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

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

Hidden Subset Sum Problem

$$
\begin{aligned}
h_1 &= \alpha_1 x_{1,1} + \cdots + \alpha_n x_{n,1} && \pmod{q} \\
&\vdots \\
h_m &= \alpha_1 x_{1,m} + \cdots + \alpha_n x_{n,m} && \pmod{q}
\end{aligned}
$$

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

Given q and $h_1, \ldots, h_m$, recover $\alpha_1, \ldots, \alpha_n$ 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 $\alpha_1, \ldots, \alpha_n$ be random integers in $\mathbb{Z}/q\mathbb{Z}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{Z}^m$ be random vectors with components in $\{0, 1\}$. Let $\mathbf{h} \in \mathbb{Z}^m$ satisfying:

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

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

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

Timeline

1998 — *Boyko, Peinado and Venkatesan*
Fast generator of random pairs $(x, g^x \pmod p)$ introducing HSSP.

1999 — *Nguyen and Stern*
First lattice based algorithm for solving the HSSP.

2020 — *Coron and G.*
Detailed analysis of the Nguyen-Stern algorithm and polynomial-time variant based on multivariate approach.

2021 — *Coron and G.*
Provable polynomial-time variant via statistical learning.

The goal

A proven polynomial algorithm for solving HSSP

The strategy

⇓

⇓

A proven polynomial algorithm for solving HSSP

A Sketch

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

Input: $\mathbf{h}$, q
Output: $\mathbf{X}, \boldsymbol{\alpha}$

A Sketch

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

Input: $\mathbf{h}$, q
Output: $\mathbf{X}, \boldsymbol{\alpha}$
Algorithm:

Step 1

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

Step 2

$$\boldsymbol{\alpha} \leftarrow \mathbf{X}$$

A Sketch

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

Input: $\mathbf{h}$, q
Output: $\mathbf{X}, \boldsymbol{\alpha}$
Algorithm:

Step 1 : orthogonal lattice attack ✓

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

Step 2 : ?

$$\begin{tikzcd} & \mathbf{C} \arrow[d, dashed] \\ \boldsymbol{\alpha} & \mathbf{X} \arrow[l, dashed] \end{tikzcd}$$

The strategy

$\Downarrow$

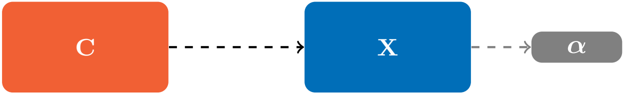
A proven polynomial algorithm for Step 2

$\Downarrow$

A proven polynomial algorithm for solving HSSP

A proven polynomial algorithm for Step 2

Step 2



Step 2

$$\mathbf{C} \dashrightarrow \mathbf{X} \dashrightarrow \alpha$$

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

$$\mathbf{X} = \mathbf{W}\,\mathbf{C}$$

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

$$\mathbf{V}\,\tilde{\mathbf{x}}_\mathbf{i} = \tilde{\mathbf{c}}_\mathbf{i}$$

$$\mathbf{V}\,\tilde{\mathbf{x}}_\mathbf{i} = \tilde{\mathbf{c}}_\mathbf{i}$$

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

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

$$\mathbf{V}\,\tilde{\mathbf{x}}_{\mathbf{i}} = \tilde{\mathbf{c}}_{\mathbf{i}}$$

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 $\mathsf{poly}(n)$ 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

$$\Downarrow$$

A proven polynomial algorithm for Step 2

$$\Downarrow$$

A proven polynomial algorithm for solving HSSP

Hidden Parallelepiped Problem [NR09]

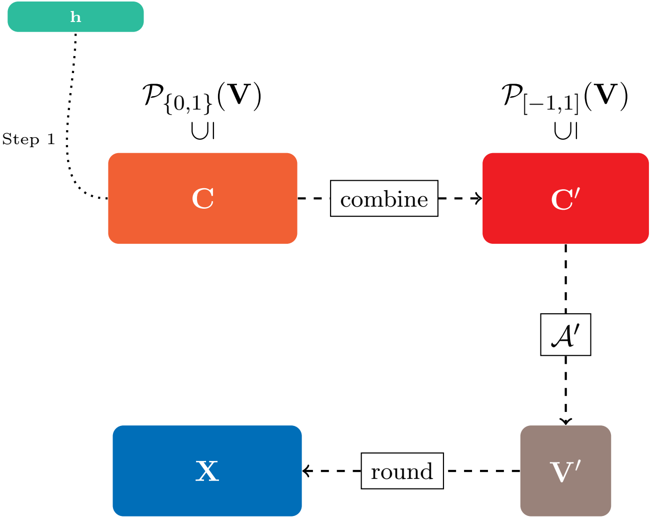
$$\mathcal{P}_{[-1,1]}(\mathbf{V}) = \{\mathbf{V}\mathbf{x} \ : \ \mathbf{x} \in [-1,1]^n\}.$$

Given a sequence of $\mathsf{poly}(n)$ 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 $\pm\mathbf{V}$.

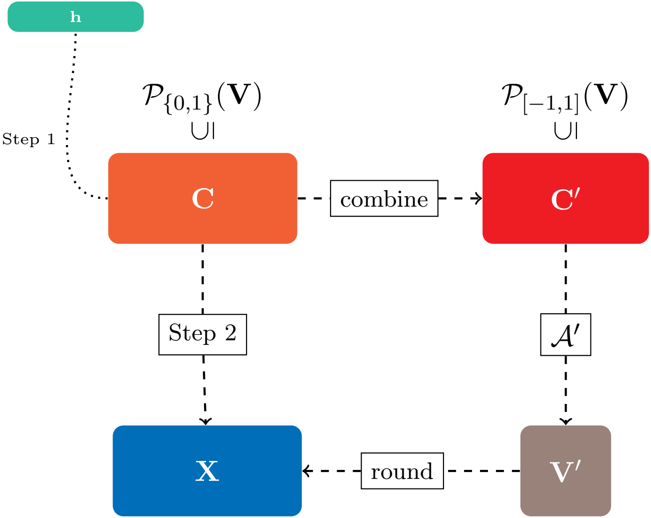
$\uparrow$

$\boxed{\textbf{Solvable via a statistical learning technique!}}$

Our algorithm for disclosing X



Our algorithm for disclosing X



The goal

A proven polynomial algorithm for solving DHPP

$\Downarrow$

A proven polynomial algorithm for Step 2

$\Downarrow$

A proven polynomial algorithm for solving HSSP

✓

Hidden Linear Combination Problem

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

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

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

$$\mathrm{h} = \boldsymbol{\alpha} \, \mathbf{X} \pmod q$$

Conclusions

Theorem

There exists an algorithm for solving the hidden subset sum problem with constant probability in polynomial time, using $\mathsf{poly}(n)$ 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	$\mathsf{poly}(n, B)$	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 $\mathsf{poly}(n, \log B)$?

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.