> 本文主要记录隐私计算中的一些计算难题，这些难题衡量了加密方案的安全性，包括大数分解、离散对数以及格密码学中的最短向量（SVP）、最近向量（CVP）、带误差学习（LWE）等问题，**仅供参考**。

一、计算难题介绍

1. 简单定义

在密码学中，计算难题指的是一类在已知输入的情况下，**无法在合理时间内求解**的数学问题，一般是指无法在**多项式时间**内求解。这些问题在**理论上可以被解出**，但由于其所需的计算资源（时间、空间）随输入规模增长的速率超过多项式表示（一般为**指数增长**），在现实中几乎不可能破解，因此被称为计算困难问题。

2. 应用意义

正是这些计算难题的存在，使得基于这些难题的加密方案**难以被暴力破解**，从而保证了加密方案的安全性。同态加密是隐私计算中的一项代表技术，本文介绍的难题是同态加密技术的基础。

二、大数分解难题（Integer Factorization Problem，IFP）

> **大数分解问题**：给定一个正整数N，求其质因数分解。密码学应用中 N 通常为两个大素数 p, q 的乘积，问题相当于给定 $N = p \times q$，求解 p, q 的值。

该问题最符合直觉的做法就是**枚举** N 的质数因子，当 N 是合数时一定能找到一个小于 $\sqrt{N}$ 的质因子，因此算法的时间复杂度为$\mathcal{O}(\sqrt{N})$，随输入规模**呈指数级增长**，即$\mathcal{O}(2^{\frac{logN}{2}})$。关于什么是输入规模不再此详细介绍。

目前最优的一些解决大数分解问题的算法都无法保证在多项式时间内得到解，因此该问题是一个计算困难问题。

著名的 **RSA 加密算法**就是基于该难题构造的，如果该问题容易解决，那么RSA加密算法将不再安全。基于量子计算提出的 Shor 算法可以在多项式时间内解决该问题，这意味着 RSA 加密在量子计算中不再安全，因此后量子密码学也是目前的研究热点。

三、离散对数难题（Discrete Logarithm Problem，DLP）

> **离散对数问题**：在一个有限循环群 G 中，给定生成元 g 和元素 y，求解整数 x 使得 $g^x = y$。密码学应用中一般取 G 为模素数群 $\mathbb{Z}_p^*$，问题相当于求模意义下的对数。

> **椭圆曲线离散对数问题**：给定定义在某个有限域上的椭圆曲线 E，及其一个生成元 P，已知曲线上一个点 Q，求解标量 k 使得 $Q = kP$。

本文不对椭圆曲线密码学作过多的介绍。离散对数问题是很多密码学方案的基础，例如 Diffie-Hellman 密钥交换协议可以让两个参与方安全协商一个**共享密钥**，即使信道被监听也不会暴露隐私。

Diffie-Hellman 密钥交换协议在模素数群 $\mathbb{Z}_p^*$ 下的思想很简单，设参与方为 Alice 和 Bob，g 为群的一个生成元，则协议流程可以简单的描述为：

1. Alice 选择随机私钥 a，计算 $A = g^a$；
2. Bob 选择随机私钥 b，计算 $B = g^b$；
3. Alice 发送 A 给 Bob，Bob 发送 B 给 Alice；
4. Alice 和 Bob 计算共享密钥 $K = B^a = A^b = g^{ab}$。

显然信道传输的消息只有 A 和 B，根据离散对数难题可知监听方无法破解 a 和 b，因此也无法获取密钥 K。

从 Diffie-Hellman 密钥交换协议背后可以看出两个离散对数问题的衍生问题：**计算 Diffie-Hellman 问题**（CDH）和 **判定 Diffie-Hellman 问题**（DDH）。

> **计算 Diffie-Hellman 问题**：在一个有限循环群 G 中，给定生成元 g 以及 g^a 和 g^b，计算 g^{ab}。

> **判定 Diffie-Hellman 问题**：在一个有限循环群 G 中，给定生成元 g 以及 g^a、g^b 和 T，判定 $T = g^{ab}$ 是否成立。

显然如果离散对数问题解决，那么计算 Diffie-Hellman 问题就能解决，那么判定 Diffie-Hellman 问题就能解决，因此一般认为判定 Diffie-Hellman 问题难度相对较弱。在带**双线性映射**（pairing）的群中，判定 Diffie-Hellman 问题通常可以轻松解决，因此若使用该问题难解假设必须构造特殊的群。

除 Diffie-Hellman 密钥交换协议外，离散对数问题也是 **ElGamal 加密**算法的安全基础。

四、判断 N 次剩余问题（Deciding Composite Residuosity，DCR）

之前的文章中介绍了二次剩余的概念，N 次剩余也类似：在模 p 意义下，若 $a = x^n$，则称 a 为模 p 意义下的 n 次剩余。

> **判断 N 次剩余问题**：给定 $n = pq$（其中 p, q 为大素数）和一个数 $a \in \mathbb{Z}_{n^2}^*$，判断 a 是否是模 n^2 的 n 次剩余。

该问题目前没有多项式时间内的解法，**Paillier 加密方案**就采用了该问题难解的假设。

五、子群判定问题（Subgroup Decision Problem，SDP）

> **子群判定问题**：给定一个阶为 $n = pq$ 的群 G（其中 p, q 为大素数）和一个群元素 x，判断 x 是否属于阶为 p 的子群 G_p。

由大数分解难题可知难以获取 p, q 的值，该问题目前没有多项式时间内的方法解决。**BGN 加密方案**使用了该问题难解的安全假设。

六、格密码难题

在介绍一些格密码难题前首先需要了解一下格的概念，这里的格并非数学上序理论中的格，而是群论意义上的格，理解起来也很简单。由于很多格密码问题不依赖于传统密码学的假设（如上面提到的能够抵抗量子攻击，是后量子密码学的基础。

1. 格的介绍及相关属性

格密码中的格通常是指欧几里德空间中的一个无限**向量集合**，可以理解为“点集”，它由 n 个线性无关的向量张成：

给定 n 个**线性无关**的向量 $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathbb{R}^m$（通常 $m = n$），它们张成的格 $\mathcal{L}$ 的定义如下：

$$\mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_n) = \left\{ \sum_{i=1}^{n} z_i \mathbf{b}_i \mid z_i \in \mathbb{Z} \right\}$$

这是一个 **无限集合**，因为每个 z_i 都可以取**任意整数**，所以组合方式是无限多的，每一种组合都可以理解为一个坐标，因此格可以看作一个无穷点集。以群论的角度理解，格是一种加法阿贝尔群。

- 其中 n 为格 $\mathcal{L}$ 的 **维度** 或 **秩**。
- 这 n 个线性无关的向量组成的向量组称为格 $\mathcal{L}$ 的一组**基**。
- 每个基向量为一列可组成组成一个矩阵 $B \in \mathbb{R}^{n \times n}$，格 $\mathcal{L}$ 的 **行列式** 或 **体积** 为 $det(\mathcal{L}) = |det(B)|$。
- 格中的元素称为**格点**，本质是一个向量，具有加法阿贝尔群的性质。
- 格点本质是向量，因此有 **范数** 和 **距离** 的定义。一般范数取欧式范数，可理解为向量长度；距离取欧式距离。

2. 格问题

下面介绍一些格密码方案中依赖的一些常见难题（并不全面）。格密码难题常用于构造有限级数**全同态加密**方案，并通过 Bootstrapping 思想转换为全同态加密方案。

2.1 最短向量问题（Shortest Vector Problem，SVP）

> **最短向量问题**：求格 $\mathcal{L}$ 中欧式范数最小的非零向量（可理解为长度最短的非零向量）。

该问题看似简单，实际上目前还没有多项式时间内的解决方法。在格的维度较低时可以通过穷举解决，但穷举复杂度是指数级的，无法解决高维格中的 SVP 问题。有些算法可以在特定情况下以多项式的时间求解**近似最短向量**，即求解长度不超过最短向量若干倍的向量。

2.2 最近向量问题（Closest Vector Problem，CVP）

> **最近向量问题**：对于一个格 $\mathcal{L}$，任意指定一个向量 x（不要求在格中），求与其距离最近的格中向量 $y \in \mathcal{L}$。

上述问题也称为搜索 CVP 问题，该问题还有一些难度稍弱的变种，即**优化 CVP 问题**和**决策 CVP 问题**。优化 CVP 问题只需求得最短距离，而决策 CVP 问题只需要判断最短距离是否不大于某个数。

2.3 带噪声还原问题（Learning With Errors，LWE）

> **带噪声还原问题**：由模数 p 确定元素取值空间 $\mathbb{Z}_p$，对于一个矩阵 $\mathbf{A} \in \mathbb{Z}_p^{n \times m}$，秘密向量 $\mathbf{s} \in \mathbb{Z}_p^m$，噪声向量 $\mathbf{e} \in \mathbb{Z}_p^n$，有 $\mathbf{b} = \mathbf{A} \cdot \mathbf{s} + \mathbf{e} \pmod p$，其中噪声向量 $\mathbf{e}$ 采样于某种概率分布（如离散高斯分布）且其无穷范数很小。给定 $\mathbf{A}$ 和 $\mathbf{b}$，求解 $\mathbf{s}$。

LWE 问题这里翻译为带噪声还原问题（好像没有确定的翻译方式），该问题看上去只是一个线性代数的方程式，似乎和格没有什么关系。从**线性代数角度**看，当没有噪声向量时，方程式 $\mathbf{b} = \mathbf{A} \cdot \mathbf{s}$ 显然可以通过**高斯消元**求解，并不难解决。但正因引入了噪声向量 $\mathbf{e}$，使得问题的求解变得困难。

从**格的角度**看，可以把矩阵 $\mathbf{A}$ 每一列当作一个**基向量**张成一个格，而秘密向量 $\mathbf{s}$ 可以理解为**格点 $\mathbf{As}$** 的**坐标**，当引入一个很小的噪声向量 $\mathbf{e}$ 后得到的向量 $\mathbf{b}$ 可以看作离格点 $\mathbf{As}$ **非常近**的向量。那么 LWE 问题中：给定向量 $\mathbf{b}$ 和 格基向量矩阵 $\mathbf{A}$，求解秘密向量 $\mathbf{s}$，就可以转化为上面提到的**最近向量（CVP）问题**。相当于秘密向量 $\mathbf{s}$ 是向量 $\mathbf{b}$ 在由基向量矩阵 $\mathbf{A}$ 张成的格上的**最近向量**。当 CVP 问题难解时，LWE 问题也是难解的。

上面的理解方式可能存在一些问题，因为矩阵 $\mathbf{A}$ 的所有列向量不一定能作为格的基向量（**有可能线性相关**）。而在实际应用中，矩阵 $\mathbf{A}$ 通常是**随机生成**的，因此基本可以认为其列向量是线性无关的。（补充说明：作者没有仔细研究过，因此不了解矩阵 $\mathbf{A}$ 的列向量如果线性相关，或行数远大于列数的情况是否会导致该问题变得容易解决。）

上述问题也称为搜索 LWE 问题（SLWE），类似地也存在对应的**决策 LWE 问题**（DLWE），即**判断**向量 $\mathbf{b}$ 是否为 LWE 问题中由噪声掩盖的线性变换生成的向量，还是仅仅为一个随机生成的向量。通常在以 LWE 问题为假设的加密方案**安全证明**时会用到 DLWE 问题的困难性。LWE 问题是许多全同态加密方案（如 GSW 方案）的基础，此外在许多方案中会把 LWE 问题扩展至**多项式环**中，称为 **Ring-LWE** 问题。通过将元素表示为多项式并在多项式环中进行运算，能够实现传统 LWE 难以实现的效率和灵活性。

2.4 短整数解问题（Short Integer Solution，SIS）

> **短整数解问题**：由模数 p 确定元素取值空间 $\mathbb{Z}_p$，给定一个矩阵 $\mathbf{A} \in \mathbb{Z}_p^{n \times m}$，求解一个向量 $\mathbf{x}$，满足$\mathbf{A} \cdot \mathbf{x} \equiv \mathbf{0} \pmod p$，且向量 $\mathbf{x}$ 的范数（通常为2-范数或∞-范数）不超过某个给定的界。

类似地，该问题可以把矩阵 $\mathbf{A}$ 当作格中基向量组成的矩阵。不正式地说，该问题相当于在某个特定格（满足$\mathbf{A} \cdot \mathbf{x} \equiv \mathbf{0}$ 方程）中**寻找最短向量**。该问题被证明与最短向量问题（SVP）在最坏情况下具有等价的难度。

SIS 问题可以应用于哈希函数、数字签名、零知识证明等领域，类似于 LWE 问题，该问题也可以扩展至多项式环。

本文为作者在学习相关知识时的一种记录，便于以后的回顾。作者并没有系统地学习过密码学，因此在表述上可能会存在不严谨甚至出错的地方，文章仅供参考，欢迎大家与我交流，一起进步！

其他平台：

- 知乎（Totoro）：https://www.zhihu.com/people/totoro-14-60
- CSDN（_Totoro_）：https://blog.csdn.net/orz_Totoro
- B站（Totoro_134）：https://space.bilibili.com/279377771
- Github（Totoro134）：https://github.com/Totoro134
- 公众号（知识长生所）