几本书获得的频率分析来获得的，同时，统计双字母组合和三字母组合的使用频率也非常有用。

有了破解的方法后，密码学中也会相对应的出现防止破解的方法。

- 多名或同音代替密码

与简单代替密码类似，只是映射是一对多的，每个明文字母可以加密成多个密文字母。用已知明文攻击比较容易破解。

- 多字母代替密码

每次对 L 个字母进行代换，隐藏或均匀化字母的自然频度，用于抵抗频率分析。比如 hill 密码，用已知明文攻击比较容易破解。

- 多表代替密码

多代替密码可以说是古典密码学的巅峰之作，是以一系列（两个以上）代换表依次对明文消息的字母进行代换的加密操作。先把明文分成多份，代换表有很多个，根据序列依次更换代换序列，对每个序列进行加密。典型的代换密码：vigenere cipher，博福特密码，enigma 密码机（这个密码机可以说是古典密码学中最厉害的一种，但是在 1940 年时，被图灵给破解了）。

古典密码中还有比较偏的，比如二战时期，在太平洋战场上，日军总能用各种方法破译美军的密电码，这令美军在战场上吃尽了苦头。为了改变这种局面，1942 年，29 名印第安那瓦霍族人被征召入伍。因为他们的语言没有外族人能够听懂，所以美军将他们训练成专门的译电员，人称「风语者」，他们的语言到现在都还没有人能够破解。

总结古典密码学的特点：

- 计算强度小
- 出现在 DES 之前
- 数据安全基于算法的保密。这和现代密码有很大的差距，只要知道加密方法，就能轻易的获取明文。现代的密码基于秘钥的加密，算法都是公开的，而且公开的密码算法安全性更高，能被更多人评论和使用，加强漏洞的修补。
- 以字母表为主要加密对象。古典密码大多数是对有意义的文字进行加密，而现代密码是对比特序列进行加密。这也是现代密码和古典密码的区别，而且古典密码的分析方法也是用字母频率分析表来破解的。
- 替换和置换技术
- 密码分析方法基于字母与字母组合的频率特性以及明文的可读性

现代密码学

现代密码学有三个代表事件：

- 1976：由 Diffie 和 Hellman 在《密码学的新方向》（《New Directions in

Cryptography》）提出了公钥密码学体制的思想

- 1977年：美国国家标准局颁布数据加密标准 DES（Data Encryption Standard）
- 1978年：第一个公钥算法 RSA 算法（由 Ron Rivest、Adi Shamir 和 Leonard Adleman 的姓氏首字母组成）

现代密码学的意义是让密码学成为了一门科学，研究方向从军事和外交走向了民用和公开。古典密码学更像一门艺术，为什么是艺术？因为古典密码学是需要用一种非常精妙的方式对明文加密，现代密码学可以通过形式化验证来证明它的安全性。

现代密码学主要有三个方向：私钥密码（对称密码）、公钥密码（非对称密码）、安全协议。

私钥密码

私钥密码也称对称密码，是对文字的加密转换成对比特序列的加密（相对于古典密码），用同一个密钥进行加密和解密操作，这个密钥发送方和接收方都是要保密的，所以称为私钥密码。它的两个基本操作就是代换和置换就是来源于古典密码学的。

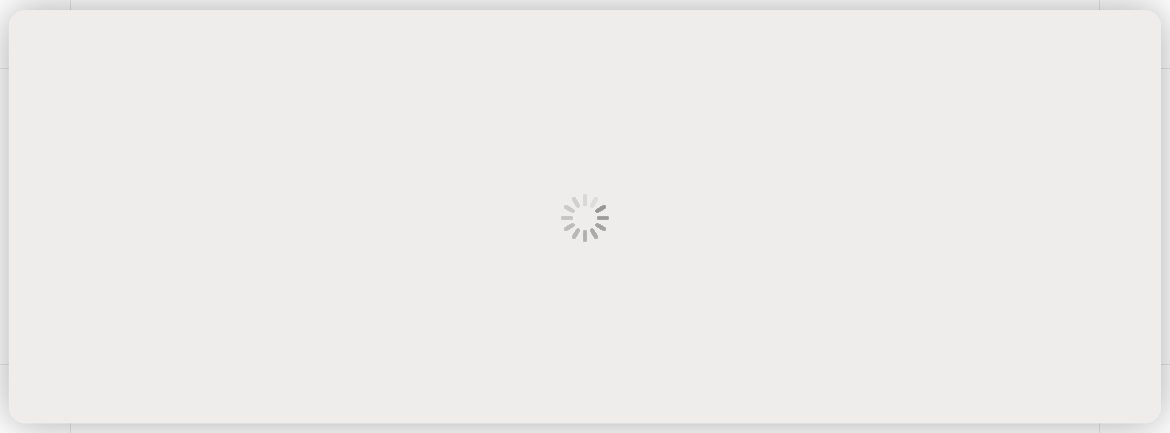
对称密码有两个设计原则，一个是扩散（Diffusion）：明文的统计结构被扩散消失到密文的长程统计特性，使得明文和密文之间的统计关系尽量复杂。

另一个是混乱（confusion）：使得密文的统计特性与密钥的取值之间的关系尽量复杂。

对称密码的代表有 DES 算法和 AES 算法，AES 算法是一个算法标准，是从 15 个对称加密算法中进行竞争选出来的，我国的 SM4 算法是我们现在使用的对称密码算法。

公钥密码

在讲解公钥密码学之前，先说明下公钥密码学的数学基础。先简单介绍下费马，来自法国的业余数学家之王。他有个有趣的故事，在 1637 年，费马在看一本书时，在书的边沿空白处写下一个看起来类似勾股定理的公式：



然后他又在旁边写了个结论「当 N 大于 2 时，这个方程式整数解」，并且自称知道怎么证明，但是书的空白处写不下证明过程。这个数学公式后来成为了数学界的三大猜想之一：费马猜想，其余两个猜想分别是哥德巴赫猜想和四色猜想。费马猜想在 1994 年被数学家安德鲁怀尔斯和他的学生理查泰勒证明，因此他们两人也获得了数学界的诺贝尔奖——Fields Medal。

以上是费马大定理，它和密码学的关系不大，和密码学关系比较紧密的是费马小定理，费马小定理中隐约有了「群」的雏形。欧拉小定理则是 RSA 算法的核心原理。

费马小定理和欧拉定理：

群论由法国数学家伽罗瓦提出，伽罗瓦是一位很年轻就去世的数学家，后世称他脾气暴躁才导致自己早亡。在 1830 年法国七月革命爆发，他批评校长将他们困在学校的保守做法，因此被校长劝退。劝退后，在社会上发表过一些比较激烈的政治言论，两次入狱。

第二次入狱后，在狱中认识一位医生的女儿，两人陷入热恋。出狱后不久，医生女儿的另一位追求者要求和他进行决斗。在决斗前一晚，他还在疯狂的记录他的数学成果，有可能是认为自己活不下来，果然第二天就离开世间。当然这个是比较富有浪漫色彩的说法，也是被广泛流传的说法。

伽罗瓦可以说是绝对的天才，只学了 5 年的数学，就发明了群论。他离世后，他的朋友将伽罗瓦写过两篇论文寄给卡尔·弗里德里希·高斯与雅各比(又翻译作：雅可比)，但是都石沉大海，一直到 1843 年，才由刘维尔肯定伽罗瓦结果的正确、独创与深邃，并在 1846 年将它发表。

伽瓦罗可以说是过早离开的天才。后来他的群论就成为未来密码学的基础，现在来研究下群论是什么。群论是一种代数结构，代数结构就是有若干集合，比如群：（G，*）。

需要有以下四个性质

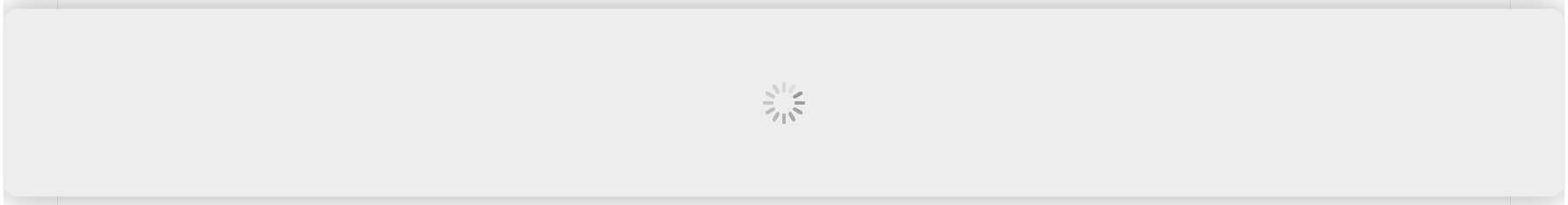
- 封闭性：群中任意两个元素经过*运算后，结果仍然是群中的元素
- 结合律：(a*b)*c = a*(b*c)
- 单位元：存在单位元（幺元），与任何元素相乘，结果不变；
- 逆元：每个元素都存在逆元，元素与其逆元相乘，得到幺元



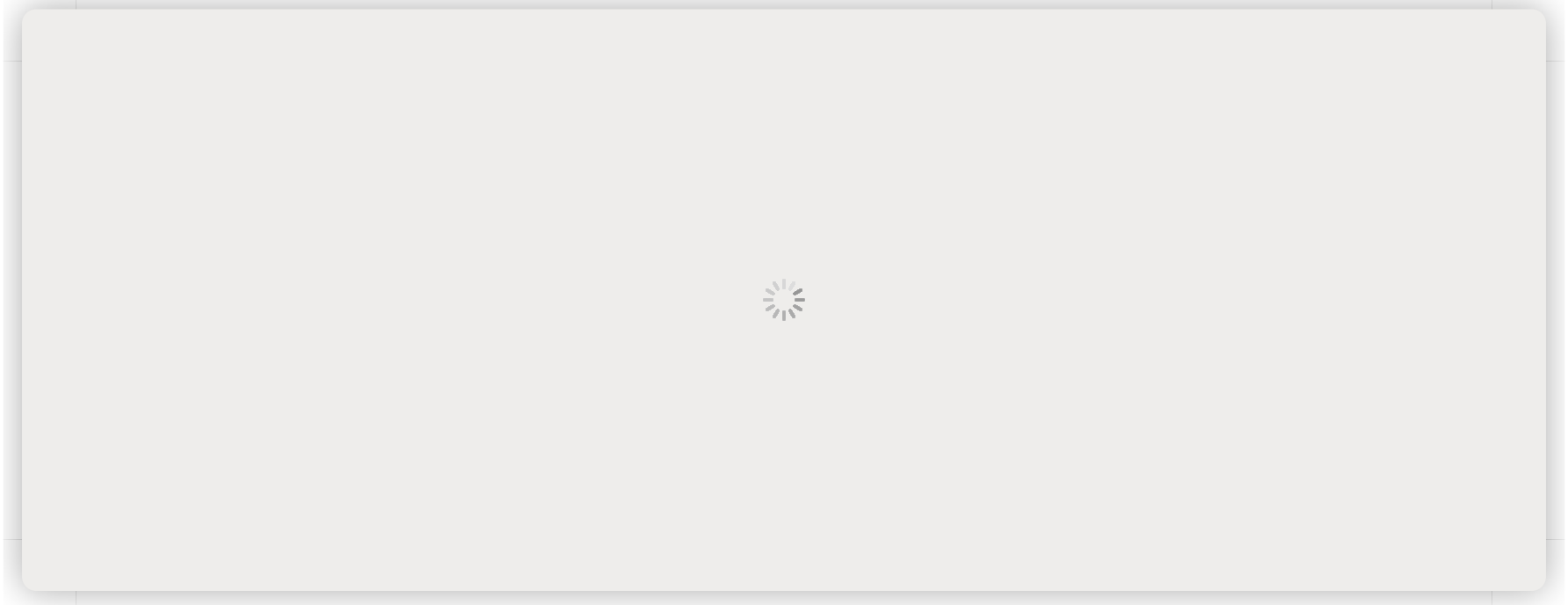
DH 密钥交换协议

DH 密钥交换协议是在 1976 年 Diffie 和 Hellman 提出的密码学新方向中提出的协议，它主要解决的是密钥分发问题，是公钥密码学的开端，它的安全性是基于计算 Diffie-Hellman 问题的难解性。

计算 Diffie-Hellman 的难解性：



看下这个协议的具体描述：



RSA 算法

RSA 算法是第一个公钥密码算法，也是第一个数字签名算法。具有乘法同态性，也可以称为第一个乘法同态特性的算法，它被提出的时间最早，关于它的研究最为广泛，因此也是理论上最成熟的密码学算法。算法如下：

数字签名算法的用处有很多，如数字货币交易时，需要用到私钥签名。RSA 签名算法是第一个数字签名算法，算法如下：

除了 RSA 数字签名之外，现代密码学中还包含很多特殊的数字签名。

盲签名：签名者不知道签名的消息就可以对这个消息进行签名。盲签名最早出现在电子银行和电子现金的设计中，目的是防止银行知道资金的流向，现在很多公平交易协议都是用盲签名来做的。

环签名：一种特殊的群签名方式，在一个签名群体中，任意某个人的签名会隐匿在这个群体当中，外人只能看到这个群体里的成员对这个信息进行签名，但并不知道来自具体的哪个成员，RingCT 和门罗币中就是运用环签名技术来保护用户隐私，Intel 的 SGX 也使用了群签名技术。

聚合签名：多个签名和多个私钥来对不同的信息进行签名，最后聚合成一个签名，这样的好处是加快验证速度，减少存储空间。

多重签名：多个签名者对同一条消息进行签名，最终生成一个签名。

门限签名：在一个签名群体中，超过门限门槛的签名者对消息进行签名后，会公布出代表这个群体的一个确定性的签名，Dfinity 的 VRF 就是使用门限签名实现的。

一次性签名：只能对消息进行一次签名，如果第二次签的话，私钥就会暴露，计算快速，一般使用于传感器之类的计算资源有限的场景中。

密码学热点

接下来我们介绍一下目前密码学的研究热点。

后量子密码算法：当量子计算机理论逐渐成熟，好多像基于离散对数问题、大数分解困难问题的公钥密码学算法都因为量子计算机的出现而被攻破，为了应对这种情况的发生，就出现了后量子密码算法。现在一般是基于格或 hash、超奇异椭圆曲线的同源问题来设计后量子密码算法。

生物密码学：生物密码是 RSA 里的 A，阿德尔曼在 1994 年提出的利用生物的生化特性来做 DNA 计算，可以破解 DES 加密，有人利用生物的加密特性来设计一次性密码本的加密技术。

同态加密：现在主要是 Intel 在研究，现在已经有全同态加密算法了，缺点是效率比较低，但是同态加密在云计算中还是很有用的。

区块链：今年欧密会（密码学三大顶级会议之一）就有 5 篇关于区块链的论文。这表明了许多密码学者开始研究密码学在区块链中的应用，区块链为密码学的发展带来新的活力，密码学也为区块链的发展提供了有力保障。

安全协议：零知识证明协议，安全多方计算协议等。

基于双线性对的密码学：这个是近年研究的一个热点。像属性基加密就是基于双线性对的公钥密码算法，另外配合秘密分享方案也可以用来设计门限签名算法。

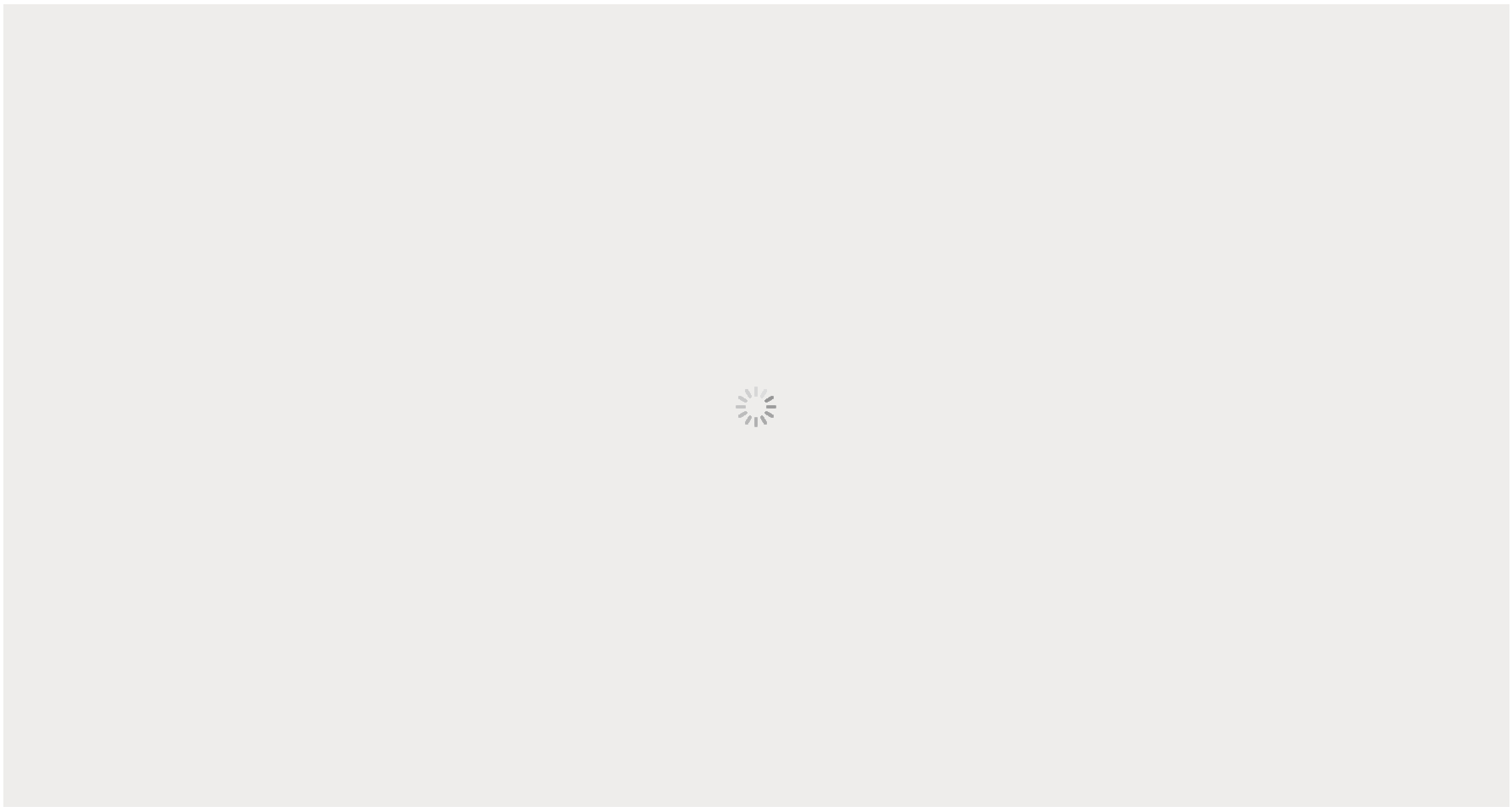
这是密码学系列文章的第一期。下期我们会给出一个基于 BLS 签名的门限签名算法，并且有白板推演的完整视频哦，敬请期待 ~

推荐阅读

- *拜占庭容错的代表 PBFT*

- *BLS 签名和基于 BLS 签名的门限签名*
- *密码学入门系列（一）*
- *CITA架构与云计算服务*
- *白话零知识证明（一）*
- *白话零知识证明（二）*
- *Authenticated Dynamic Dictionaries in Ethereum*
- *区块链与共同知识库*
- *P2P 网络思考*
- *智能合约的一种设计结构*

如果您对技术感兴趣，欢迎关注我们开源代码库（点个 Star 呗）：
GitHub：https://github.com/citahub



阅读原文