

Provably Solving the Hidden Subset Sum Problem via Statistical Learning

Jean-Sébastien Coron and **Agnese Gini**

University of Luxembourg

MathCrypt 2021

August 15th, 2021



Subset Sum Problem

$$h = \alpha_1 x_1 + \cdots + \alpha_n x_n \pmod{q}$$

with $x_1, \dots, x_n \in \{0, 1\}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{Z}/q\mathbb{Z}$.

Given q, h and $\alpha_1, \dots, \alpha_n$, recover x_1, \dots, x_n .

Hidden Subset Sum Problem

$$h_1 = \alpha_1 x_{1,1} + \cdots + \alpha_n x_{n,1} \pmod{q}$$

\vdots

$$h_m = \alpha_1 x_{1,m} + \cdots + \alpha_n x_{n,m} \pmod{q}$$

with $x_{i,j} \in \{0, 1\}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{Z}/q\mathbb{Z}$.

Given q and h_1, \dots, h_m , recover $\alpha_1, \dots, \alpha_n$ and $x_{i,j}$ for $i \in [n]$ and $j \in [m]$.

The weights α_i 's are hidden!!

Hidden Subset Sum Problem

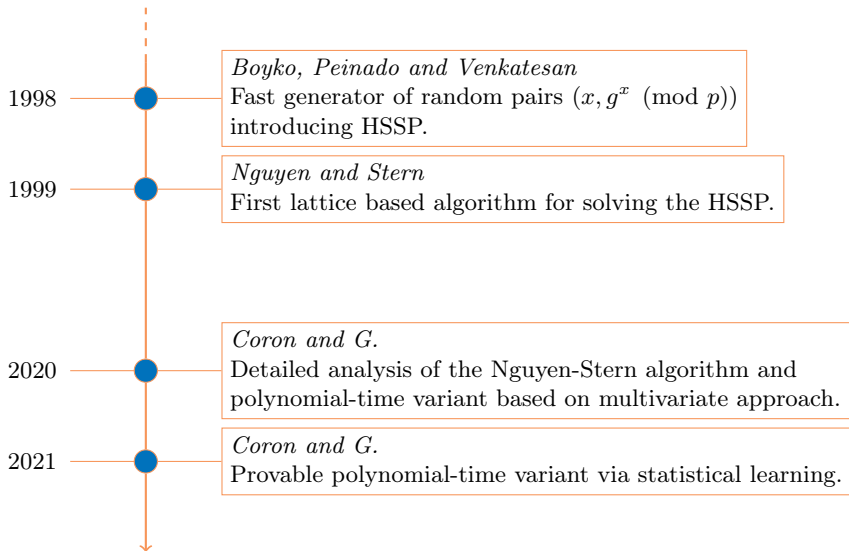
Let q be an integer, and let $\alpha_1, \dots, \alpha_n$ be random integers in $\mathbb{Z}/q\mathbb{Z}$. Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{Z}^m$ be random vectors with components in $\{0, 1\}$. Let $\mathbf{h} \in \mathbb{Z}^m$ satisfying:

$$\mathbf{h} = \alpha_1 \mathbf{x}_1 + \dots + \alpha_n \mathbf{x}_n \pmod{q}$$

Given q and \mathbf{h} , recover the integers α_i 's and the vectors \mathbf{x}_i 's.

A diagram illustrating the equation $\mathbf{h} = \alpha \mathbf{X} \pmod{q}$. The variable \mathbf{h} is enclosed in a green rounded rectangle. The variable α is enclosed in a pink rounded rectangle. The variable \mathbf{X} is enclosed in a blue rounded rectangle. The equation is written as $\mathbf{h} = \alpha \mathbf{X} \pmod{q}$.

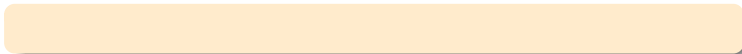
Timeline



The goal

A proven polynomial algorithm for solving HSSP

The strategy



A proven polynomial algorithm for solving HSSP

A Sketch

$$\mathbf{h} = \alpha \mathbf{X} \pmod{q}$$

Input: \mathbf{h}, q

Output: \mathbf{X}, α

A Sketch

$$\mathbf{h} = \alpha \mathbf{X} \pmod{q}$$

Input: \mathbf{h}, q

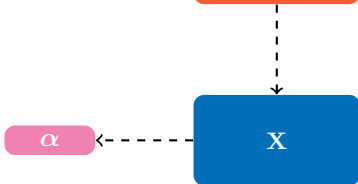
Output: \mathbf{X}, α

Algorithm:

Step 1

$$\mathbf{h} = \gamma \mathbf{C} \pmod{q}$$

Step 2



A Sketch

$$\mathbf{h} = \alpha \mathbf{X} \pmod{q}$$

Input: \mathbf{h}, q

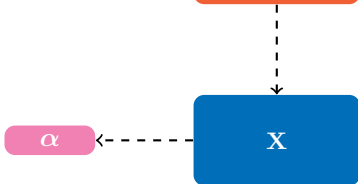
Output: \mathbf{X}, α

Algorithm:

Step 1 : orthogonal lattice attack ✓

$$\mathbf{h} = \gamma \mathbf{C} \pmod{q}$$

Step 2 : ?



The strategy



A proven polynomial algorithm for Step 2



A proven polynomial algorithm for solving HSSP

A proven polynomial algorithm for Step 2

Step 2



Step 2



1. The \mathbf{x}_i 's generate a full-rank sublattice of $\mathcal{L}(C)$.

$$X = W C$$

A diagram representing the equation $X = W C$. On the left is a blue rounded rectangle labeled X . To its right is an equals sign. Further right is a purple rounded rectangle labeled W , followed by an orange rounded rectangle labeled C .

2. If $V = W^{-1}$, then the pairs of columns satisfy

$$V \tilde{x}_i = \tilde{c}_i$$

A diagram representing the equation $V \tilde{x}_i = \tilde{c}_i$. On the left is a grey rounded rectangle labeled V . To its right is a blue rounded rectangle labeled \tilde{x}_i . To the right of that is an equals sign. On the far right is an orange rounded rectangle labeled \tilde{c}_i .

A diagram illustrating the relationship between a matrix \mathbf{V} , a vector $\tilde{\mathbf{x}}_i$, and a column $\tilde{\mathbf{c}}_i$ of a matrix \mathbf{C} . The matrix \mathbf{V} is represented by a gray rounded square. The vector $\tilde{\mathbf{x}}_i$ is represented by a blue rounded rectangle. The column $\tilde{\mathbf{c}}_i$ is represented by an orange rounded rectangle. An equals sign is placed between the vector and the column, indicating that the product of \mathbf{V} and $\tilde{\mathbf{x}}_i$ equals $\tilde{\mathbf{c}}_i$.

$$\mathbf{V} \tilde{\mathbf{x}}_i = \tilde{\mathbf{c}}_i$$

3. The m columns $\tilde{\mathbf{c}}_i$ of \mathbf{C} are samples from the *discrete parallelepiped* associated to \mathbf{V} :

$$\mathcal{P}_{\{0,1\}}(\mathbf{V}) := \{\mathbf{V}\mathbf{x} \mid \mathbf{x} \in \{0,1\}^n\}.$$

A diagram illustrating the relationship between a matrix \mathbf{V} , a sample vector $\tilde{\mathbf{x}}_i$, and a column vector $\tilde{\mathbf{c}}_i$. The matrix \mathbf{V} is represented by a grey rounded rectangle. The sample vector $\tilde{\mathbf{x}}_i$ is represented by a blue rounded rectangle. The column vector $\tilde{\mathbf{c}}_i$ is represented by an orange rounded rectangle. An equals sign is placed between the sample vector and the column vector, indicating that $\tilde{\mathbf{x}}_i = \mathbf{V} \tilde{\mathbf{c}}_i$.

3. The m columns $\tilde{\mathbf{c}}_i$ of \mathbf{C} are samples from the *discrete parallelepiped* associated to \mathbf{V} :

$$\mathcal{P}_{\{0,1\}}(\mathbf{V}) := \{\mathbf{V}\mathbf{x} \mid \mathbf{x} \in \{0,1\}^n\}.$$

Discrete Hidden Parallelepiped Problem

Given $\text{poly}(n)$ independent samples from the uniform distribution over $\mathcal{P}_{\{0,1\}}(\mathbf{V})$, recover the columns of \mathbf{V} .

The strategy

A proven polynomial algorithm for solving DHPP



A proven polynomial algorithm for Step 2



A proven polynomial algorithm for solving HSSP

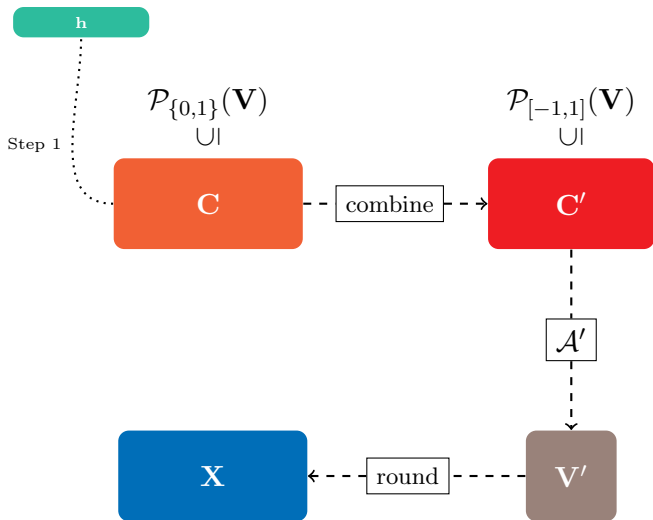
Hidden Parallelepiped Problem [NR09]

$$\mathcal{P}_{[-1,1]}(\mathbf{V}) = \{\mathbf{V}\mathbf{x} : \mathbf{x} \in [-1, 1]^n\}.$$

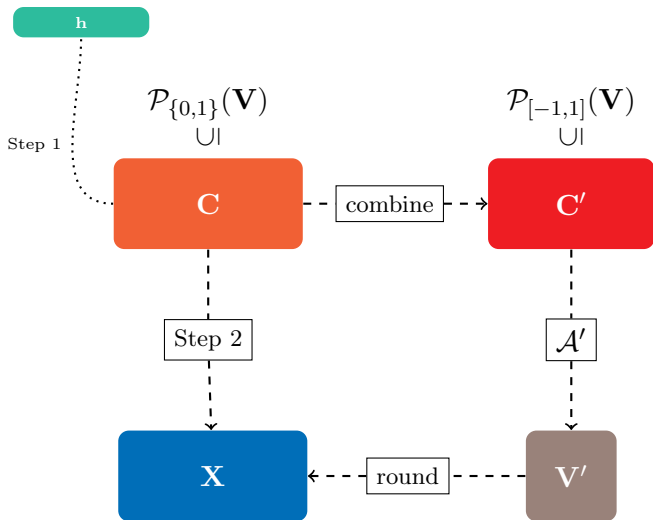
Given a sequence of $\text{poly}(n)$ independent samples from the uniform distribution over $\mathcal{P}_{[-1,1]}(\mathbf{V})$, the goal is to recover a good approximation of the columns of $\pm\mathbf{V}$.

↑
Solvable via a statistical learning technique!

Our algorithm for disclosing X



Our algorithm for disclosing X



The goal

A proven polynomial algorithm for solving DHPP



A proven polynomial algorithm for Step 2



A proven polynomial algorithm for solving HSSP



Hidden Linear Combination Problem

Let q be an integer, and let $\alpha_1, \dots, \alpha_n$ be random integers in $\mathbb{Z}/q\mathbb{Z}$. Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{Z}^m$ be random vectors with components in $\{0, \dots, B\}$. Let $\mathbf{h} \in \mathbb{Z}^m$ satisfying:

$$\mathbf{h} = \alpha_1 \mathbf{x}_1 + \dots + \alpha_n \mathbf{x}_n \pmod{q}$$

Given q, B and \mathbf{h} , recover the integers α_i 's and the vectors \mathbf{x}_i 's.



The diagram illustrates the equation $\mathbf{h} = \alpha \mathbf{X} \pmod{q}$ using colored boxes. On the left, a green rounded rectangle contains the vector \mathbf{h} . This is followed by an equals sign. To the right of the equals sign is a pink rounded rectangle containing the scalars α , followed by a blue rounded rectangle containing the vectors \mathbf{X} . To the right of the blue box is the expression \pmod{q} .

Conclusions

Theorem

There exists an algorithm for solving the hidden subset sum problem with constant probability in polynomial time, using $\text{poly}(n)$ samples, for any prime integer q of bitsize at least $4n^2 \log(n)$.

- Attacks for Hidden Linear Combination Problem

approach	complexity	status
lattice	$2^{\Omega(n)} \cdot \log^{\mathcal{O}(1)} B$	heuristic
multivariate	$\mathcal{O}(n^{B+1})$	heuristic
statistical	$\text{poly}(n, B)$	heuristic/ proven for $B = 1$

Thank you for your attention!

Full paper at
<https://ia.cr/2021/1007>

Bonus question: can we find an attack $\text{poly}(n, \log B)$?

References

- [CG20] Jean-Sébastien Coron and Agnese Gini. A polynomial-time algorithm for solving the hidden subset sum problem. In *Advances in Cryptology - CRYPTO 2020 - 40th Annual International Cryptology Conference, CRYPTO 2020, Santa Barbara, CA, USA, August 17-21, 2020, Proceedings, Part II*, 2020. Full version available at <https://eprint.iacr.org/2020/461>.
- [NR09] Phong Q. Nguyen and Oded Regev. Learning a parallelepiped: Cryptanalysis of GGH and NTRU signatures. *J. Cryptology*, 22(2):139–160, 2009.
- [NS99] Phong Q. Nguyen and Jacques Stern. The hardness of the hidden subset sum problem and its cryptographic implications. In *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 1999, Proceedings*, pages 31–46, 1999.