改进的原象采样算法

1. 补充知识

1.1 格、理想格

格: 设$b_1, b_2, \cdots, b_m$是$\mathbb{R}^n$上的一组线性无关向量, 由所有$b_1, b_2, \cdots, b_m$的整数线性组合构成的集合构成一个格, 向量组$b_1, b_2, \cdots, b_m$构成格的一组格基. m, n分别为格的秩和维数.

> 一个格可以有不同的基, 例如, 基$(1,0)^T$和$(0.1)^T$生成格$\mathbb{Z}^2$, 这个格包含所有整数点, 同时基$(1,1)^T$与$(2,1)^T$也可以生成格$\mathbb{Z}^2$.
>
> 物理意义: 高维空间中排列规律的点阵

循环格: 向量$a = (a_0, a_1, \cdots, a_{n-1})^T$的一次循环移位记为$rot(a) = (a_{n-1}, a_0, \cdots, a_{n-2})^T$, 若对于格$L$和$\forall a \in L$, 都有$rot(a) \in L$, 则称格$L$是循环格. 记循环矩阵为

$$Rot(a) = \begin{bmatrix} a_0 & a_{n-1} & \cdots & a_1 \\ a_1 & a_0 & \cdots & a_2 \\ \cdots & \cdots & \cdots & \cdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 \end{bmatrix}.$$

> 提出循环格的定义解决了基于一般格上的密码方案中密钥量大, 运行效率低的问题. 循环格能够用一个向量来表示.

理想格: 将$\mathbb{Z}^n$构造成一个环, 定义乘法运算. 定义乘法的通常方法是取一个n次整系数首一多项式$f(x)$作为模多项式, 即考虑$\mathbb{Z}^n$到环$Z[x]/f(x)$的同构
$\Phi : (\nu_0, \nu_1, \cdots, \nu_{n-1}) \to \nu_0 + \nu_1 x + \cdots + \nu_{n-1} x^{n-1}$, 对于环$Z[x]/f(x)$上由$g_1, g_2, \cdots, g_m$生成的理想$I$, 任意的$h \in I$都可以表示为$g_1, g_2, \cdots, g_m$的模$f(x)$倍式组合:
$h(x) = u_1(x)g_1(x) + u_2(x)g_2(x) + \cdots + u_m(x)g_m(x)(\bmod f(x))$.

环$\mathbb{Z}^n$的理想I又是格$\mathbb{Z}^n$的子格. 称这个子格为格$\mathbb{Z}^n$的理想格.

> 理想格是循环格的概念推广, 一般格式群的子群, 理想格式环的理想
>
> 理想格式为了降低格表示的空间尺寸, 可以用几个生成元来表示原来需要n^2个元素表示的$n \times n$的矩阵代表的基.
>
> 理想格对内乘法封闭, 对外乘法吸收

1.2 高斯分布

高斯函数: 以c为中心, 方差参数是s的n维高斯函数为$\rho_{s,c}(x) = \exp\left(\frac{-\pi\|x-c\|^2}{s^2}\right)$, 其中$\forall x \in \mathbb{R}^n, s \in R$

以向量$c > 0$为中心,实参数为$s > 0$可以定义n维格Λ得离散高斯分布: $D_{\Lambda,s,c}(x) = \frac{\rho_{s,c}(x)}{\rho_{s,c}(\Lambda)}$

2. 原象采样算法

2.1 新的gadget采样器

使用一个复合模数$Q = pq$, 其中p 和q是正整数. 选择一个特点向量$\mathbf{g} = (1, b, \ldots, b^{w-1}) \in \mathbb{Z}^w$, 其中$b$是一个小整数, $w = \lceil log_b(q) \rceil$.

(Algorithm 1： $ApproxGadget(\mathbf{t}, r, p, q)$)

1. 给定目标$t \in \mathbb{Z}_Q$, 正实数$r > 0$和整数$p, q > 0$使得$Q = pq$.
2. 通过双射$\tau : \mathbb{Z}_Q \to \mathbb{Z}_p \times \mathbb{Z}_q$将$t$分解为$(t_p, t_q)$.
3. 从$D_{\Lambda^{\perp}_{t_q}(\mathbf{g}^t), r}$中采样$\mathbf{z}$.
4. 返回$\mathbf{z}$.

2.2 改进的原象采样器

定义全局参数$\Gamma = (n, m, p, q, Q, \chi)$, 模数$p$和$q$, 全局模数$Q = pq$, 以及秘密分布$\chi$. $\mathbf{A} \in \mathbb{Z}_Q^{n \times m}$ 是$m > n$的矩阵.

定义为矩阵$\mathbf{T} \in \mathbb{Z}^{m \times n}$使得, $\mathbf{A} \cdot \mathbf{T} = \mathbf{F} = \mathbf{I}_n \otimes \mathbf{f} \bmod Q$, 其中$\mathbf{f} = p \cdot \mathbf{g} = p \cdot (1, b, \ldots, b^{w-1})$ 是Gadget向量.

(Algorithm 2： $ApproxPreSamp(\mathbf{A}, \mathbf{T}, \mathbf{u}, r, \Sigma)$)

1. 从$D_{Z^m, \sqrt{\Sigma_p}}$中高斯采样扰动向量$\mathbf{p}$.
2. 计算目标$\mathbf{v} = \mathbf{u} - \mathbf{A} \cdot \mathbf{p} \mod Q$.
3. 使用gadget采样器$ApproxGadget(\mathbf{v}, r, p, q)$采样$z$使得$\langle \mathbf{f}, \mathbf{z} \rangle = v - e \mod Q$.
4. 返回近似原象$\mathbf{x} = \mathbf{p} + \mathbf{T} \cdot \mathbf{z}$.

2.3 高斯采样

针对$ApproxGadget$的 `第3步` 以及$ApproxPreSamp$中的 `第1步` 过程中, 分别有两次高斯采样, 文中定义为$SampleP(msk)$和$SampleZ(c)$.

SampleP: 从$D_{Z^m, \sqrt{\Sigma_p}}$中高斯采样扰动向量$\mathbf{p}$.

输入: 主私钥msk, **输出**: 扰动向量$\mathbf{p}$.

1. 在R^w空间中采样一个向量$\mathbf{p}'$, 其协方差矩阵为$\sigma_2^2 \cdot I_{w \times w} - r^2$, 即$\mathbf{p}' \leftarrow D_{\mathbb{Z}^{w \cdot n}, \sqrt{\sigma_2^2 - r^2}}$.
2. 使用$\mathbf{p}'$来计算中心向量$c = (c_0, c_1)$.
3. 使用c和Σ_2来递归地采样p_0和p_1. 这个过程涉及到更新协方差矩阵和中心点，直到采样出所有需要的p的分量. 这个过程实际上是使用了Peikert's sampler, 文中记为 `SampleFz` 函数.
4. 最后, 将所有采样得到的分量$p_0, p_1, \ldots, p_{w+1}$合并成最终的扰动向量$\mathbf{p}$并返回.

SampleZ: 从$D_{\Lambda^{\perp}_{t_q}(\mathbf{g}^t),r}$中采样$\mathbf{z}$.

输入: 中心c, 输出: 样本$z+\lfloor c \rfloor$.

1. 计算$d=c-\lfloor c \rfloor$并初始化$z+$为一个基于固定高斯分布(文中记为 BaseSample 函数)的样本.
2. 随机选择一个比特b.
3. 根据b和$z+$计算z.
4. 使用z和$z+$计算一个概率值x.
5. 如果一个随机数r大于$exp(x)$, 如果接受, 则返回$z+\lfloor c \rfloor$作为最终样本.

SampleFz: 生成与中心c和协方差d相关的随机样本

输入: 协方差d,中心c, 输出: 随机样本z.

1. 在$\mathbb{R}^n$空间中生成一个具有协方差$d-\bar{r}^2$的连续高斯向量y, 即标准化.
2. 计算$c'=c+\sqrt{d-\bar{r}^2}\cdot y$, 将高斯样本平移至中心$c$.
3. 调用 SampleZ 函数对每个系数c'_i进行采样, 得到环上的随机样本z.
4. 返回最终的环上随机样本z.

BaseSample: 从半整数值的高斯分布$D^{+}_{Z,\bar{r}}$中采样

输入: none, 输出: 高斯分布样本$z+$.

1. 选择一个随机的 72 位比特数u.
2. 通过循环 13 次, 使用预先定义的累积分布表(cumulative distribution table, CDT)来计算并累加样本$z+$.
3. 利用u和CDT来确定非负整数$z+$的值.
4. 返回计算得到的高斯分布样本$z+$.

2.4 参数选择

文中表格2和3给出了合适的算法参数含义与选择.

3. 算法理解

> 原象采样算法的作用就是生成与用户身份相关联的私钥, 对应IBE算法中的Extract.

假设我们有一个身份id, 并通过初始化生成了主密钥mpk和msk, 我们想要利用这个身份标识通过改进的原象采样算法, 生成一个IBE算法中的用户私钥sk_{id}.

1. 选择一个大的复合模数$Q=pq$, 其中p,q是正整数.
2. 生成一个gadget向量$\mathbf{g}=(1,b,\ldots,b^{w-1})\in\mathbb{Z}^w$, 其中$b$是一个小整数, $w=\lceil log_b(q)\rceil$.

3. 计算id的哈希值$u = H(id)$.
4. 随机采样一个扰动向量$\mathbf{p}$, 这个向量是从以$\Sigma_{\mathbf{p}}$为协方差的高斯分布中采样得到的(即$SampleP(msk)$.), 以主私钥msk作为输入参数.
5. 使用$\mathbf{u}$和$\mathbf{p}$计算目标$\mathbf{v} = \mathbf{u} - \mathbf{A} \cdot \mathbf{p} \mod Q$. 其中$\mathbf{A} = [1, a, \mathbf{b}]$, a是由一个理想的可扩展输出函数XOF和一个32字节的种子$seed_a$生成的.
6. 对于$\mathbf{v}$中的每个部分都作为参数, 使用gadget采样器$ApproxGadget(\mathbf{v}, r, p, q)$从$D_{\Lambda^{\perp}_{t_q}(\mathbf{g}^t),r}$中采样到相应的$z$部分(即$SampleZ(c)$), 满足$\langle \mathbf{f}, \mathbf{z} \rangle = v - e \mod Q$. 其中$\mathbf{f}$是gadget的扩展向量, e是一个小的误差项. r是高斯分布的标准差.
7. 将采样得到的z每部分组合起来, 形成最终的$\mathbf{z}$向量.
8. 计算用户私钥$\mathbf{y} = \mathbf{p} + \mathbf{T} \cdot \mathbf{z}$, 其中$\mathbf{T}$是与$\mathbf{A}$相关的陷门矩阵.
9. 最终，输出的用户私钥是$\mathbf{y}$的一部分, $sk_{id} = (\mathbf{y_1}, \cdots, \mathbf{y_{w+1}})$.