密码学原语如何应用？解析密码学特有的数据编解码

原创 廖飞强 微众银行区块链 5月13日

来自专辑

WeDPR隐私保护周三见

隐私保护方案的工程实现，如何关联到学术论文中天书一般的公式符号？密码学工程中，有哪些特有的数据编解码方式、存在哪些认知误区和注意事项、需要克服哪些限制和挑战？

作为支撑隐私保护方案的核心技术，如何运用数据编解码，将密码学论文中抽象的数学符号和公式具象成业务中具体的隐私数据，是学术成果向产业转化需要跨过的第一道门槛。

学术论文中所使用的数学语言与工程中所使用的代码编程语言，差异非常大。不少在数学上容易定义的属性和过程，若要在工程上提供有效实现，颇具挑战。实现不当的话，甚至可能破坏学术方案中的安全假设，最终导致方案失效、隐私数据泄露。

常用的密码学算法拥有多种标准化编解码方式，其应用到隐私保护方案，可以分别解决相应问题。以下将逐一展开。

业务应用难题：类型不匹配
工程实现之道：数据映射

在实际业务中，隐私数据可以表现为五花八门的数据类型，这些类型通常不满足密码学协议中特定的类型要求，无法被直接使用，这就是我们需要解决的第一个问题：数据类型不匹配。

例如，业务系统中，交易的金额是一个长整型整数，而常见的密码学算法可能要求输入为有限循环群中的一个元素，如果直接使用长整型整数的值，可能该值并不在对应的有限循环群中；在椭圆曲线系统中，单个数值还需要转化成曲线上的点坐标，需要将一个数值转化成两个数值的坐标形式。

针对以上问题，密码学工程实现中，一般通过**数据映射**进行类型转换处理。具体而言，是将用户的隐私数据，通过一定的方法，变换到具体密码协议要求的数据类型。

下面以密码学中的椭圆曲线(Elliptic Curve)加解密为例，介绍一种常见的数据映射方式。

椭圆曲线可以简单理解为定义了一个特定点的集合，例如下面这种公式定义了比较常见的一类椭圆曲线：

$$y^2 = x^3 + ax + b\ (mod\ p)$$

其中满足公式成立的点(x, y)都在椭圆曲线上。椭圆曲线密码通过在限定的点集上定义相关的点运算，实现加解密功能。

在椭圆曲线加解密过程中，首先面临的问题是『如何将待加密的数据嵌入到椭圆曲线上，通过点运算来完成加密操作』。这需要将明文数据m映射到椭圆曲线上的一个特定点M(x,

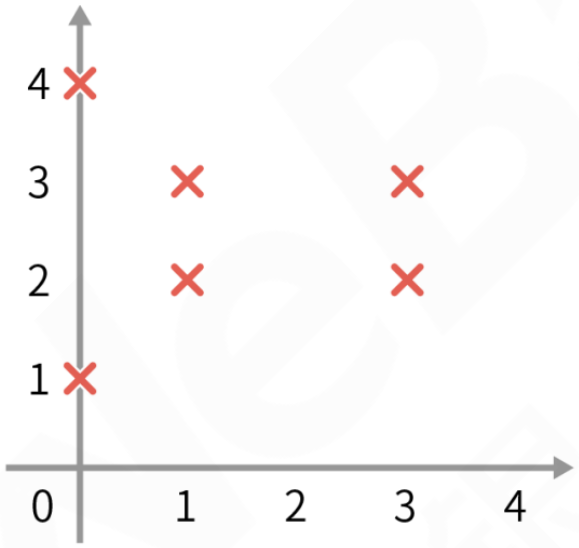
y)。

数据编码方式是将明文数据m通过进制转换到椭圆曲线上某点的x坐标值，然后计算m^3 + am + b的完全平方数，得到y，这样就将m转换到了点M(x, y)。

数据解码方式比较直白，解密还原出明文数据点M之后，读取M的x坐标值，再通过进制转换还原为明文信息m。

然而，密码椭圆曲线是定义在有限域上的，即曲线上是一个离散的点集合。这样会导致计算完全平方数不一定存在，即x没有对应的y在椭圆曲线上，那么，部分明文数据无法转换到椭圆曲线上的点，从而导致部分数据无法被直接加密。

$$y^2 = x^3 + 2x + 1\ (mod 5)$$



x，y的有效值域：0,1,2,3,4

当x等于2和4时，不存在满足曲线的关系的整数y，由此可能会导致数据编解码失败。

在实际工程化的方案中，为了保证椭圆曲线加解密的可用性，会加入其它更复杂的扩充编码机制，以应对明文数据转换失败的情况。

一般而言，密码学协议中所定义的类型要求越多，数据映射的工程实现也会越复杂，如果缺乏高效的数据编解码算法和配套的硬件优化支持，即便密码学协议的理论计算复杂度再低，最终也是难以实用化。

具体的数据映射涉及到很多流程细节和算法参数，一旦存在微小的差异，由不匹配的编码算法所产生的数据，都会极大概率无法解码，导致隐私数据丢失、业务中断。

所以，在具体工程实现时，数据映射需要严格按照已有工程标准的实现要求，以国密SM2为例，可以参考GM/T0009-2012《SM2密码算法使用规范》、GM/T0010-2012《SM2密码算法加密签名消息语法规范》等一系列相关技术标准。

业务应用难题：数据太长
工程实现之道：数据分组

除了类型不匹配，密码学协议中使用的核心算法对输入的数据长度往往也有一定要求。但在实际应用中，需要处理源自不同业务需求的隐私数据，难以限定其长度，难免会出现数据长度超出核心算法处理长度的情况。

例如，对称加密AES算法AES-128、AES-256，表明其使用的密钥位数分别是128位和256位，但加密过程中单次进行核心密码运算时处理的数据固定为128位。

针对以上问题，密码学工程实现中一般通过**数据分组**进行处理，即化整为零，将长数据切分为多个较短且符合长度要求的数据块。

典型的例子是分组加密，例如AES、DES等。分组加密顾名思义就是，将输入的数据分组为固定长度的数据块，然后以数据块为单位作为核心密码算法的处理单元进行加解密处理。

为了在数据分组之后，依旧保持方案的安全性，数据分组技术不仅仅是简单地对数据进行划分，还需要引入额外的流程操作。

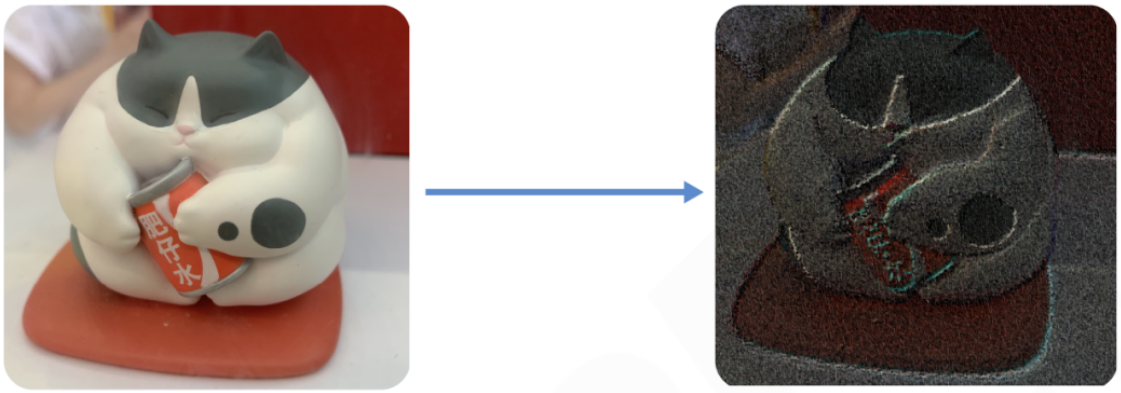
下面以AES 256位密钥加密为例，介绍其中典型的分组加密模式ECB、CBC和CTR。

ECB模式 (Electronic Code Book)

ECB是最简单的分组加密模式，也是不安全分组模式的典范。

假定有1280位待加密的数据，ECB模式将其平均分为10个128位数据块。每个数据块使用相同的密钥单独加密生成块密文，最后块密文进行串联生成最终的密文。

ECB模式下的分组加密



明文　　密文

ECB模式泄露了密文数据之间的关联性，使得密文数据在一定程度上依旧可读。

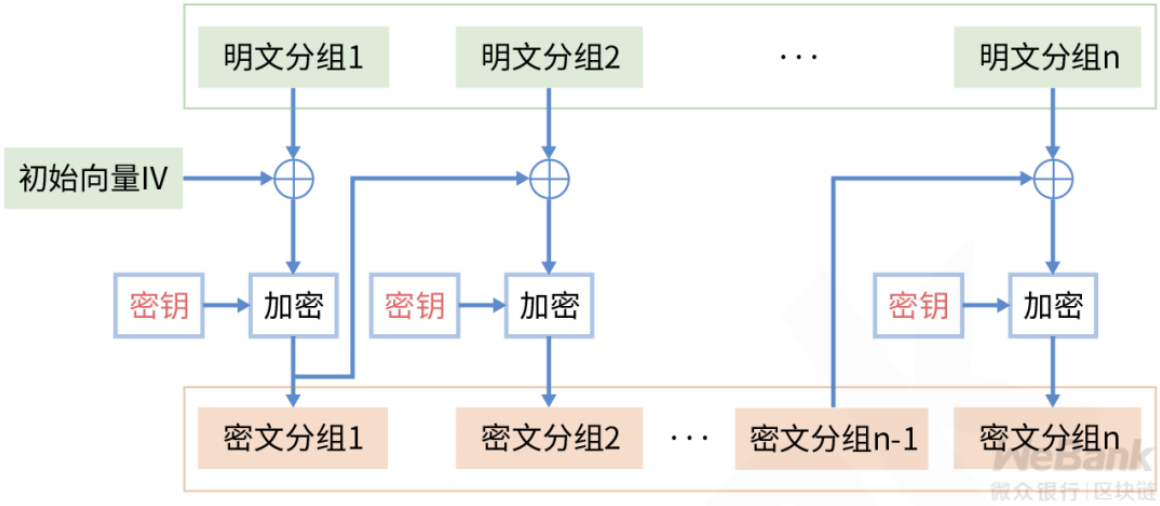
ECB模式的加密特点是在相同的明文和密钥情况下，其密文相同，因此泄露了明文数据与密文数据之间的关联性，不推荐用于任何隐私保护方案中。

CBC模式 (Cipher Block Chaining)

CBC模式通过前后数据块的数据串连避免ECB模式的缺点。

与ECB模式类似，CBC模式中，每个明文块先与前一个密文块进行异或后，再进行加密。在这种方法中，每个密文块都依赖于它前面的所有明文块。同时，为了保证每个数据密文的随机性，在第一个块中需要使用一个随机的数据块作为初始化向量IV。

CBC模式下的分组加密

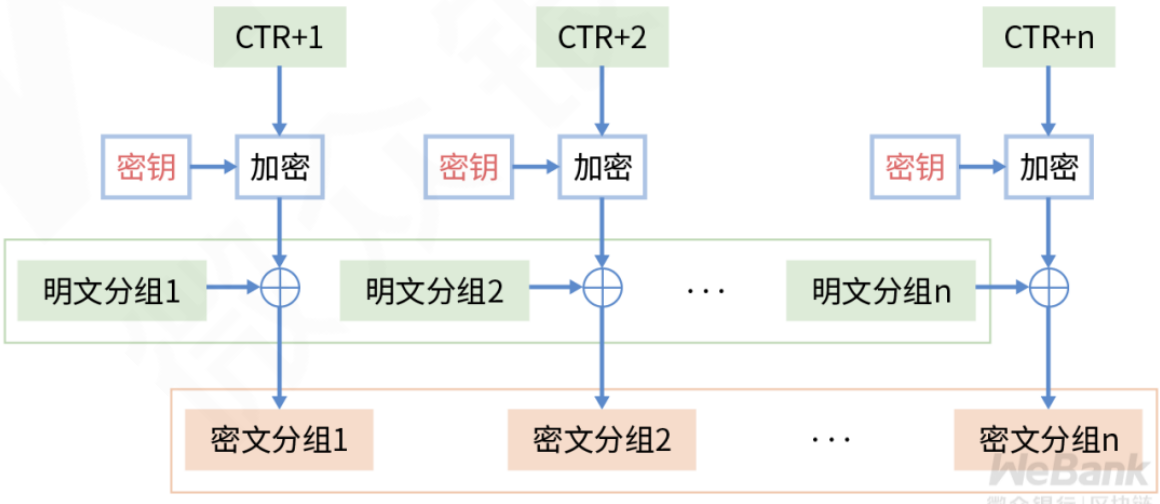


CBC模式解决了ECB模式的安全问题，但也带来了一定的性能问题。其主要缺点在于每个密文块都依赖于前面的所有明文块，导致加密过程是串行的，无法并行化。

CTR模式 (CounTeR)

CTR模式的出现让分组加密更安全且并行化，通过递增一个加密计数器以产生连续的密钥流，使得分组密码变为流密码进行加密处理，安全性更高。

CTR模式下的分组加密



CTR加密和解密过程均可以进行并行处理，使得在多处理器的硬件上实现高性能的海量隐

私数据的并发处理成为了可能，这是目前最为推荐的数据分组模式。

密码学协议中的数据分组与传统大数据处理中的数据分组有很大区别。理想情况下，数据分组不应该弱化隐私保护的强度，不能为攻击者获取未授权的信息提供可乘之机。这往往会涉及精心的数据分组方案设计，不能简单看作是数据分块之后的批处理。

业务应用难题：数据太短
工程实现之道：数据填充

数据太长是个问题，数据太短往往也是问题。

在以上分组处理的过程中，最后一个数据块中数据长度不足，密码学协议中的核心算法也可能无法工作。

假定一个密码协议处理的数据块长度要求为6字节，待加密的隐私数据长度为7字节。用两个十六进制数代表一个字节数据，其示例如下：

b1 b2 b3 b4 b5 b6 b7

7字节长于数据块的处理长度6字节，因此该数据将被分组，且可以分为两个数据块。分组示例如下：

第一个数据块：b1 b2 b3 b4 b5 b6

第二个数据块：b7

其中第一个数据块刚好是6个字符，第二个数据块只有1个字节，这个数据块就太短了，不满足处理要求。

针对以上问题，密码学工程实现中一般通过**数据填充**进行处理，即将短的数据块填充补位到要求的字节长度。示例中第二个数据块需要进行数据填充，为其补上缺少的5个字节。

与数据分组类似，这里的数据填充也不是普通的数据填充，也应该满足一定的安全性要求。最常用的数据填充标准是PKCS#7，也是OpenSSL协议默认采用的数据填充模式。

PKCS#7填充

需要填充的部分都记录填充的总字节数。应用于示例中第二个数据块，则补5个字节都是5的数据，其填充效果如下：

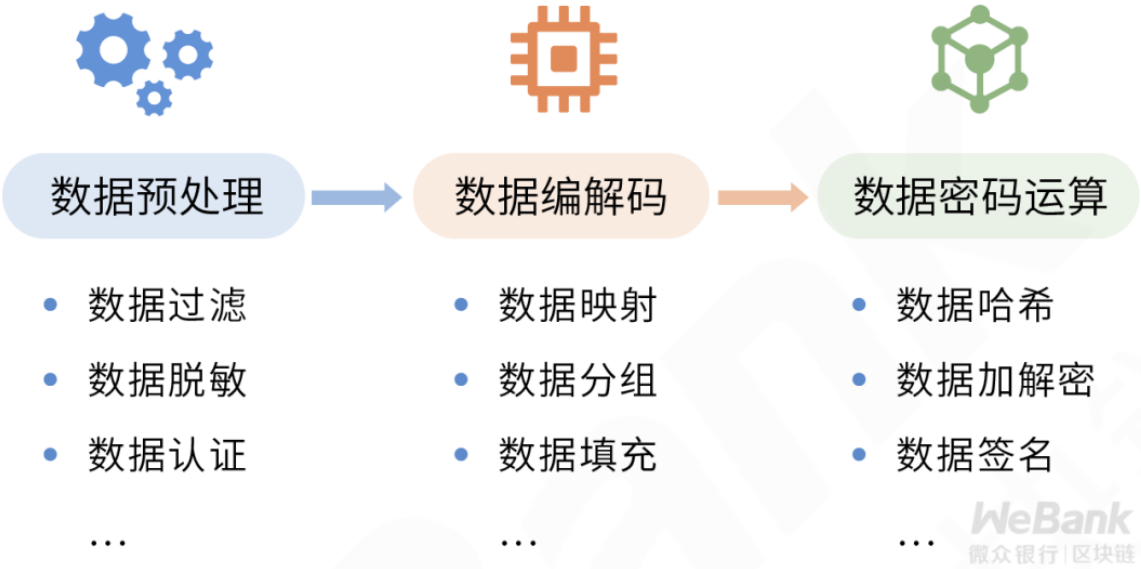
b7 05 05 05 05 05

这里还存在一个问题：如果一个隐私数据的最后一个分组，刚好就是一个符合其填充规则的数据，在事后提取原始数据时，如何分辨是原始数据还是填充之后的数据?

避开这种歧义情况的关键是，任何长度的原始数据，在最后一个数据块中，都要求进行数据填充。

值得注意的是，对隐私数据加密时，按特定填充模式进行处理，那么填充的数据也将被加密，成为加密前明文数据的一部分。解密时，其填充模式也需要和加密时的填充模式相同，这样才可以正确地剔除填充数据，提取出正确的隐私数据。

在隐私保护方案的编解码过程中，以上提到的数据映射、数据分组、数据填充，都是保证隐私数据安全的必要环节。此外，在特定的合规要求下，实际业务系统还需要引入更多的相关数据预处理环节，如数据脱敏、数据认证等，使得数据在进入密码学协议前，尽早降低潜在的隐私风险。



正是：理论公式抽象赛天书，工程编码巧手点迷津！

学术论文的公式符号与隐私保护方案的可用工程实现之间，存在一条不小的技术鸿沟，而密码学特有的数据编解码，正是我们建立桥梁实现学术成果产业转化的基石。

安全高效的数据编解码技术，对于处理以5G、物联网为爆点的海量隐私数据应用意义重大，是隐私数据进出业务系统的第一道防线，其重要性不亚于其他密码学原语。

了解完数据编解码之后，接下来将进入具体应用相关的密码学原语，欲知详情，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

上下滑动查看更多

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