A Note on Cryptanalysis of the Preliminary Version of the NTRU Signature Scheme

Ilya Mironov*
mironov@cs.stanford.edu

Abstract

In this paper a preliminary version of the NTRU signature scheme is cryptanalyzed. The attack exploits a correlation between some bits of a signature and coefficients of a secret random polynomial. The attack does not apply to the next version of the signature scheme.

1 Introduction

Since the introduction of public key cryptography in the late seventies, many cryptosystems have been put forth but only few of them have survived. Finding a new cryptosystem that overcomes deficiencies of existing ones is a challenging task of paramount importance.

In 1996 a new encryption scheme called NTRU was proposed [HPS96, HPS98]. It is a highly efficient cryptosystem based on a problem of finding small vectors in certain lattices. After an increase in key size to preclude an improved lattice attack, its encryption/decryption speed still has a big safety margin over conventional public key cryptosystems. [HS00] introduced a complimentary signature scheme. The NTRU signature scheme is based on a similar problem and also much faster than other known signature schemes.

The hardness of underlying problems is necessary but not sufficient condition of security of the NTRU and NSS schemes. This relation between a hard problem and a cryptosystem is not unique and may be found in other cryptosystems (e.g, the DSS). It opens door to attacks on a cryptosystem that do not attempt to solve the underlying hard problem. In this paper we exhibit one such an attack on the preliminary version of the NTRU signature scheme.

The current version of the NTRU signature scheme [HPS00] is resistant to our attack.

2 A brief description of the preliminary version of the NSS

We briefly describe the preliminary version of the NTRU signature scheme (NSS) as it was presented in [HS00].

The parameters of the schemes are

$$(N, p, q, q_f, q_w, K, B).$$

[HS00] does not give clear guidelines on the choice of these parameters except for these conditions:

$$p \text{ divides } q_f \text{ and } q_w$$
$$q \text{ divides } q_f q_f$$

*Supported by NSF contract #CCR-9984259

$\left(\frac{2\pi e B^2}{q^2 N}\right)^{N/2}$ has the order of the security parameter of the scheme, e.g, 2^{-80}.

We will use the following parameters that according to [HS00] "provide a security level that is at least as great as an RSA 1024-bit modulus:"

$$(N, p, q, q_f, q_w, K, B) = (251, 2, 128, 16, 8, 6, 393). \tag{1}$$

All polynomials in the description of the NSS have degree $N-1$ and the multiplication is done using the rule $X^N = 1$. The multiplication in the ring $\mathbb{Z}[X]/(X^N - 1)$ is denoted by $*$. The norm of a polynomial $a = a_{N-1}X^{N-1} + \cdots + a_1 X + a_0$ is

$$||a|| = \sqrt{(a_0 - \mu_a)^2 + \cdots + (a_{N-1} - \mu_a)^2},$$

where $\mu_a = \sum_{i=0}^{N-1} a_i/N$ is the mean of its coefficients. The norm of a polynomial that is only determined modulo q is computed by putting first its coefficients in a preselected interval $[A, A+q-1]$ by adding or subtracting q.

The scheme is designed as follows

- **Private and public keys:** Bob chooses four polynomials of degree $N-1$. Two of them, f_1 and g_1, are random with coefficients reduced modulo p (if $p = 2$, these polynomials are binary). Two others, f_2 and g_2, have exactly K coefficients equal to 1 with the rest equal to 0. Then Bob computes f and g as

  $$f = f_1 + q_f f_2 \quad \text{and} \quad g = g_1 + q_f g_2$$

  and inverses of f modulo p and q

  $$F_p \equiv f^{-1} \pmod{p} \quad \text{and} \quad F_q \equiv f^{-1} \pmod{q}.$$

  The **private** key is (f, F_p). The **public** key is the product

  $$h \equiv F_q * g \pmod{q}.$$

- To **sign** a message M the signer does the following:
  1. Hash M and encode the result as a polynomial m modulo p.
  2. Compute $w_1 = F_p * m \pmod{q}$.
  3. Randomly add or subtract multiples of p to some coefficients of w_1. For example, if $p = 2$, replace two zero coefficients with 2, two with -2, and two of the 1 coefficients with -1.
  4. Randomly choose w_2 with exactly K coefficients equal to 1 and the rest to 0.
  5. Compute $w = w_1 + q_w w_2$.
  6. Compute $s = f * w \pmod{q}$.
  7. The signature on M is the the pair (m, s).

- To **verify** a signature (m, s) on M:
  1. Check that $s \equiv m \pmod{p}$.
  2. Check that $||s|| \leq B$.
  3. Compute $t = h * s \pmod{p}$ and check that $||t|| \leq B$.

The norm of s and t that are determined modulo q can be computed given A. A is chosen so that the means of the coefficients of s and t lie in the middle of the interval $[A, A+q-1]$. We let

$$A = \frac{N}{4} + \frac{q_f K}{2} + \frac{q_w K}{2} - \frac{q}{2}.$$

The following observation is crucial for the proof of completeness of the NSS (and our attack):

$$s \equiv f * w = (f_1 + q_f f_2) * (w_1 + q_w w_2) \equiv f_1 * w_1 + q_f f_2 * w_1 + q_w f_1 * w_2 \pmod{q},$$

because q divides $q_f q_w$. Now for our choice of parameters the resulting polynomial will have almost all of its coefficients lying in an interval of length q. When one computes its norm and puts its coefficients in the interval $[A, A+q-1]$, one almost restores this polynomial in integers (to within some additive constant).

3 Attack on the NSS

We show that a passive attacker who may only intercept signatures can recover the private key. For the concrete parameters (1) just a few dozens signatures are sufficient to reveal part of the key. The attack does not even need the messages corresponding to the intercepted signatures.

We present the attack in three steps. First, a correlation between two coefficients of s and w is proved. Second, we show how this correlation can be exploited resulting in a partical recovery of f_2 and g_2. The two polynomials are recovered to within a circular shift of their coefficients. Third, the attack is extended to recovery of more bits of the secret key.

3.1 Useful correlation

We write $p[k]$ to denote the kth coefficient of a polynomial p. For $-N < k < 0$ let $p[k] \stackrel{\text{def}}{=} p[k+N]$.

There are K coefficients of f and g that are at least q_f (they are K non-zero coefficients of f_2 and g_2). Denote positions of these coefficients in f by $a_1, \ldots, a_K$ and in g by $b_1, \ldots, b_K$. We call these coefficients "large," since all others are less than p, which is less than q_f.

A valid signature is $s = f * w \pmod{q}$. Its kth coefficient is

$$s[k] = \sum_{i=0}^{N-1} f[i]w[k-i] = \sum_{i=0}^{N-1} (f_1[i]w_1[k-i] + q_f f_2[i]w_1[k-i] + q_w f_1[i]w_2[k-i]) \pmod{q}. \tag{2}$$

Notice that all but K coefficients of f_2 are zero. It means that a fairly large summand $q_f f_2[a_i] = q_f$ in the sum (2) is only present if $w_1[k-a_i]$ is set to one. We conjecture that if $s[k]$ is large, then $w[k-a_i]$ is most likely equal 1. To quantify this observation we want to estimate the correlation between $s[k]$ and $w[k-a_i]$.

Fix f and let w_1, w_2 be sampled independently at random according to their respective distributions. We ignore step 3 of the signing algorithm that introduces some distortion in the distribution of the coefficients of w. The expected value of a coefficient of s is approximately

$$E(s[k]) \approx \left(\sum_{i=0}^{N-1} f_1[i]\right)\frac{1}{2} + q_f\frac{K}{2} + q_w\left(\sum_{i=0}^{N-1} f_1[i]/N\right)K = \frac{d_{f_1} + Kq_f}{2} + \frac{Kq_w d_{f_1}}{N},$$

where $d_{f_1} = \sum_{i=0}^{N-1} f_1[i]$. If $w_1[k-a_i] = 1$, then the expected value of $s[k]$ conditioned on this event is

$$E(s[k] \mid w_1[k-a_i] = 1) \approx \frac{d_{f_1} + f_1[k-a_i]}{2} + q_f \frac{K+1}{2} + \frac{K q_w d_{f_1}}{N}.$$

The variance of a single coefficient is approximately

$$\mathrm{Var}(s[k]) \approx \left(\sum_{i=0}^{N-1} f[i]^2 \right) \frac{1}{4} + K \frac{d_{f_1}}{N} (1 - \frac{d_{f_1}}{N}) q_w^2. \tag{3}$$

Now we can compute the covariance and the correlation coefficient of two random variables. One random variable is $s[k]$ as a function of w and another is $w_1[k-a_i]$ for a fixed i.

$$\begin{aligned} \mathrm{cov}(s[k], w_1[k-a_i]) &= E(s[k] w_1[k-a_i]) - E(s[k]) E(w_1[k-a_i]) \\ &= \frac{1}{2} E(s[k] \mid w_1[k-a_i] = 1) - E(s[k]) \frac{1}{2} = \frac{1}{4} (f_1[k-a_i] + q_f). \end{aligned}$$

The correlation coefficient is

$$\rho(s[k], w_1[k-a_i]) = \mathrm{cov}(s[k], w_1[k-a_i]) / \sqrt{\mathrm{Var}(s[k]) \mathrm{Var}(w_1[k-a_i])}.$$

The maximum on the variance (3) is delivered by $\sum_{i=0}^{N-1} f[i]^2 = K(q_f+1)^2 + (N-K)$ and $d_{f_1} = N/2$. The correlation coefficient is at least

$$\rho(s[k], w_1[k-a_i]) \geq \frac{q_f/4}{\sqrt{(\frac{1}{4}(K(q_f+1)^2 + (N-K)) + \frac{1}{4} K q_w^2) \frac{1}{4}}}.$$

Therefore for the concrete parameters (1)

$$\rho(s[k], w_1[k-a_i]) \geq 0.42.$$

This correlation is very high and can be detected in a few trials. The problem now is to exploit this dependency since w_1 is kept secret and picked at random anew for every signature.

3.2 Partial recovery of the key

The previous section established a lower bound on the correlation between $s[k]$ and $w_1[k-a_i]$ for some i. It means that when $s[k]$ is large then $w_1[k-a_i]$ is more likely to be one. It also works the other way round. If $s[k]$ is small, then $w_1[k-a_i]$ is zero with probability more than 1/2. Clearly, this argument also applies to $t \equiv g * w \pmod{q}$, since the product has exactly the same form as $s \equiv f * w \pmod{q}$.

Take two coefficients $s[k]$ and $t[k']$ such that $s[k]$ is smaller than the median and $t[k']$ is larger than the median. Fix some integer i and j between 1 and K. We expect that $w_1[k-a_i]$ is more likely to be zero than one, while $w_1[k'-b_j]$ tends to be one rather than zero. But it cannot be true simultaneously if $k - a_i = k' - b_j$ and we thus have that

$$k - a_i \neq k' - b_j,$$

hence

$$b_j - a_i \neq k' - k. \tag{4}$$

These inequalities are true only probabilistically, i.e., $b_j - a_i \neq k' - k$ with probability better than $1 - 1/N$, which is the case when the coefficients are chosen uniformly and independently at random.

Notice that in (4) the unknown and known variables are separated and we can conjecture about $b_j - a_i$ by observing $k' - k$. It suggests the following line of attack on the scheme.

Partial recovery of $a_1, \ldots, a_K$ and $b_1, \ldots, b_K$.

Input: A list of valid signatures $(s_1, m_1), \ldots, (s_M, m_M)$ (hashes $m_1, \ldots, m_M$ are ignored by the algorithm).

Output: An unordered list of length K^2 that contains all $a_i - b_j$ for $1 \leq i, j \leq K$, where a_i are positions of the non-zero coefficients of f_2 and b_j are positions of the non-zero coefficients of g_2.

Step 1. Compute $t_l \equiv h * s_l \bmod q$ for all $1 \leq l \leq M$.

Step 2. Put coefficients of s_l and t_l into the interval $[A, A+q-1]$, where $A = \frac{N}{4} + \frac{q_f K}{2} + \frac{q_w K}{2} - \frac{q}{2}$.

Step 3. Choose $k_1, \ldots, k_{N/2}$—the positions of the $N/2$ largest coefficients of s_l and $k'_1, \ldots, k'_{N/2}$—the positions of the $N/2$ smallest coefficients of t_l.

Step 4. For every pair k_i and k'_j compute their difference modulo N. Keep track of frequences of all possible differences (numbers between 0 and $N - 1$).

Step 5. Aggregate statistics collected on step 0 over all pairs (s_l, t_l).

Step 4. Choose the K^2 *least* frequent numbers. This is a tentative list of $a_i - b_j \bmod N$ for all $1 \leq i, j \leq K$.

We do not give a formal proof of correctness of this algorithm. The intuition behind the algorithm follows from (4). We know that the difference of two coefficient k and k' selected as in step 3 is less likely to coincide with $a_i - b_j \bmod N$ than with a random number between 0 and $N - 1$. Therefore after sufficiently large number of trials the differences $a_i - b_j \bmod N$ will emerge from the list $1, \ldots, N-1$—they will occur less frequently among differences $k - k' \bmod N$.

Experimental evidence suggests that the probabilities converge very fast and for the parameters (1) less than a hundred signatures are sufficient to find out $a_i - b_j \bmod N$ for $1 \leq i, j \leq K$.

3.3 Revealing more bits of the key

The result of the partial recovery of the key presented in the previous section is a complete list of $a_i - b_j \bmod N$ (and probably a few more numbers, since there are K^2 numbers in total) for unknown values $\{(a_i, b_i)\}_{i=1}^K$. Our goal is to restore the unknown values.

Notice that every circular shift $\{((a_i + M) \bmod N, (b_i + M) \bmod N)\}_{i=1}^K$ or a swap $(\{(N-1-b_i, N-1-a_i)\}_{i=1}^K)$ results in the same pattern of $a_i - b_j \bmod N$. Therefore the best we can do is to come up with a list of $2N$ possibilities for the unknown values. A simple backtrack algorithm finds one solution (and immediately other $2N - 1$) surprisingly fast for the parameters (1). The problem of studying the performance of this algorithm or even proving that for N and K of interest there are at most $2N$ possible solutions appears to be hard.

Suppose we know the list $\{(a_i, b_i)\}_{i=1}^K$, which encodes f_2 and g_2. [HPS98] presents a lattice attack on a NTRU private key that also applies to the NSS. The $2N \times 2N$-matrix

$$\left(\begin{array}{cccc|cccc}
\alpha & 0 & \ldots & 0 & h_0 & h_1 & \ldots & h_{N-1} \\
0 & 0 & \ldots & 0 & h_{N-1} & h_0 & \ldots & h_{N-2} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \alpha & h_1 & h_2 & \ldots & h_0 \\
\hline
0 & 0 & \ldots & 0 & q & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & q & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & q
\end{array}\right)$$

defines a lattice parameterized by α that contains the vector $(\alpha f, g)$. This vector is most likely the shortest vector in the lattice but the parameters of the scheme are chosen to make the task of finding this vector infeasible. If f_2 and g_2 are known, then the vector $(\alpha f_2, g_2)$ is a very good approximation to the shortest vector. It might be sufficient for an attack but to the best of our knowledge no algorithm exists that may take advantage of this approximation.

Instead, we use our knowledge of f_2 to conjecture about w_1 based on this formula:

$$s \equiv f_1 * w_1 + q_f f_2 * w_1 + q_w f_1 * w_2 \pmod{q}.$$

If, as we have shown, there is a strong correlation between coefficients $s[k + a_i]$ and $w_1[k]$, we can guess $w_1[k]$ by observing $s[k+a_1]$,...,$s[k+a_K]$. As we know w_1, we may then detect a more subtle correlation between coefficients of f_1 and s. Experiments show that we can thus reveal over two thirds of the bits of f_1. It still falls short from a total break of the system, but demonstrates that a partial key exposure may be used to uncover more bits of the key.

4 Conclusion

We presented an attack on the preliminary version on the NSS that reveals a significant part of the private key. This attack requires interception of less than a hundred signatures. We may conclude that the NSS as presented in [HS00] is completely insecure.

Our attack fails against the next version of the NSS [HPS00]. The reason for it is that in the new NSS the polynomials forming the private key do not have a few large coefficients, which undermines our method of detecting differences between coefficients. In addition, a special algorithm for choosing w deliberately hides "features" of the private polynomials.

References

[HPS96] J. Hoffstein, J. Pipher, J.H. Silverman, "NTRU: A new high speed public key cryptosystem," preprint; presented at the rump session of Crypto'96.

[HPS98] J. Hoffstein, J. Pipher, J.H. Silverman, "NTRU: A new high speed public key cryptosystem," in Algorithmic Number Theory (ANTS III), LNCS 1423, pp. 267–288, 1998.

[HPS00] J. Hoffstein, J. Pipher, J.H. Silverman, "NSS: The NTRU signature scheme," preprint; available from `www.ntru.com`, November 2000.

[HS00] J. Hoffstein, J.H. Silverman, "NSS: The NTRU signature scheme. Preliminary version—August, 2000," preprint.