

# TITLE OF DOCUMENT

NAME OF AUTHOR

## 1. THE CHI-SQUARE TEST FOR UNIFORM DISTRIBUTION

We briefly recall the property and usage of the chi-square test for uniform distributions over a finite set  $S$ .

Suppose  $M$  samples  $y_1, \dots, y_M \in S$ . We partition  $S$  into  $r$  subsets

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

For each  $1 \leq j \leq r$ , we compute the expected number of samples that would fall in the  $j$ -th subset:  $c_j := |S_j|M/|S|$ . Then we compute the actual number of samples,  $t_j := |\{1 \leq i \leq M : y_i \in S_j\}|$ . Finally, the  $\chi^2$  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 distribution, we can either look up a table of  $\chi^2$  values, or use an approximation rule: when  $df$  is large, the  $\chi^2$  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  $2\Phi(c) - 1$ .

If  $P, Q$  are two probability distributions on  $S$ , then their *statistical difference* is defined as

$$d(P, Q) = \frac{1}{2} \sum_{t \in S} |P(t) - Q(t)|,$$

and their  $l_2$  distance is

$$d_2(P, Q) = \left( \sum_{t \in S} |P(t) - Q(t)|^2 \right)^{\frac{1}{2}}.$$

We have  $d(P, Q) \leq \frac{\sqrt{|S|}}{2} d_2(P, Q)$ .

## 2. THE $\chi^2$ 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  $q^f$  possibilities 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 distribution 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 chi-square test. On the other hand, if the guess is correct, then we expect the errors  $\bar{e}'$  to fail the test.

Let  $N = q^f$  denote the cardinality of  $F$ . Note that  $N$  is also the number of 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 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  $\beta$  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** chi-square-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 **INSUFFICIENT-SAMPLES**

```

1:  $\alpha \leftarrow 1 - \frac{1}{10N}$ .
2:  $\omega \leftarrow \Phi^{-1}((1 + \alpha)/2)$ 
3:  $G = \emptyset$ 
4: for  $s$  in  $F$  do
5:   for  $a, b$  in  $S$  do
6:      $E \leftarrow \emptyset$ .
7:      $\bar{a}, \bar{b} \leftarrow a \pmod{\mathfrak{q}}, b \pmod{\mathfrak{q}}$ .
8:      $\bar{e} \leftarrow \bar{b} - \bar{a}s$ .
9:     add  $e$  to  $E$ .
10:   end for
11:   Run  $\chi^2$  test on  $E$  and obtain the value  $\chi^2(E)$ .
12:   if  $|\chi^2(E) - B - 1| > \omega\sqrt{2B - 2}$  then
13:     add  $s$  to  $G$ 
14:   end if
15: end for
16: if  $G = \emptyset$  then
17:   return NOT RLWE
18: else if  $G = \{g\}$  then
19:   return  $g$ 
20: else
21:   return INSUFFICIENT-SAMPLES
22: end if

```

Note that the time complexity of the attack is  $O(q^f)$  since we have  $q^f$  possible values for  $s \pmod{\mathfrak{q}}$ . The number of samples need for the attack is also  $O(q^f)$ . The correctness of the attack is captured in the following theorem.

**Theorem 2.1.** Assume  $q \gg 1$ . Let  $\Delta$  denote the statistical distance between the error distribution  $D_{\mathcal{R}, \mathfrak{q}} := D_{\mathcal{R}} \pmod{\mathfrak{q}}$  and the uniform distribution on  $R/\mathfrak{q}$ . Then the above attack succeeds with probability at least

$$p = 0.904(1 - \Phi(\frac{\omega\sqrt{2(N-1)} - \lambda}{\sqrt{2(N-1) + 4\lambda}})),$$

where  $\Phi$  is the cumulative distribution function for the standard Gaussian distribution,  $\omega = \Phi^{-1}(1 - \frac{1}{20N})$ , and  $\lambda = 4M\Delta$ .

*Proof.* The chi-square value for uniformity on samples from  $D_{\mathcal{R}, \mathfrak{q}}$  follow a noncentral chisquare distribution with the same degree of freedom and a parameter  $\lambda$  given by

$$\lambda = MNd_2(D_{\mathcal{R}, \mathfrak{q}}, U(R/\mathfrak{q}))^2.$$

(fixme: cite the chisquare paper.) In particular, we have  $\lambda \geq 4M\Delta$ . We approximate a noncentral chi-square distribution with degree of freedom  $k$  and parameter  $\lambda$  with a Gaussian distribution of mean  $k + \lambda$  and variance  $2k + 4\lambda$ . The result now follows from the fact that the argument of  $\Phi$  is a decreasing function of  $\lambda$ .  $\square$

To get a sense of how the constants behave, we give a table containing some  $p$  values for various choices of  $N$  and  $\Delta$ , computed using theorem. We fix the number of samples to be  $M = 5N$ . Note that 0.904 is the upper bound of the success rate.

TABLE 2.1. Success rates of chi-square attack

$N(=q^f)$	$d(D_{\mathcal{R},\mathfrak{q}}, U(R/\mathfrak{q}))$	$p$
257	0.025	0.84
4093	0.005	0.551
$67^2$	0.005	0.610
12289	0.005	0.903
$307^2$	0.001	0.27