第六届（2021年）

全国高校密码数学挑战赛

参赛论文

赛题一：环上序列的截位还原

战队名称 Se\_P3t

学 校 杭州电子科技大学

学 院 通信工程学院

专 业 通信工程

参赛选手 方佳乐

指导教师 钟华

二零二一年六月

摘 要

序列密码由于其高效性在密码学中被广泛研究和应用. 对于一个密码学安全的伪随机序列而言, 除了需要满足统计学上的伪随机外, 它还需要具备一定的不可预测性, 即给出一段输出和必要参数, 还原生成器的初态是一个NP问题.

在本次密码学数学挑战赛中, 我们研究了环上序列的截位还原问题. 基于格基约化算法, 使用 python 语言 / SageMath 软件调用开源库 fpylll [3] 与 g6k [1], 我们取得了以下成果:

对于第一类挑战, 在已知模数和本原多项式的条件下, 我们使用 BKZ 算法和 G6K 求解 approx-SVP, 在数分钟内解决第1-6 级挑战; 对于第7级挑战, 我们使用 G6K - GPU Tensor [4] 中实现的 gpu\_sieve 也在两小时以内还原了初态.

对于第二类挑战, 给出模数和级数, 受 Sun. 等人[14]的研究启发, 我们提出一种基于零空间的新解法, 能够通过几次BKZ调用并在数秒内分别完成所有9级挑战.

对于第三类挑战, 只给出了模数的比特长度以及级数, 我们采用第二类挑战中的算法直接求出初态, 进而求出模数和本原多项式. 但是由于格的维数过大, 对于一台普通个人笔记本, 只能在数分钟内完成前3级挑战.

**关键词**：线性递归序列，整数剩余类环，截位序列，序列还原，格基约化算法，最短向量问题

目 录

[第一章 引言 1](#_Toc75940161)

[1.1 背景 1](#_Toc75940162)

[1.2 赛题分析 1](#_Toc75940163)

[1.3 解题方法与结果 2](#_Toc75940164)

[第二章 预备知识 3](#_Toc75940165)

[2.1 格的基本概念和性质 3](#_Toc75940166)

[2.2 格中困难问题 4](#_Toc75940167)

[2.3 格基约化算法 4](#_Toc75940168)

[2.3.1 LLL 算法 4](#_Toc75940169)

[2.3.2 BKZ算法 5](#_Toc75940170)

[第三章 主要结果 6](#_Toc75940171)

[3.1 初步分析 6](#_Toc75940172)

[3.2 第一类问题 7](#_Toc75940173)

[3.3 第二类问题 8](#_Toc75940174)

[3.3.1 Sun等人的解法 9](#_Toc75940175)

[3.3.2 基于零空间的新解法 10](#_Toc75940176)

[3.4 第三类问题 11](#_Toc75940177)

[第四章 实验结果 12](#_Toc75940178)

[4.1 第一类挑战 12](#_Toc75940179)

[4.2 第二类挑战 12](#_Toc75940180)

[4.3 第三类挑战 13](#_Toc75940181)

[第五章 参考文献 14](#_Toc75940182)

[第六章 谢辞 15](#_Toc75940183)

[第七章 附录 16](#_Toc75940184)

[7.1 各类挑战使用的运行参数及输出 16](#_Toc75940185)

[7.1.1 第一类挑战 16](#_Toc75940186)

[7.1.2 第二类挑战 19](#_Toc75940187)

[7.1.3 第三类挑战 24](#_Toc75940188)

[7.2 求解结果 26](#_Toc75940189)

[7.2.1 第一类挑战 26](#_Toc75940190)

[7.2.2 第二类挑战 28](#_Toc75940191)

[7.2.3 第三类挑战 29](#_Toc75940192)

# 引言

## 背景

伪随机序列生成器为序列密码提供了随机性, 需要具备一定的不可预测性. 其中, 线性同余生成器是一个典例. 基于线性同余的截位序列早在上世纪被Stern[13]证明不是密码学安全的.

随后, 线性同余生成器的一般化类型, 线性递归生成器, 也被广泛研究.

**定义1 (环上序列的截位还原问题).** 对于一个在整数剩余类环 上的 阶本原多项式 , 其对应的线性递归生成器定义为

其中 为生成器的内部状态. 令 且 , 那么该生成器的截位高 比特的输出为 . 环上序列的截位还原问题定义为给出生成器的一段长为 的连续的截位输出 , 求出初态 .

Sun等人 [14] 基于 Stern [13] 对于线性同余生成器的攻击方法, 将其拓展至高阶本原多项式, 给出了环上序列截位还原问题的解决方法.

## 赛题分析

本届挑战赛赛题一主要围绕环上序列的截位还原问题, 其中包含三类挑战, 每类挑战分为9级. 三类挑战类似Sun等人 [14] 的主要研究结果: 未知初态, 未知本原多项式和未知模数.

对于第一类挑战, Sun等人 [14, Section 3.5] 将生成器内部状态之间的关系转换成一组线性同余方程组, 并考虑使用Frieze等人 [5] 的理论求解初态. 虽然题目不全满足他们给出的求解条件, 但是我们可以通过使用其他格基约化算法 (如 BKZ 算法) 来优化条件. 另外, 这组同于方程组可以看作一个多元隐藏数问题 [7], 我们还考虑了基于傅里叶分析的方法求解. 对于一元的情形, 即最初的截位线性同余生成器, 使用第9级的参数(截取位数, 模数及输出个数), 我们确实能在很短的时间内一次还原部分高位比特; 但是对于多元的情形, 由于隐藏数的乘数的选取属于一特定子群, 我们无法使用此类方法还原[6, Section 6.2].

对于第二类挑战, 题目还未给定内部状态之间的线性关系. 根据前三级问题给出参考参数取值, 我们很容易就可以发现截为输出个数恰等于参数 , 之和, 由此可以推出给定条件无法满足 Sun 等人提出的方法的基本条件: . 这也意味着我们需要尝试优化他们的解法或者提出新的解法.

对于第三类挑战, 题目未明确给出模数, 而只给出了模数的比特长度. 仅仅观察给出的输出个数也能够感受到格基维数的巨大, 这对算力有很大的要求. 若非提出足够大的优化方法, 对于一台普通的个人笔记本电脑而言, 就只能经历而为了.

## 解题方法与结果

对于第一类挑战, 我们将其转换为SVP问题,使用 BKZ 算法和 G6K 求解 approx-SVP, 在数分钟内解决第1-6 级挑战; 对于第7级挑战, 我们使用 gpu\_sieve 也在两小时以内还原了初态.

对于第二类挑战, 我们提出了一种基于零空间的新解法, 较 Sun. 等人[<sun2020>]的解法快, 能够通过几次BKZ调用并在数秒内分别完成所有9级挑战.

对于第三类挑战,我们采用第二类挑战中的算法直接求出初态, 进而求出模数和本原多项式. 但是由于格的维数过大, 对于一台普通个人笔记本, 只能在数分钟内完成前3级挑战.

# 预备知识

所有的向量均用小写粗体字母表示, 并且默认为行向量. 矩阵使用大写粗体字母表示. 表示向量 的欧几里得范数.

## 格的基本概念和性质

**定义2 (格).** 令 为 内的一组线性不相关(行)向量(). 那么由 生成的格为 的整系数线性组合, 即

其中向量 称为格的基, 格的秩为 , 且格的维数为 . 如果 , 那么 是一个满秩格.

格 的基矩阵 是将行作为基向量 形成的 矩阵. 因此 为行向量 的第 个元素, 且

格  的行列式是 的任一基矩阵 的基本平行六面体的体积, 即

当格 满秩时, 其行列式 .

**定义3 (逐次最小值).** 令 是一个秩为 的格. 的逐次最小值为 使得, 对于 , 是满足存在 个线性无关的向量 且 的最小值. 特别的, 是 的最短非零向量长度.

**定义4 (Minkowski定理).** 令 是一个秩为 的格, 则其逐次最小项 满足

**推论4.1.** .

**定义5 (高斯启发式 [**10**]).** 高斯启发式指出, 在 中维数为 的一个随机格的最短非零向量长度期望为

其中, 表示半径为 的 -维欧几里得球.

**定义6 (Hermite根因子).** 对于维数为 的格的一个基矩阵 , Hermite根因子定义为

## 格中困难问题

令 是 中的一个格.

**定义7 (最短向量问题).** 给出格 的一个基矩阵 , 求出一个非零向量 使得它的欧几里得范数 是最小的, 即 .

**定义8 (最近向量问题).** 给出格 的一个基矩阵 和一个向量 , 求出一个非零向量 使得 **||** 是最小的.

**定义9 (近似最短向量问题).** 给定 , 给出格 的一个基矩阵 , 求出一个非零向量 使得 .

## 格基约化算法

格基约化的目的是将给定的格基转换为由短且接近正交的向量组成的”优质”基. 要实现这一点, 既需要一个合适的”优质”基的定义, 也需要一种计算满足该定义的基的有效算法.

### LLL 算法

1982 年 Lenstra 等人提出了 LLL 算法 [9], 这在实际应用中是一个非常重要的算法.

**定义10 (LLL-规约基).** 令 是格的一组基. 设其对应的Gram-Schmidt正交基为 , 并记 . 对于 令

为Gram-Schmidt过程中的系数. 给定约化参数 . 我们称 为LLL-规约基当且仅当下列条件满足:

* (size条件) 对于 有 .
* (Lovász条件) 对于 有 .

对于一组LLL-规约基, 它满足一下性质.

**定理1.** 设 , ,将格 的任一基矩阵 作为LLL算法的输入, 则其输出的格基 满足:

1. 对于 , 有
2. .

### BKZ算法

1994年, Schnorr和Euchner [11] 提出了一个更加实用的算法 BKZ. 该算法基于KZ基, 具体定义参考 [12] ,这里我们不再赘述. 给定一个维数为 的随机格, 对于 block-size 为 的 BKZ 算法输出的最短向量 的欧几里得范数满足

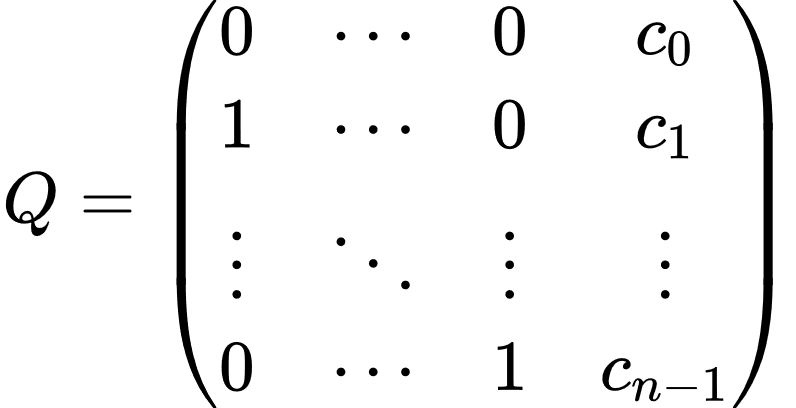
其中 .

# 主要结果

已知在整数剩余类环 上的 阶本原多项式 , 给出其对应的线性递归生成器的一段长为 的连续的截位高 比特输出 , 我们讨论在三种不同已知条件下还原初态 : 未知初态, 未知本原多项式和未知模数.

## 初步分析

已知

令

则有

那么设 的第一列为 , 其中 是关于 的整系数多项式, 我们得到内部状态的线性关系

令上式中 , 有任一状态 关于初态 的线性同余关系式:

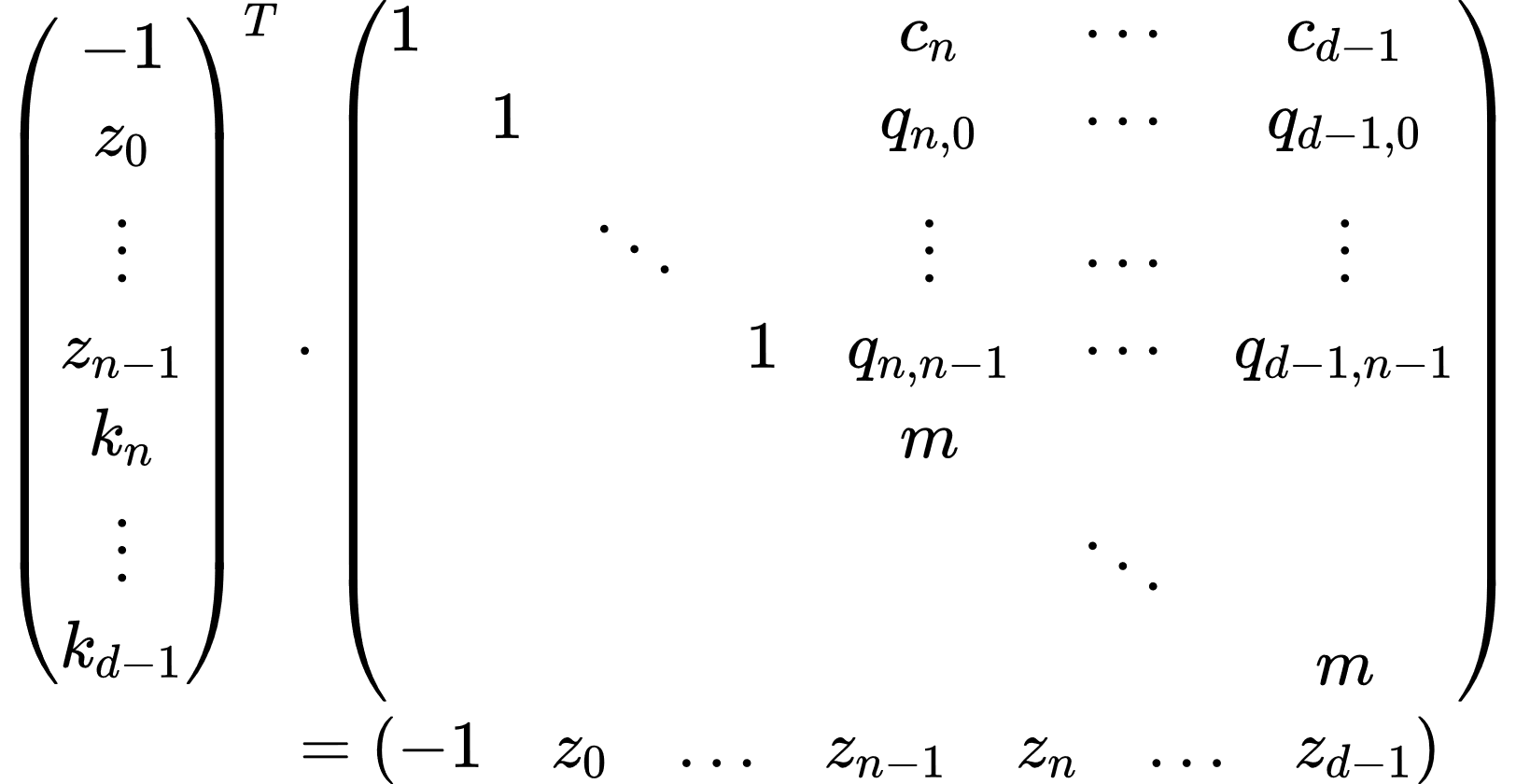
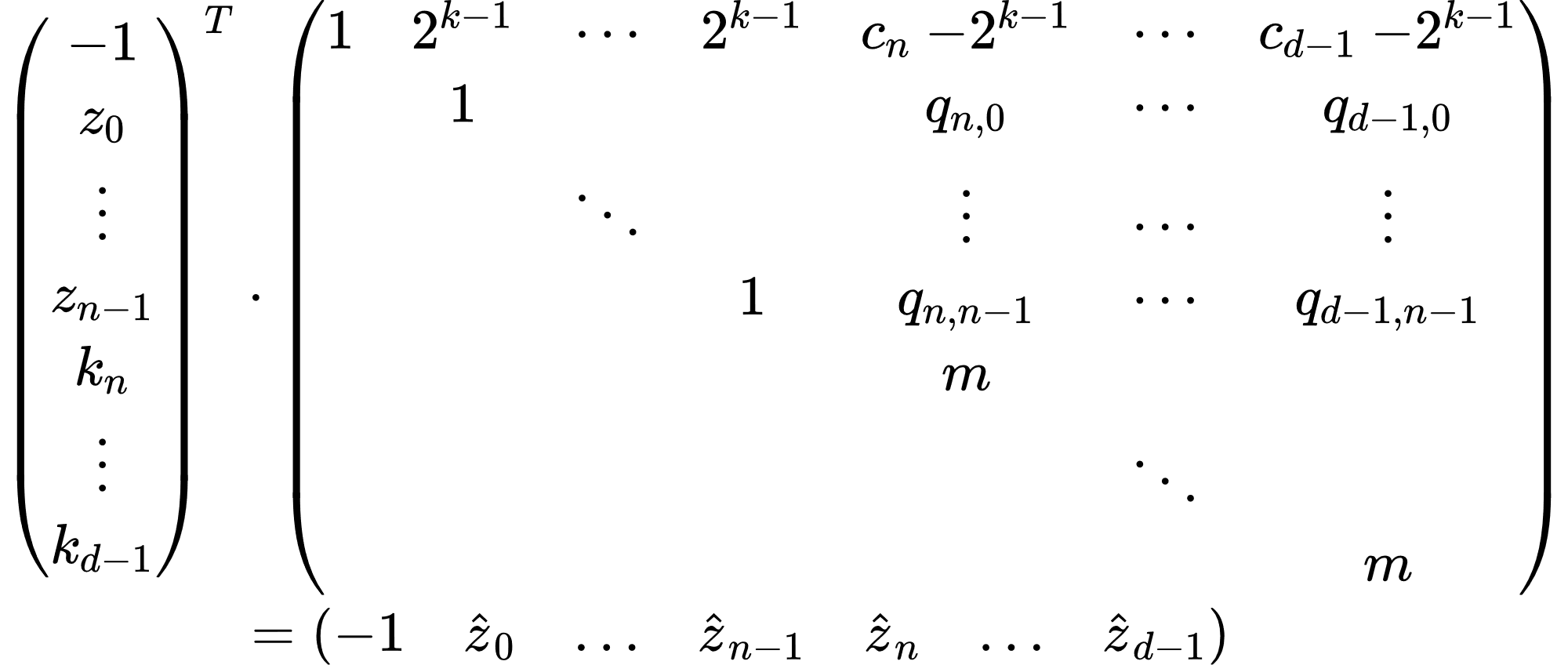
现给出截位高 比特输出 , 即有 , 代入并整理得

其中 .

## 第一类问题

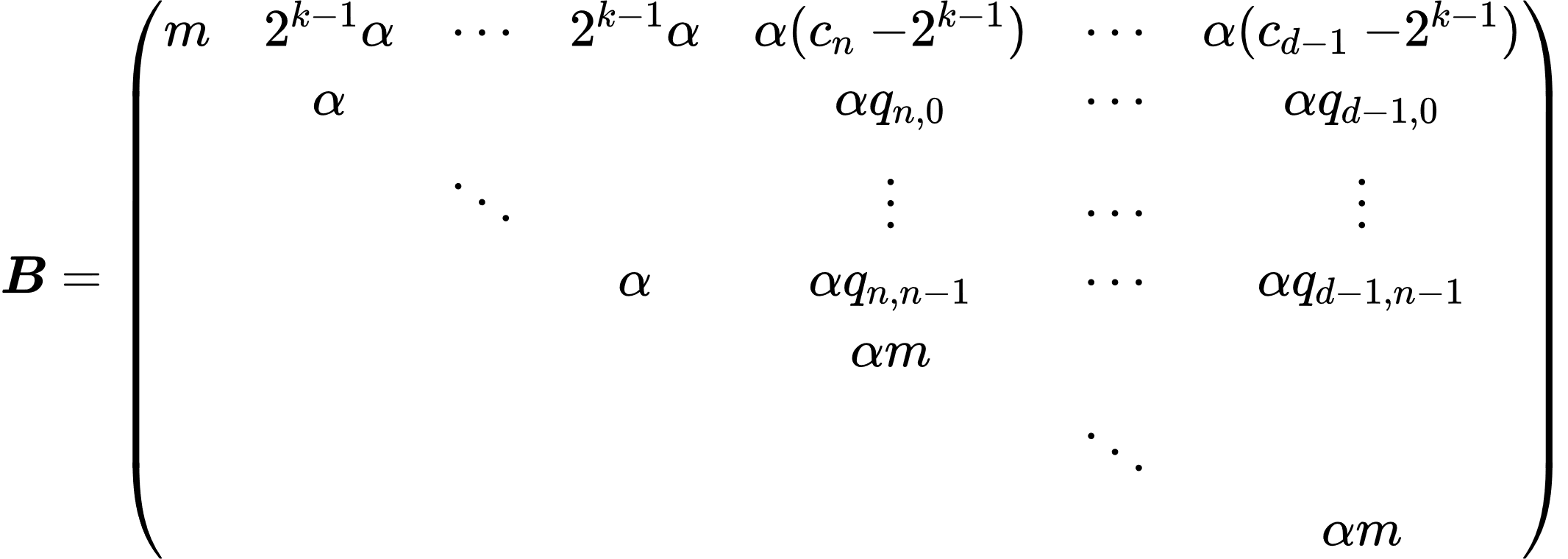
对于第一类环上序列截位还原问题, 我们除了生成器的内部状态未知, 其他参数均给出, 也就是说我们只需求出初始内部状态的低 比特 . 所有九级挑战的生成器的本原多项式以及模数相同 ().

显然, 关于 的线性同余关系能够转换成 CVP 问题, 我们可以应用Kannan的嵌入技术 [8] 将其转换成 SVP 问题. 整理得:

在讨论格基的放缩因子之前, 我们注意到未知变量 均大于等于零, 因此我们可以先进行换元以减少目标向量的长度一比特. 令 , 则有 .

随后, 根据目标向量预期大小, 我们可以取放缩因子为 , 其中 . 最终我们得出目标向量

以及基矩阵



设基矩阵 定义的格为 , 由于其行列式

根据推论4.1, 格 中的最短向量的欧几里得范数至多为

为了能够使用格基约化算法快速发现目标向量 , 一个必要条件为 的长度需要显著小于格中其他向量. 因此, 当 , 即

时, 理论上能够发现目标向量的概率不大 (注意到Martin等人 [2] 利用额外的predicate检验来突破这个界限, 但是由于算力限制, 这里我们不考虑). 在实际应用中, 这个界限要宽松些.

最终, 我们只需要求解一个SVP问题. 首先, 我们对基矩阵进行LLL规约. 然后对于较小维数的基, 我们直接完成HKZ; 如果维数过大的话, 我们先完成一个小block-size 的BKZ过程, 然后使用G6K [1] (或其GPU实现 [4] ) 筛出目标向量.

## 第二类问题

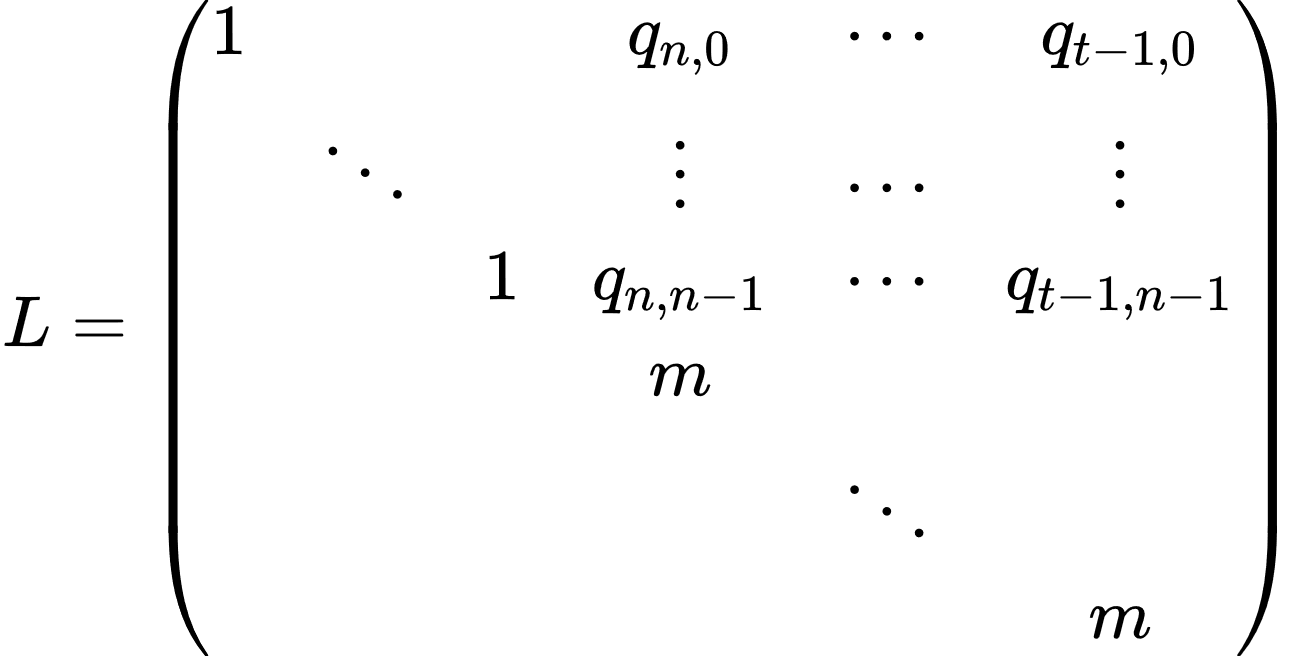
对于第二类问题, 给定某一阶数的本源多项式, 我们仅知道其模数而不知道具体系数. 在本节中, 我们首先回顾Sun等人 [14, Section 3.4] 的解法并分析其在此类问题中的限制, 随后我们提出一种新的解法, 能够在更短的时间内求出结果.

### Sun等人的解法

由于本原多项式系数未知, 此解法的关键在于利用生成器的输出导出关于内部状态的新的线性关系. 令

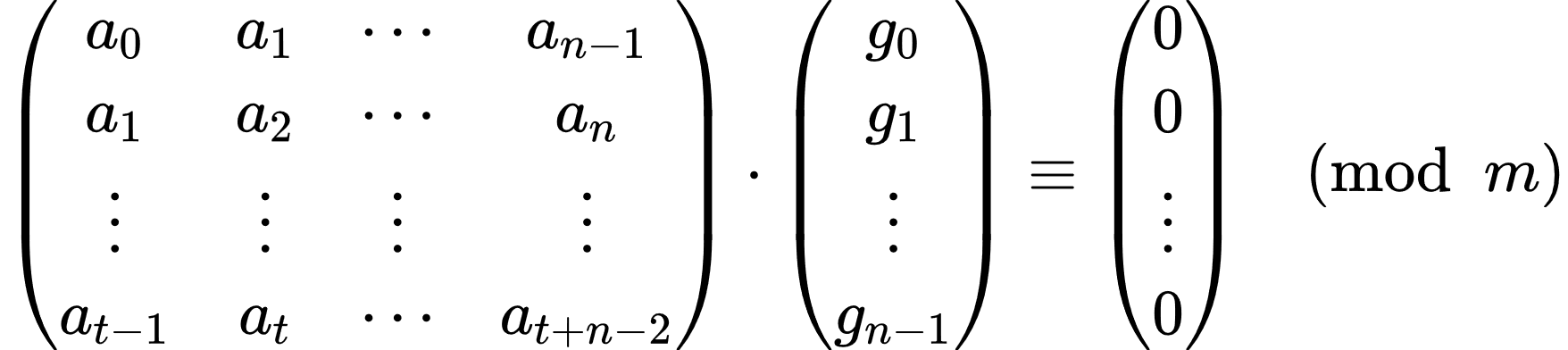
其中参数 , 的取值待定. 由这些向量构成的 矩阵 的维数至多为 , 当 远大于 时, 一般来说, 其维数恰为其最大值. 那么 的核 的维数应为 , 即存在向量 使得 .

令

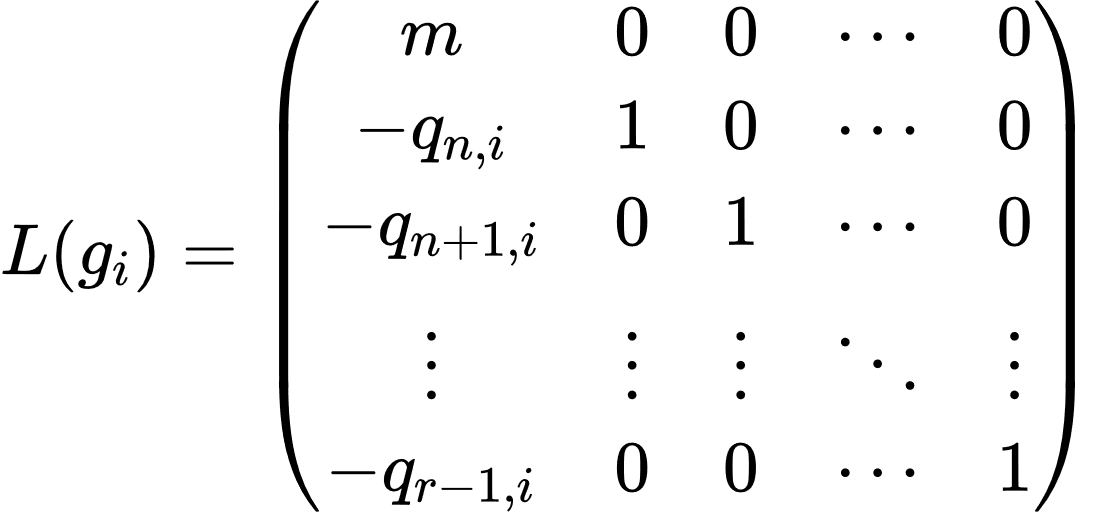
则有 , 其中 . 注意到任意 均属于同一个格, 记为 , 则

显然, . 如果 足够小使得 , 那么 .

假设现已求出 满足 , 即对于 , 有

结合[内部状态的线性关系](#内部状态的线性关系), 可以推出

其中 . 当等式左边关于 的 矩阵中 取值合适时, 它的零空间为空, 故有 . 也就是说, 对于任一 , 存在一个整数 使得

亦即 , 其中维数为 的格

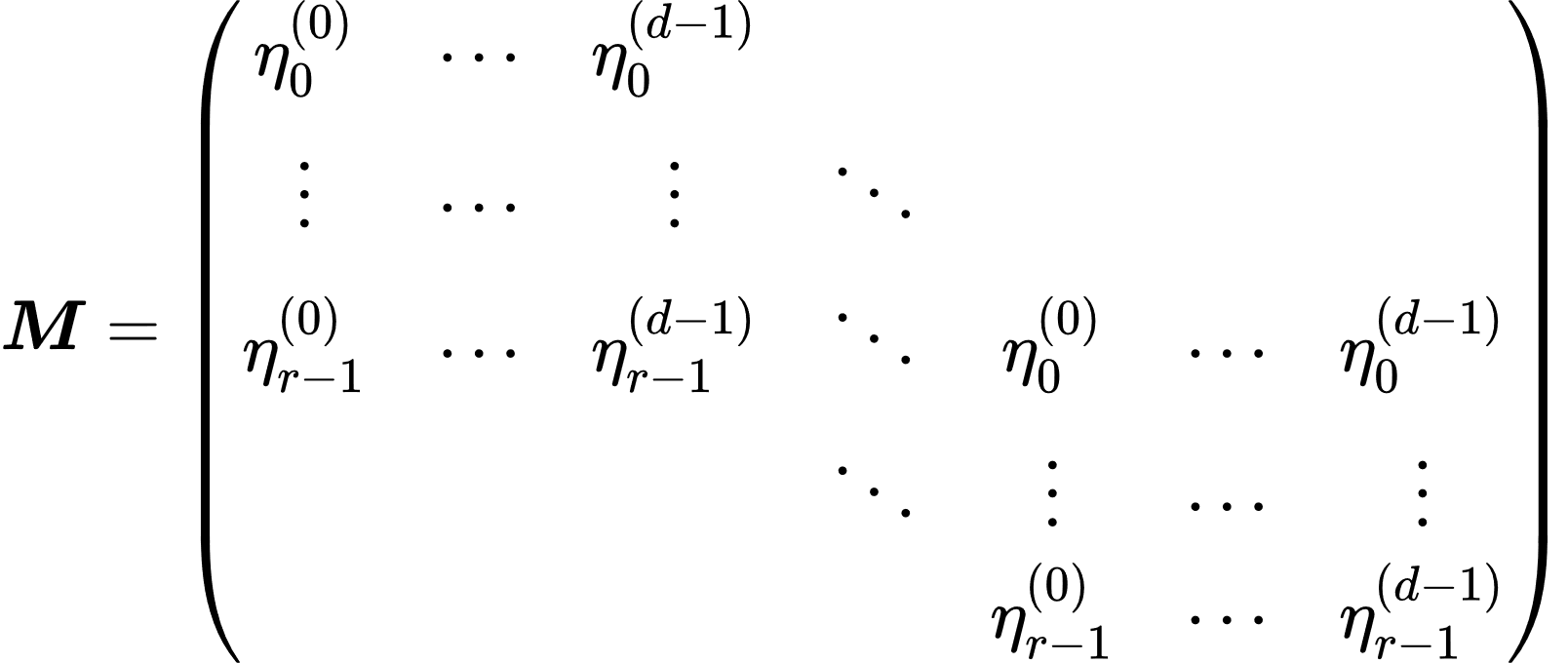
这个解法的最后一个步骤为使用不同偏移的连续输出, 找出 个不相关的满足条件的 , 即还原出 , 最终求其子格获得每个 .

接下来我们讨论本解法在解题中的限制. 根据挑战数据前三级给出的参考参数取值, 我们发现给出的输出个数 恰等于 . 那么使用参考值仅能构成一组核 , 其中至多含有 个向量 满足 . 由于 , 所以我们无法获得足够多的向量生成格 .

在提出新解法之前, 我们尝试了去优化该解法. 假设我们能够获得两个满足条件的 , 我们可以求出 个关于 的模多项式 . 随后, 我们在整数上对 个多项式两两计算结式, 最终获得两个关于某一 的单变量多项式 , . 根据结式的计算, 易知 , 根据实验我们发现, 在模 的剩余环上, 大概率有 . 对于其他系数 , 我们可以先将 的值代入, 然后类似之前解法迭代求出剩余系数. 但是, 当本原多项式的阶数以及参数的增大, 多项式 的项数也不断增多, 加大了求解结式的难度. 经过测试, 我们无法在几小时内解出第三级挑战.

### 基于零空间的新解法

假设我们已经获得了关于内部状态的新的线性关系, 即对于 , 有

对于 组符合条件的 , 我们构造如下 的矩阵

注意到向量 和向量 为两个共同属于矩阵 的左核 的线性不相关的向量. 一般的, 当 时, 的零空间维数恰为 .

记 生成的格为 , 并设格 经过LLL规约后的维数为 的格基为 . 设两向量最大元素的比特长度分别为 和 , 我们分如下四种情形讨论:

如果 , 一般的, 即当 时, 的最短向量即为所求的 ;

当 时, 我们对两向量进行代换(保证长度短且不相关), 取 , , 我们有, , 故 , 我们仍可以直接获得初始状态;

当 时, 有 , , 我们分别对代换前后进行LLL规约, 由于LLL输出接近正交的向量, 我们可以得到 , 和 , . 给最后一个向量加上 得到 , 注意到

所以我们可以求出 , 最终获得 ;

最后, 当 恰等于 时, 我们可以尝试其他代换或者试试运气. 由于挑战中并未出现该情形, 这里不做进一步讨论.

已知一定长度()的内部状态后, 我们可以在模 的剩余环上构建 个[等式](#内部状态递归关系), 并求解出 以及对应的本原多项式的系数.

## 第三类问题

在第三类问题中, 关于模数, 我们只知道其比特长度. 注意到第二类问题中提出的新解法在求解初始状态并没有使用到模数, 因此我们可以搬用之前的解法. 多次计算GCD直到达到给定的比特长度得到模数, 然后求解系数即可.

# 实验结果

本章节我们给出挑战赛的解题所需时间 (详细数据参见[附录](#_附录)). 如无特别说明, 所有的实验均在一台个人笔记本电脑上完成 (Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz). 我们主要使用 python 语言 / SageMath 软件调用开源库 fpylll [3] 与 g6k [1] 进行求解 (具体代码已在GitHub上开源).

## 第一类挑战

对于第一级和第二级挑战, 我们直接完成HKZ规约, 在一秒内求得结果. 对于第三级挑战, 我们使用30条输出, 并进行BKZ-20规约, 另外对于第四级和第五级, 我们取block-size为40, 均在数秒内完成. 对于第六级, 我们先进行BKZ-30规约, 随后使用G6K在三分钟内筛出目标向量. 对于第七级挑战, 利用10线程CPU 筛法, 连续运行十小时我们也无法获得正确的向量, 最终利用GPU筛法, 在一块Tesla T4上运行两小时左右, 我们求得了结果.

具体解题时间如下:

|  |  |  |  |
| --- | --- | --- | --- |
| 级数 | 求解方法 | 格基维数 | 解题时间 |
| 1 | HKZ | 19 | <1s |
| 2 | HKZ | 20 | <1s |
| 3 | BKZ-20 | 31 | ≈1s |
| 4 | BKZ-40 | 52 | ≈1s |
| 5 | BKZ-40 | 74 | 5.39s |
| 6 | BKZ-30 + Sieve | 121 | 2m13s |
| 7 | BKZ-20 + GPU Sieve | 151 | 99m4s |

## 第二类挑战

在第二类挑战中, 我们提出了新的解法, 能够利用一两次BKZ过程在数秒内求解出所有挑战, 具体解题时间如下:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 级数 | 阶数 | zbits (mbits=31) | r | t | BKZ: block-size | 解题时间 |
| 1 | 2 | 17 | 30 | 8 | 20 | 2.35s |
| 2 | 2 | 23 | 60 | 15 | 20 | 2.9s |
| 3 | 3 | 21 | 68 | 17 | 20 | 3.2s |
| 4 | 4 | 21 | 95 | 25 | 30 | 7.8s |
| 5 | 5 | 18 | 85 | 23 | 30 | 6.7s |
| 6 | 8 | 11 | 90 | 20 | 20 | 4.4s |
| 7 | 10 | 11 | 110 | 26 | 20 | 9.3s |
| 8 | 12 | 8 | 110 | 28 | 20 | 4.65s |
| 9 | 14 | 8 | 128 | 32 | 32 | 10.0s |

## 第三类挑战

对于三类挑战, 由于维数过大和时间限制, 我们只能解出前三级挑战:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 级数 | 阶数 | mbits | zbits | r | t | 解题方法 | 解题时间 |
| 1 | 16 | 31 | 5 | 140 | 30 | BKZ-20 | 30.8s |
| 2 | 16 | 31 | 10 | 190 | 40 | BKZ-30 | 1m14.8s |
| 3 | 16 | 31 | 14 | 265 | 70 | BKZ-30 + sieve-10-threads | 5m24.0s |

# 参考文献

[1] Martin R. Albrecht, Léo Ducas, Gottfried Herold, Elena Kirshanova, Eamonn W. Postlethwaite, and Marc Stevens. The general sieve kernel and new records in lattice reduction. *IACR Cryptol. ePrint Arch.*, 2019:89, 2019.

[2] Martin R. Albrecht and Nadia Heninger. On bounded distance decoding with predicate: Breaking the "lattice barrier" for the hidden number problem. *IACR Cryptol. ePrint Arch.*, 2020:1540, 2020.

[3] The FPLLL development team. fpylll, a Python wraper for the fplll lattice reduction library. Available at https://­github.com/­fplll/­fpylll, 2021.

[4] Léo Ducas, Marc Stevens, and Wessel P. J. van Woerden. Advanced lattice sieving on gpus, with tensor cores. *IACR Cryptol. ePrint Arch.*, 2021:141, 2021.

[5] Alan M. Frieze, Johan Håstad, Ravi Kannan, J. C. Lagarias, and Adi Shamir. Reconstructing truncated integer variables satisfying linear congruences. *SIAM J. Comput.*, 17(2):262–280, 1988.

[6] Steven D. Galbraith, Joel Laity, and Barak Shani. Finding significant fourier coefficients: Clarifications, simplifications, applications and limitations. *IACR Cryptol. ePrint Arch.*, 2016:682, 2016.

[7] Steven D. Galbraith and Barak Shani. The multivariate hidden number problem. *IACR Cryptol. ePrint Arch.*, 2015:111, 2015.

[8] Ravi Kannan. Minkowski’s convex body theorem and integer programming. *Math. Oper. Res.*, 12(3):415–440, 1987.

[9] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. 261:515–534, 1982.

[10] Phong Q. Nguyen and Damien Stehlé. LLL on the average. In Florian Hess, Sebastian Pauli, and Michael E. Pohst, editors, *Algorithmic Number Theory, 7th International Symposium, ANTS-VII, Berlin, Germany, July 23-28, 2006, Proceedings*, volume 4076 of *Lecture Notes in Computer Science*, pages 238–256. Springer, 2006.

[11] Claus-Peter Schnorr and M. Euchner. Lattice basis reduction: Improved practical algorithms and solving subset sum problems. *Math. Program.*, 66:181–199, 1994.

[12] C.P. Schnorr. A hierarchy of polynomial time lattice basis reduction algorithms. *Theoretical Computer Science*, 53(2):201–224, 1987.

[13] Jacques Stern. Secret linear congruential generators are not cryptographically secure. In *28th Annual Symposium on Foundations of Computer Science, Los Angeles, California, USA, 27-29 October 1987*, pages 421–426. IEEE Computer Society, 1987.

[14] Hong-Yu Sun, Xuan-Yong Zhu, and Qun-Xiong Zheng. Predicting truncated multiple recursive generators with unknown parameters. *Des. Codes Cryptogr.*, 88(6):1083–1102, 2020.

# 谢辞

感谢方佳乐组员对本组Se\_P3t所做的突出贡献; 感谢其室友黄安等人对我的陪伴, 在这期末考试的关键时刻, 是他们与我一切熬夜, 一起复习; 最后, 感谢杭电通信信息安全社团协会Vidar-Team以及指导老师钟华在我学习密码学初期起着关键作用的引导和支持.

# 附录

## 各类挑战使用的运行参数及输出

### 第一类挑战

#### 第一级



#### 第二级



#### 第三级



#### 第四级



#### 第五级



#### 第六级



#### 第七级



[日志文件](https://drive.google.com/file/d/1RffrzXly7PzOhaBrUOkM8JUoy2RjNe4D/view?usp=sharing): https://drive.google.com/file/d/1RffrzXly7PzOhaBrUOkM8JUoy2RjNe4D/view?usp=sharing

### 第二类挑战

#### 第一级



#### 第二级



#### 第三级



#### 第四级



#### 第五级



#### 第六级



#### 第七级



#### 第八级



#### 第九级



### 第三类挑战

#### 第一级



#### 第二级



#### 第三级



## 求解结果

### 第一类挑战

#### 第一级



#### 第二级



#### 第三级



#### 第四级



#### 第五级



#### 第六级



#### 第七级



### 第二类挑战

#### 第一级



#### 第二级



#### 第三级



#### 第四级



#### 第五级



#### 第六级



#### 第七级



#### 第八级



#### 第九级



### 第三类挑战

#### 第一级



#### 第二级



#### 第三级

