CR01-Report

Cryptanalysis of GGH

Krishna Acharya

January 2020

1 Back to the 90s

Ajtai's [1] results for SVP, CVP being NP-hard under randomized reductions have revived the crypto communities interest in Lattice based systems. Until then most public-key cryptosystems were based on dlog, integer factorization hardness assumptions. Ajtai-Dwork [2] and Goldreich-Goldwasser-Halevi cryptosystems [3] are current frontrunners for Lattice based PKC. [6] have already cryptanalszed AD, showing that realistic implementations are insecure. Nguygen in this paper attacks GGH.
Prior attempts : For dimension 200, message recovered using embedding technique(described next). The dimension 250,300,350, 400 challenges set by GGH are yet to be solved.

1.1 Embedding technique - A general method for CVP

$(\mathbf{b_1}, \ldots, \mathbf{b_n})$ is a basis of Lattice L, $\mathbf{c}$ is the vector to which we have to find a closest lattice vector.

$$L' = \begin{pmatrix} - & \mathbf{b_1} & - & 0 \\ & \vdots & & \\ - & \mathbf{b_n} & - & 0 \\ - & \mathbf{c} & - & 1 \end{pmatrix}$$

The observation here is that L' and L have same determinant, nearly same dimension, $\therefore$ a reasonable argument is that shortest vectors for both L and L' have about the same length. Say a vector $\mathbf{v} \in L$ minimizes distance to $\mathbf{c}$, $(\mathbf{c} - \mathbf{v}, 1)$ is very likely to be a shortest vector in L'. The hope is that this is indeed a shortest vector in L', so that SVP in $n+1$ dim solves CVP in dim n

2 GGH Cryptosystem

Security parameter is $(n, \sigma) \in \mathbb{N}^2$

- **Keygen**: Choose good basis (private key) * $R = \mathbf{v_1}, \ldots, \mathbf{v_n}$
  Transform R to a bad basis (public key) † $B = UR$, $U = U_1 \ldots \times U_k$, each U_k is sufficiently small and unimodular
- **Encryption**: $m \in Z^n$, but restricted to within some bounds ‡ Ciphertext $\mathbf{c} = \mathbf{m}B + \mathbf{e}$. Where $\mathbf{e}$ is error vector chosen uniformly from $\{\sigma, -\sigma\}^n$
- **Decryption**: $\lfloor \mathbf{c}R^{-1} \rceil RB^{-1}$

Of course there is a possibility that decryption does not give us back $\mathbf{m}$ but GGH showed that

Theorem 1 *For R: a private basis and $\frac{\gamma}{\sqrt{n}}$: denoting the max L_∞ norm of columns in R^{-1}, the probability of decryption error is bounded by $2ne^{-1/(8\sigma^2\gamma^2)}$*

Implication: CVP instances are harder for larger σ, but larger σ gives more possibility for error during decryption. So based on some tradeoffs and experiments, GGH stuck with $\sigma = 3$ and conjectured dimension $n \geq 300$ being secure.

*For example one with close to 1 Hadamard ratio, in literature we call it reduced basis
†referred to as non-reduced basis
‡In the GGH challenge $m \in [-128, \ldots, +127]^n$ corresponding to 8 bits.

3 Attack: Leaking Remainders

An insightful observation is that for $\mathbf{s} = (\sigma, \ldots, \sigma)$, $\mathbf{e} + \mathbf{s}$ disappears $\pmod{2\sigma}$

$$\begin{aligned} \mathbf{c} &= \mathbf{m}B + \mathbf{e} \\ \mathbf{c} + \mathbf{s} &\equiv \mathbf{m}B \pmod{2\sigma} \end{aligned} \tag{1}$$

Implication: $\mathbf{c}, \mathbf{s}, \sigma$ and B the public basis is known. B was obtained by multiplication by Unimodulars, as such it can be thought that its entries $b_{ij} \xleftarrow{\$} \mathbb{Z}_{2\sigma}$.
The classic problem of solving $\mathbf{x}B \equiv \mathbf{y} \pmod{N}$ reappears. For our case, we know there is atleast 1 solution i.e $\mathbf{m}$. In general two solutions for this system of congruences differ by the kernel of B.
Two specific cases are tackled:

- **Case 1**: $det(B)$ is coprime to $N \implies$ unique solution found by $\mathbf{x} = \mathbf{y}B^{-1} \pmod{N}$
- **Case 2**: $det(B)$ is not coprime to N but B has a kernel of small dimension.

According to the two following theorems, the cases above occur for a sufficiently high proportion of matrices $\pmod{N}$, that an attack is possible.

Theorem 2 *N is some positive integer, $p_1 \ldots p_l$ are distinct prime factors of N, For the $n \times n$ ring with entries $\in \mathbb{Z}_N$, the proportion of matrices with determinant coprime to N is equal to:*

$$prop(N, n) = \prod_{i=1}^{l} \prod_{k=1}^{n} (1 - p_i^{-k}) \tag{2}$$

In particular for $N = 6$ [§], when $\lim_{n \to \infty} prop(N, n) \approx 0.162$. This limit is achieved rapidly, and for $n \geq 200$ can be taken as known constant.

Theorem 3 *For the $n \times n$ matrix with entries in $\mathbb{F}_q$, where $\mathbb{F}_q$ is a Finite field, q: prime power.*

- *The proportion of elements with one-dim kernel:*

$$prop_1(q, n) = \frac{q}{(q-1)^2}(1 - q^{-n}) \prod_{k=1}^{n} (1 - q^{-k}) \tag{3}$$

- *The proportion of elements with two-dim kernel:*

$$prop_2(q, n) = \frac{q^2}{(q-1)^2(q^2-1)^2}(1 - q^{1-n} - q^{-n} + q^{1-2n}) \prod_{k=1}^{n} (1 - q^{-k}) \tag{4}$$

Implication: Consider if the entries of B come form $\mathbb{F}_q$, We know that d the kernel dimension is one-dim or two-dim with sufficiently high probability [¶]. We know there are p^d number of solutions(corresponding to Particular solution $(x_0) + x$, where $x \in kernel(B)$). Its important to note that the actual entries of B come from $\mathbb{Z}_N$. if N is composite, but square free, Theorem 3 is useful, we can obtain solution $\pmod{p_i}$ for of the individual primes [‖] and then recombined[**]. For non square free composites, there are other heuristic techniques but no exact estimates for proportions like earlier.

4 For GGH, CVP instance is actually easier

Let $\mathbf{m_{2\sigma}}$ denote the solution we obtained from the linear system of concurrences $\pmod{2\sigma}$.
$\therefore \mathbf{m} - \mathbf{m_{2\sigma}} = 2\sigma\mathbf{z},\ \mathbf{z} \in \mathbb{Z}^n$

$$\begin{aligned} \mathbf{c} - \mathbf{m_{2\sigma}}B &= (\mathbf{m} - \mathbf{m_{2_\sigma}})B + \mathbf{e} \\ \tfrac{\mathbf{c} - \mathbf{m_{2\sigma}}B}{2\sigma} &= \mathbf{z}B + e/2\sigma. \end{aligned} \tag{5}$$

Implication: The error vector $e/2\sigma \in \{\pm\frac{1}{2}\}^n$ and of length $\sqrt{n/4}$ v/s the when it was earlier $\in \{\pm\sigma\}^n$ and length $\sigma\sqrt{n}$. We can even avoid working with rationals by multiplying the earlier equation by 2, so the new error vector $\in \{\pm 1\}^n$, and we have doubled basis entries.

[§] Recall $2\sigma = 6$
[¶] See the main article for actual values, its nearly 0.4. To obtain particular solution embedding technique can be used
[‖] as $N = p_1 \times \ldots p_l$
[**] Chinese remainder thm

5 Results and Possible repair

Using the techniques [4] solved the challenge for $n = 200, 250, 300, 350$ challenging the claim by GGH that $n \geq 300$ is secure.
The two weaknesses for GGH were:

- W1: Error vector $\mathbf{e}$ is much shorter than the lattice vectors [††].
- W2: $\mathbf{e}$ was chosen uniformly from $\{\pm\sigma\}^n$, thus the congruence relations were formed.

No direct solution to W1, as GGH was specified to have short error vector. For repairing W2, we could instead choose $\mathbf{e} \in [-\sigma, \ldots, \sigma]^n$, but this again has other drawbacks.

6 Back to the future, Conclusion

After this papers attack, GGH lay dormant, until NTRUSign came about and used the GGH system but with more compact NTRU lattices. [7] discovered a lattice reduction technique for a class of Hypercubic lattices, this was important as transcript analysis of GGH and NTRUSign gave rise to these lattices. Inspired by this [5] came up with their "Learning a Paralleopiped" i.e given many random points uniformly distributed over an unknown n-dimensional parallelepiped they recovered the parallelepiped or an approximation of it. This naturally gave rise to a transcript attack and they were able recover the secret key in the signature analogue of all the GGH encryption challenges almost half of NTRUSign parameters. This method of learning proved to be quite useful, in fact another adaptation of GGH for signing called DRS sign was subject to a similar statistical attack [8]. So much for GGH! atleast it led to such elegant cryptanalysis techniques.

References

[1] Miklós Ajtai. The shortest vector problem in l2 is np-hard for randomized reductions (extended abstract). *Electronic Colloquium on Computational Complexity (ECCC)*, 4, 1997.

[2] Miklós Ajtai and Cynthia Dwork. A public-key cryptosystem with worst-case/average-case equivalence. In *STOC '97*, 1996.

[3] Oded Goldreich, Shafi Goldwasser, and Shai Halevi. Public-key cryptosystems from lattice reduction problems. *Electronic Colloquium on Computational Complexity (ECCC)*, 3, 1996.

[4] Phong Nguyen. Cryptanalysis of the goldreich-goldwasser-halevi cryptosystem from crypto '97. In Michael Wiener, editor, *Advances in Cryptology — CRYPTO' 99*, pages 288–304, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

[5] Phong Q. Nguyen and Oded Regev. Learning a parallelepiped: Cryptanalysis of ggh and ntru signatures. *Journal of Cryptology*, 22(2):139–160, Apr 2009.

[6] Phong Q. Nguyen and Jacques Stern. Cryptanalysis of the ajtai-dwork cryptosystem. In *CRYPTO*, 1998.

[7] Michael Szydlo. Hypercubic lattice reduction and analysis of ggh and ntru signatures. In *EUROCRYPT*, 2003.

[8] Yang Yu and Léo Ducas. Learning strikes again: the case of the drs signature scheme. In *IACR Cryptology ePrint Archive*, 2018.

[††] Lattice gap, leads to embedding based attack