

一种基于 LWE 采样算法的实现与优化

王柯翔,黎琳,彭双和

(北京交通大学 计算机与信息技术学院,北京 100044)

摘 要:基于带错误学习问题(Learning With Errors,LWE)构造的密码体制能够抵御量子攻击,它的应用效率与 LWE 问题的采样过程密切相关.而在 LWE 问题采样中,对其中的错误因子(Error Factor)采样占采样过程绝大部分时间,本文对 LWE 问题中的错误因子的采样算法进行研究,将在高斯分布上效率较高的金字塔(Ziggurat)采样算法,应用到了一种高效的 LWE 问题采样算法中.基于在连续域上的采样比离散域上采样效率高的思路,对 LWE 问题采样算法在离散域上采样的过程进行了优化,提出了一种将连续域上的采样结果进行取整的方法,对优化前后的两种 LWE 问题的采样算法进行了对比实验,结果表明:改进后的算法在不占用大量内存并且保证安全性的情况下,将采样速度提高了 38%~200%.

关键词:格;带错误学习问题;高斯分布;错误因子;采样

中图分类号:TP309 **文献标志码:**A

Realization and optimization of a LWE sampling algorithm

WANG Kexiang,LI Lin,PENG Shuanghe

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044,China)

Abstract: The crypto system constructed with Learning With Errors (LWE) can resist quantum attacks, and its application efficiency is closely related to the sampling process of LWE problem. In the LWE problem sampling, the error factor sampling which accounted for most of the sampling process. This paper studies the sampling algorithm of the error factor in the LWE problem, and applies the Gaussian distribution (Ziggurat) sampling algorithm to an effective sampling algorithm of the LWE problem. Based on the idea of high sampling efficiency on the sampling domain in the continuous domain, this paper deals with the LWE problem sampling algorithm on the discrete domain. The sampling process is optimized, and a method of rounding the sampling results in the continuous domain is proposed and applied to the LWE problem sampling algorithm. We have compared the two LWE sampling algorithms before and after optimization. The experimental results show that the improved algorithm increases the sampling speed by 38% ~ 200% in the condition of not using a lot of memory and ensuring the safety of sampling.

Keywords: lattice; learning with errors; Gaussian distribution; error factor; sampling

收稿日期:2016-10-19
基金项目:国家自然科学基金青年基金项目(61402035);中央高校基础科研业务费专项资金(2014JBM033)
Foundation items:National Natural Science Foundation of China(61402035);Fundamental Research Funds for the Central Universities (2014JBM033)
第一作者:王柯翔(1992—),男,广西桂林人,硕士.研究方向为密码学及其应用.email:14120358@bjtu.edu.cn.
通信作者:黎琳(1978—),女,山东济南人,讲师,博士. email: lilin@bjtu.edu.cn.
引用格式:王柯翔,黎琳,彭双和.一种基于 LWE 采样算法的实现与优化[J].北京交通大学学报,2017,41(5):32-36.
WANG Kexiang,LI Lin,PENG Shuanghe. Realization and optimization of a LWE sampling algorithm[J].Journal of Beijing Jiaotong University, 2017, 41(5): 32-36.(in Chinese)

在量子计算不断发展与突破的情形下,传统的密码学将面临巨大的挑战.目前研究人员已提出针对大整数因子分解和计算离散对数的量子算法,这将使得未来的量子计算机可攻击现有的基于数论问题的公钥密码算法,如 RSA (Rivest-Shamir-Adleman) 提出的加密算法,DSA (Digital Signature Algorithm, 数字签名算法), ECDSA (Elliptic Curve Discrete Signature Algorithm, 椭圆曲线数字签名算法) 等,而量子计算机的出现只是时间问题,因此设计新型、高效和能抵抗量子攻击的密码算法已经成为了如今密码学发展的趋势^[1].

在抵抗量子计算机攻击的密码体制研究中格理论中的困难问题被应用在很多密码体制设计中^[2].目前应用在密码体制里主要有小整数解 (Small Integer Solution, SIS) 问题^[3] 和学习错误 (LWE) 问题^[4], SIS 被视为 LWE 问题的对偶问题. 2005 年,文献^[4]提出的 LWE 问题因为已经被证明为 NP 完全问题,从而在 NP 不等于 P 的假设下,基于 LWE 问题的公钥密码方案可以抵御量子计算机的攻击,并且有可证明安全的特性,被应用在很多公钥密码体制中.而很多在格上的基于 LWE 问题的密码体制,通常需要从某种分布上采样获得错误因子,一般是在格上的离散高斯分布上采样^[5].因而研究出有效的格上的离散高斯分布的采样算法对 LWE 问题的应用有着重要的意义.

基于高斯随机数的生成器^[6],目前在离散高斯分布上的采样算法主要有:文献^[7]的离散 Ziggurat 采样算法较优.文献^[8]的 Kunth-Yao 算法是基于一种完美的随机比特产生器和随机比特模型来进行采样,DDG-tree 的概念在其中起着关键作用,Kunth-Yao 算法的采样速度较快,但是需要占用较大内存,综合占用内存和运行速度离散 Ziggurat 算法性能较优.文献^[9-10]的拒绝采样算法 (rejection sampling) 和取反算法^[11-12] (inversion method) 是采样的两种最基本的算法.文献^[13-14]的二分法算法是将这两种算法相结合.

在文献^[5]中提出了一种对错误因子采样的算法,本文作者将在连续域上与离散域上的 Ziggurat 采样算法分别应用在了文献^[5]提出的算法中,通过对连续域上的 Ziggurat 算法进行改进,在对比实验下发现本文在连续域上高斯分布的 Ziggurat 改进算法相比较于离散 Ziggurat 采样算法,可以使得 LWE 问题的采样算法的采样速度提高 38% ~ 200%.

1 预备知识

1.1 符号说明

本文中, \mathbf{Z} 表示整数域, \mathbf{Z}_q 表示模 q 的整数域; \mathbf{R} 表示实数域;如果 X 是一个集合,用 $\mathbf{x} \leftarrow_{\mathbf{R}} X$ 表示从 X 上均匀采样获得的向量 \mathbf{x} ,如果 X 是一个分布, $\mathbf{x} \leftarrow_{\mathbf{R}} X$ 表示从 X 上采样获得满足 X 分布的 \mathbf{x} .

1.2 基本概念

定义 1 LWE 问题. LWE 存在两个方面的问题: 1) 判断问题. 是对于当给定相互独立的采样 $(\mathbf{A}_i, \mathbf{b}_i^T) \in \mathbf{Z}_q^{n \times m} \times \mathbf{Z}_q^m$, 其中 \mathbf{A}_i 表示 i 维矩阵, \mathbf{b}_i^T 表示 i 维向量, n 和 m 均为正整数, q 为素数. 对于 $\mathbf{A}_i \leftarrow_{\mathbf{R}} \mathbf{Z}_q^{n \times m}$, $\mathbf{e}_i \leftarrow_{\mathbf{R}} \chi$, 有 $\mathbf{b}_i^T = \mathbf{s}^T \mathbf{A}_i + \mathbf{e}_i^T \bmod q$, $\mathbf{s} \in \mathbf{Z}_q^n$, 其中 \mathbf{e}_i 为错误因子, \mathbf{s}^T 为秘密因子. 在 \mathbf{Z}_q^n 上的某一分布 χ , χ 一般为离散高斯分布, 很难判断出采样 $(\mathbf{A}_i, \mathbf{b}_i^T)$ 是从 $\mathbf{Z}_q^{n \times m} \times \mathbf{Z}_q^m$ 中均匀随机采样获得的. 2) 搜索问题. 则是很难从采样 $(\mathbf{A}_i, \mathbf{b}_i^T)$ 中找出 \mathbf{s}^T .

定义 2 格. 格是 \mathbf{R}^n 的一个子群, 对于 n 个线性无关的向量组成的基 $\mathbf{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\} \subset \mathbf{R}^n$. 定义由基 \mathbf{B} 构成的 n 维格为

$$\Lambda = L(\mathbf{B}) = \{\mathbf{B} \times \mathbf{c} = \sum_{i=0}^n \mathbf{b}_i \cdot c_i : c \in \mathbf{Z}^n\} \quad (1)$$

有满秩矩阵 $\mathbf{A} \in \mathbf{Z}_q^{n \times m}$, 其中 $n \in \mathbf{Z}$, 则定义所有向量正交的奇偶校验矩阵 \mathbf{A} 组成的 m 维格为

$$\Lambda^\perp(\mathbf{A}) = \{\mathbf{e} \in \mathbf{Z}^m \mid \mathbf{A}\mathbf{e} = 0 \bmod q\} \quad (2)$$

其中, 对于向量 $\mathbf{v} \in \mathbf{Z}_q^n$, 任意 $\mathbf{x} \in \mathbf{Z}^m \Lambda_v^\perp(\mathbf{A})$ 表示 $\Lambda^\perp(\mathbf{A})$ 到 $\mathbf{x} + \Lambda^\perp(\mathbf{A})$ 的转换, 满足 $\mathbf{A}\mathbf{x} = \mathbf{v} \bmod q$, 定义为

$$\Lambda_v^\perp(\mathbf{A}) = \{\mathbf{e} \in \mathbf{Z}^m \mid \mathbf{A}\mathbf{e} = \mathbf{v} \bmod q\} \quad (3)$$

2 Ziggurat 采样算法

2.1 连续域上的 Ziggurat 采样算法

Ziggurat 采样算法是文献^[10]提出的拒绝采样算法的一种改进, 将高斯概率密度分布函数中 $x \geq 0$ 部分划分为 m 个矩形, 为了保证选取到每一个矩形的概率相同, 将所有矩形面积设为相同. 对于任一矩形 R_i 都与分布曲线相交分为两部分, 其中一部分为完全在高斯分布以内, 另一部分则在分布以外, 如图 1 所示.

图 1 中, 每次分好块后将每个矩形的右下角的点 (x_i, y_i) 记下后, 开始采样: 1) 随机产生一个整数 i , $1 < i < m - 1$, 来选中第 i 个矩形. 2) 在 $[0, x_i]$ 上随机产生一个 x' , 如果 $x' \leq x_{i+1}$, 说明 x' 必在分布以内; 否则 $x_i < x' \leq x_{i+1}$, 需要采用拒

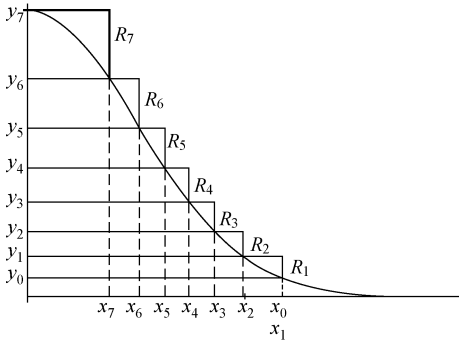


图 1 划分块数 $m=8$ 时的 Ziggurat 采样方法

Fig.1 Ziggurat sampling method
when divided blocks $m = 8$

绝采样. 3) 随机产生一个 $y' \in [y_{i-1}, y_i]$, 如果 $y' + y_{i-1} \leq f(x')$ ($f(x')$ 为连续高斯分布概率密度函数), 则表示选中了在分布以内的值, 就选中 x' , 否则拒绝这个 x' , 返回到第 1) 步产生 i 值.

2.2 离散域上的 Ziggurat 采样算法

离散 Ziggurat 采样算法^[7]在 Ziggurat 算法的基础上进行了改进, 在算法思路与连续域 Ziggurat 算法相同, 主要是将高斯分布上的坐标值和矩形的面积进行了变换, 做出了如下改进:

1) 把每一块矩形的面积规定为对应的 x 轴坐标 x_i 值向上取整乘以矩形的高度, $(1 + \lfloor x_i \rfloor) \times (y_i - y_{i-1})$.

2) 矩形的各点的计算方法不一样. m 个矩形块有相同的面积 S , 规定: $y_m := 0, x_0 := 0, x_m := t\sigma$, 其中, t 表示高斯分布的边界, σ 表示高斯分布的方差. 通过迭代运算可以得到

$$y_{m-1} = S / (1 + \lfloor x_m \rfloor) \quad (4)$$

$$x_{m-1} = \rho_{\sigma}^{-1}(y_{m-1}) \quad (5)$$

对于 $i = m - 2, \dots, 1$,

$$y_i = S / (1 + \lfloor x_{i+1} \rfloor) + y_{i+1} \quad (6)$$

$$x_i = \rho_{\sigma}^{-1}(y_i) \quad (7)$$

$$y_0 = S / (1 + \lfloor x_1 \rfloor) + y_1 \quad (8)$$

式中, $\rho_{\sigma}^{-1}(y)$ 表示离散高斯分布概率密度函数.

3) 在 $[\bar{y}_i, \bar{y}_{i-1}]$ 中获取随机数 \bar{y} , 为此定义 $\bar{h}_i := \bar{y}_{i-1} - \bar{y}_i$ 为第 i 个矩形的高, 在 $\{0, 1, \dots, 2^{\omega} - 1\}$ 中均匀采样获得 y' , 将 \bar{y} 的值定义为 $\bar{y} = \bar{h}_i y' \in [0, 2^{\omega} \bar{h}_i]$, 其中 ω 为正整数. 通过上述变换使得均匀采样得到的值更为准确.

3 LWE 采样算法的改进

3.1 一种 LWE 采样算法的实现

在文献[5]中提出了一种从初始矩阵 G 上对

LWE 问题中错误因子的采样算法, 为了方便在之后做比较, 将该算法描述为算法 1, 其内容如下:

算法 1:

输入: G, v

输出: $t = (t_0, \dots, t_{k-1})^T$

对于 $i = 0, \dots, k - 1$, 循环计算

$$t_i \leftarrow D_{2Z+a_i, r} \quad (9)$$

$$a_{i+1} = (a_i - t_i) / 2 \quad (10)$$

式中: G 满足 $G_{nk} = I_n \otimes g^T$, 其中 I_n 表示单位向量, $g^T = (1, 2, \dots, 2^{k-1})$; $D_{2Z+a_i, r}$ 为整数域上的离散高斯分布. 该算法初始化一整数 a_0 为 v 中任意元素 v_i 满足 $v_i \in \mathbf{Z}_q$, 输出的 t 满足格上分布 $t \in \Lambda_{v_i}^{\perp}(g^T)$, 向量 t 即为采样结果.

算法 1 的优点在于将较为复杂的格上的高斯分布采样转化为整数域上以 $2Z + a_0$ 为中心, r 为方差的离散高斯分布的采样, 使得采样效率能较大提高. 因而决定该算法采样效率好坏的关键是在整数域上的离散高斯分布采样的过程, 在前文中提到了离散 Ziggurat 采样算法是目前综合性能较优的算法, 因而将该算法应用到了对算法 1 的实现, 与之后提出的算法进行比对.

3.2 整数域上采样过程的改进

算法 1 通过将格上的离散高斯分布采样转化到整数域上的离散高斯分布采样的方法使采样效率得到提升, 而考虑到对于一种分布的采样, 连续分布上的采样效率要比离散分布上采样效率高, 基于这一思路, 将连续高斯分布的采样通过取整的处理方式转化为离散高斯分布的采样, 形成了算法 2, 其内容如下:

算法 2:

输入: G, v

输出: $t = (t_0, \dots, t_{k-1})^T$

对于 $i = 0, \dots, k - 1$, 循环计算

$$t_i' \leftarrow \lfloor \text{Ziggurat}(a_i, r) \rfloor \quad (11)$$

$$\text{If}(t_i' \bmod 2) = (a_i \bmod 2) \quad (12)$$

$$t_i = t_i' \quad (13)$$

$$a_{i+1} = (a_i - t_i) / 2 \quad (14)$$

如果式(12)不满足, 则返回再执行式(11).

为了验证算法 2 的正确性, 对该算法进行了证明:

1) 由式(14)可知 $t_i = a_i - 2a_{i+1}$;

2) 若向量 t 满足 $t \in \Lambda_{v_i}^{\perp}(g^T)$, 则满足 $g^T t = v_i \bmod q$, 由 $g^T = (1, 2, \dots, 2^{k-1})$ 可得

$$g^T t = t_0 + 2t_1 + 2^2 t_2 + \dots + 2^{k-1} t_{k-1} \quad (15)$$

3) 将 $t_i = a_i - 2a_{i+1}$ 代入式(15), 可得

$$g^T t = a_0 - 2a_1 + 2(a_1 - 2a_2) +$$

$$2^2(a_2 - 2a_3) + \dots + 2^{k-1}(a_{k-1} - 2a_k) \quad (16)$$

$$\mathbf{g}^T \mathbf{t} = a_0 - 2^k a_k = v_i - 2^k a_k \quad (17)$$

式中, $2^k a_k$ 中 $k = \lceil \ln q \rceil$, 则 $2^k \bmod q = 0$, 从而 $\mathbf{g}^T \mathbf{t} = v_i \bmod q$ 成立。

通过上述证明可知算法 2 在采样结果上是正确的, 采样得到的向量 \mathbf{t} 均能满足 $\mathbf{t} \in \Delta_{v_i}^\perp(\mathbf{g}^T)$, 所以采样结果均满足在格上的分布, 因而算法 2 的采样结果既具有正确性, 也具有安全性。

4 实验结果及分析

实验是在 Intel(R) Core(TM) i3-4170 3.70 GHz 的处理器, 8 GB 内存, 操作系统为 ubuntu 14.04 的计算机上实现的. 对于算法中的中心和方差等参数, 选择依照了文献[1]中提出的参数: $n \geq 256$; q 为 2 的次方, 且 $q \geq 2^{19}$; $r \geq 2 \times$

$\sqrt{\ln\left(2n\left(1+\frac{1}{\epsilon}\right)\right)}/\pi$. 两种算法都调用了基于 linux 的时间函数 clock_gettime 的计数器 CLOCK_PROCESS_CPUTIME_ID 来计算运行效率。

通过验证得到向量 \mathbf{t} 是否满足 $\mathbf{g}^T \mathbf{t} = v_i \bmod q$ 来验证实验结果是否正确. 得到的采样都满足上述条件. 算法 1 中使用的离散 Ziggurat 采样算法, 在进行采样算法前需要预先计算将高斯分布划分矩形的块数, 实验使用了文献[5]中对数据精度和方差等参数的选取方法; 同时由于 Ziggurat 采样算法本身具有划分矩形的块数越多, 采样速度越快且占用内存也越高的特性, 从而将算法 1 中离散 Ziggurat 采样算法在不同划分块数下与算法 2 进行了效率对比结果如图 2 所示。

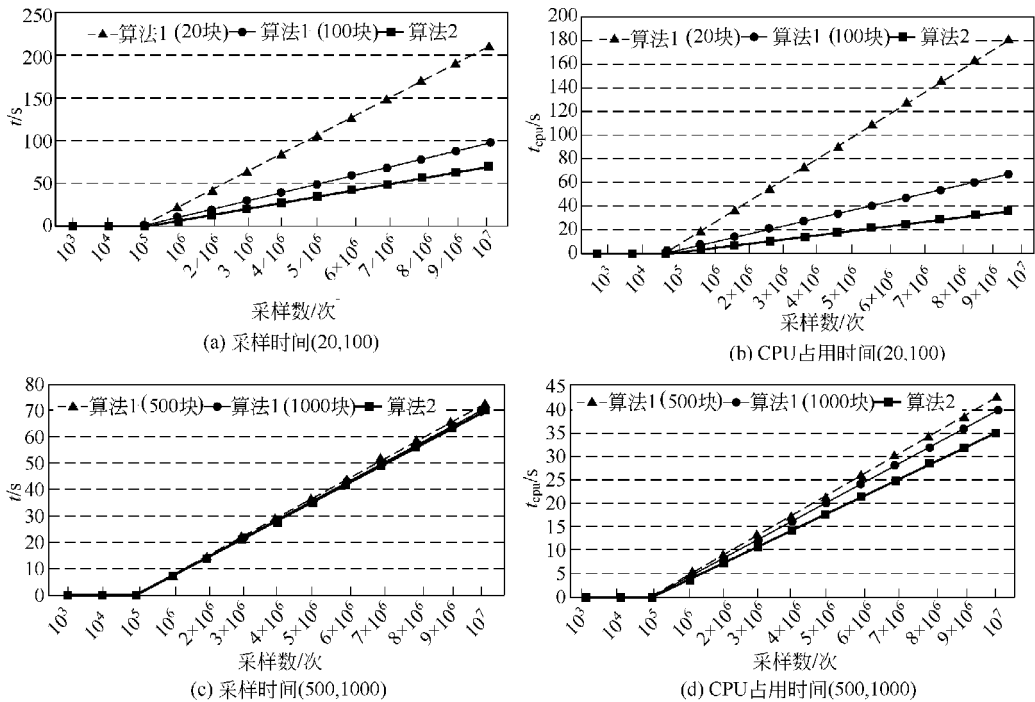


图 2 算法 1 和算法 2 实验结果对比

Fig.2 Experimental results comparisons contrast between algorithm 1 and algorithm 2

图 2(a)中显示了算法 1 中使用离散 Ziggurat 采样算法在分块数为 20 和 100 时与算法 2 采样所需时间的对比, 从图 2 中可以看出, 在采样次数到达 10^7 次时, 算法 2 采样所需的时间分别是算法 1 采样所需时间的 33% 和 72%。从图 2(b) 可以看到, 此时算法 1 中使用离散 Ziggurat 采样算法在分块数为 20 和 100 时比算法 2 中占用 CPU 的时间更长, 约是算法 2 占用时间的 514% 和 190%。图 2(c) 显示了算法 1 中使用离散 Ziggurat 采样算法在分块数达到 500 和 1 000 时与算法 2 的采样所需时间对比, 可以看到此时算法 1 与算法 2 采样所需时间接近。

从图 4(d)中可以看到, 此时算法 2 比算法 1 中使用离散 Ziggurat 采样算法分块数为 500 和 1 000 时所占用 CPU 的时间更短。

从算法采样速度快慢来看, 尽管将连续高斯分布采样结果直接取整的方法, 不能够准确转化成为对应整数域上的离散高斯采样结果, 但是通过实验结果表明, 将此方法应用在本文提到的 LWE 问题的采样算法中是可行的. 应用到本文提到的 LWE 问题的采样算法中, 在采样次数相同且数量较大的情况下, 将连续高斯分布采样结果取整的算法应用

在LWE问题的采样中比使用离散Ziggurat采样算法在分块数不是特别高的情况下所需的采样时间少,使得算法的采样速度提升了38%~200%左右,而在使用分块数很大的离散Ziggurat采样算法时,与连续高斯分布采样结果取整的方法的采样速度十分接近。

从算法采样占用内存情况来看,实验结果表明了在分块数为正常数值20和100的情况下,使用将连续高斯分布取整的方法比使用离散Ziggurat采样算法在LWE问题采样中占用CPU的时间短,只有到分块到500和1000这样非常大的数值的时候,两种算法的占用CPU的时间才较为接近,而此时离散Ziggurat采样算法在预计算上花费的时间很长,会很大的影响LWE问题采样算法的整体效率,因此在离散Ziggurat选择分块数为500和1000时,算法2的采样效率仍然高于算法1。

5 结论

本文在高斯分布上采样效率较高的Ziggurat采样算法基础上,分析了一种将LWE问题采样从格上离散高斯分布转化到整数域上的离散高斯分布上采样的算法,并将离散Ziggurat采样算法应用其中.本文作者提出了一种新的改进算法,使用将连续高斯分布的采样结果取整的采样方法替代了离散高斯分布采样,为了验证改进的采样算法的采样效率,对两种采样算法进行了实验,实验记录的采样所需时间显示,在产生采样数量大于等于 10^7 且占用内存不大的情况下,改进的LWE问题的采样算法所需时间是原有的采样算法的33%~72%,即采样速度上提升了38%~200%左右,实现了对原有算法的优化。

参考文献(References):

- [1] FOLLÁTH J. Gaussian sampling in lattice based cryptography[J]. Tatra Mountains Mathematical Publications, 2015, 60(1): 1–23.
- [2] CHARLES F, KARNEY F. Sampling exactly from the normal distribution[J]. ACM Transactions on Mathematical Software, 2013, 42(3): 1–10.
- [3] WANG S B, ZHU Y, DI M A, et al. Lattice-based key exchange on small integer solution problem[J]. Science China Information Sciences, 2014, 57(11): 1–12.
- [4] REGEV O. On lattices, learning with errors, random linear codes and cryptography[C]// 37th Annual ACM Symposium on Theory of Computing, 2005: 84–93.
- [5] BANSARKHANI R E, DAGDELEN Ö, BUCHMANN J. Augmented learning with errors: the untapped potential of the error term[J]. Financial Cryptography and Data, 2015: 333–352.
- [6] THOMAS D, LUK W, LEONG P H W, et al. Gaussian random number generators [J]. ACM Computing Surveys, 2007, 39(4): 415–416.
- [7] BUCHMANN J, CABARCAS D, GÖPFERT F, et al. Discrete ziggurat: a time-memory trade-off for sampling from a Gaussian distribution over the integers[C]// Selected Areas in Cryptography, 2014: 402–417.
- [8] ROY S S, VERCAUTEREN F, VERBAUWHEDE I. High precision discrete Gaussian sampling on FPGAs [C]// Selected Areas in Cryptography, 2014: 383–401.
- [9] GENTRY C, PEIKERT C, VAIKUNTANATHAN V. Trapdoors for hard lattices and new cryptographic constructions[C]// Proceedings of the Annual ACM Symposium on Theory of Computing, 2008, 14: 197–206.
- [10] DWARAKANATH N C, GALBRAITH S D. Sampling from discrete Gaussians for lattice-based cryptography on a constrained device[J]. Applicable Algebra in Engineering Communication & Computing, 2014, 25(3): 159–180.
- [11] PEIKERT C. An efficient and parallel Gaussian sampler for lattices[C]// Advances in Cryptology-Crypto, 2010: 145–166.
- [12] MICCIANCIO D, PEIKERT C. Trapdoors for lattices: simpler, tighter, faster, smaller [C]// International Conference on the Theory & Applications of Cryptographic Techniques, 2012, 7237: 700–718.
- [13] DUCAS L, DURMUS A, LEPOINT T, et al. Lattice signatures and bimodal Gaussians[C]// Advances in Cryptology - CRYPTO 2013, 2013, 8042: 40–56.
- [14] MARSAGLIA G, TSANG W W. The ziggurat method for generating random variables [J]. Journal of Statistical Software, 2000, 5(8): 1–7.
- [15] DUCAS L, NGUYEN P Q. Faster Gaussian lattice sampling using lazy floating-point arithmetic[C]// International Conference on the Theory & Application of Cryptology & Information Security, 2012, 7658: 415–432.