TITLE OF DOCUMENT

NAME OF AUTHOR

1. The χ^2 test for uniform distribution

We briefly recall the property and usage of the χ^2 test for uniform distributions over a finite set S. Suppose M samples $y_1, \cdots, y_M \in S$. We partition S into r subsets

$$S = \sqcup_{j=1}^{r} S_i,$$

For each $1 \leq j \leq r$, we compute the expected number of samples that would fall in the i-th subset: $c_j := |S_j|M/|S|$. Then we compute the actual number of samples, $t_j := |\{1 \leq i \leq r : y_i \in S_j\}|$. Finally, the χ^2 value is computed as

$$\chi^2(S, y) = \sum_{j=1}^{r} \frac{(t_j - c_j)^2}{c_j}.$$

Note that degree of freedom in this test is $d = r - 1$. To decide whether the samples are from a uniform distribtuion, we can either look up a table of χ^2 values, or use an approximation rule: when df is large, the χ^2 distribution can be well-approximated by a normal distribution $N(d, 2d)$; for example, if it turns out that $\chi^2 \notin (d - c\sqrt{2d}, d + c\sqrt{2d})$, then the confidence we have that the samples are not taken from a uniform distribution is $2\Phi(c) - 1$.

2. The χ^2 Attack on $SRLWE(\mathcal{R}, \mathfrak{q})$

Let $\mathcal{R}$ be an RLWE instance with error distribution $D_{\mathcal{R}}$ and $\mathfrak{q}$ be a prime ideal above q. Our attack relies on the assumption that the distribution $D_{\mathcal{R}} \pmod{\mathfrak{q}}$ is distinguishable from the uniform distribution on the finite field $F = R/\mathfrak{q}$. More precisely, the attack loop through all q^f possitbilities of $\bar{s} = s \pmod{\mathfrak{q}}$. For each guess s', it computes the values $\bar{e}' = \bar{b} - \bar{a}s' \pmod{\mathfrak{q}}$ for every sample $(a, b) \in S$. If the guess is wrong, or if the samples are taken from the uniform distirbution in $(R_q)^2$ instead of an RLWE instance, the values $\bar{e}'$ would be uniformly distributed in F and it is likely to pass the χ^2 test. On the other hand, if the guess is correct, then we expect the errors $\bar{e}'$ to fail the χ^2 test.

Let $N = q^f$ denote the cardinality of F. Note that N is also the number of χ^2 tests we run in the attack. For the attack to be successful, we need the $(N - 1)$ tests corresponding to wrong guess of $s \pmod{\mathfrak{q}}$ to pass, and the one test corresponding to the correct guess to fail. Therefore, we need to choose the confidence interval of our χ^2 test so that it is unlikely for a set of samples coming from uniform distribution to fail the test. In practice, we choose the confidence level to be $\alpha = 1 - \frac{1}{10N}$. Let β denote the probability that the sample errors fails the uniform test with probability Then the probability that our algorithm will success is $p = (1 - \frac{1}{10N})^{N-1}\beta$. Note that when N is large, $(1 - \frac{1}{10N})^{N-1}$ is about $e^{-1/10} \approx 0.904$.

Algorithm 1 χ^2-test attack of $SRLWE(\mathcal{R}, \mathfrak{q})$

Require: $R = (K, q, \sigma, s)$ – an RLWE instance.
$R = \mathcal{O}_K$ – the ring of integers of K.
n: the degree of K.
$\mathfrak{q}$: a prime ideal in K above q.
N: the cardinality of $R/\mathfrak{q}$.
S: a collection of M ($M = \Omega(N)$) RLWE samples $(a, b) \sim \mathcal{R}$.
Ensure: a guess of the value $s \pmod{\mathfrak{q}}$, or **NON-RLWE**, or **INSUFFIICNET-SAMPLES**

$\alpha \leftarrow 1 - \frac{1}{10N}$.
$\omega \leftarrow \Phi^{-1}((1+\alpha)/2)$
$G = \emptyset$
for s in F **do**
  for a, b in S **do**
    $E \leftarrow \emptyset$.
    $\bar{a}, \bar{b} \leftarrow a \pmod{\mathfrak{q}}, b \pmod{\mathfrak{q}}$.
    $\bar{e} \leftarrow \bar{b} - \bar{a}s$.
    add e to E.
  end for
  Run χ^2 test on E and obtain the value $\chi^2(E)$.
  if $|\chi^2(E) - B - 1| > \omega\sqrt{2B-2}$ **then**
    add s to G
  end if
end for
if $G = \emptyset$ **then**
  return NOT RLWE
else if $G = \{g\}$ **then**
  return g
else
  return INSUFFIICNET-SAMPLES
end if

Proposition 2.1. *The time complexity of the attack is $O(q^{2f})$. Let Δ be the l_2 distance betweenthe distribution $D_{R,\sigma} \pmod{\mathfrak{q}}$ and the uniform distribtuion on $R/\mathfrak{q}$. Then attack succeeds with proability at least*

$$p = 0.904(1 - \Phi(\frac{\omega\sqrt{2(N-1)} - cN^2\Delta}{\sqrt{2(N-1) + 4cN^2\Delta}}))$$

, where Φ is the cumulative distribution function for the standard Gaussian distribution, ω is as in the algorithm, and $c = M/N$.

To get a sense of how the constants behave, consider a hypothetical scenario, where $q = 4091$, $f = 1$ and $D_{R,\sigma} \pmod{\mathfrak{q}}$ takes value from $(-q/4, q/4)$ with probability $1/q + \mu$, and other values with probability $1/q - \nu$. Then the l_2 distance from uniform is equal to $\mu^2(q^2+q)/(q-1)$. Let $M = 5q$ be the number of samples. The following table summarises some μ and p values:

N	statisitical distance($\leq N\Delta$)	p
4093	0.02	0.58