The Modular Subset Product Problem and Obfuscation

Lukas Zobernig

The University of Auckland

Outline

- Motivation
- Preliminaries & Obfuscation
- Modular Subset Product Problem
- Constructions
- Post-Quantum Hardness

Motivation

Can we:

- securely **encode** and **match** fingerprints ...
- ... as well as other biometric features (iris scans, DNA, etc.)?

Example

$$x = (F, i, N, g, e, r, p, R, i, n, t),$$
$$y = (F, I, n, g, e, r, p, r, i, n, t)$$

Some keywords:

- Secure sketch.
- Fuzzy extractor.
- View as obfuscation problem.

Types of Obfuscators

Denote obfuscator by $\mathcal{O}$, adversary by $\mathcal{A}$, simulator by $\mathcal{S}$, negligible function by ϵ.

Definition (Distributional Virtual Black-Box Obfuscator)

For every $\mathcal{A}$, there exists $\mathcal{S}$, such that for every predicate φ:

$$\left| \Pr_{P \leftarrow D_\lambda, \mathcal{O}, \mathcal{A}} [\mathcal{A}(\mathcal{O}(P)) = \varphi(P)] - \Pr_{P \leftarrow D_\lambda, \mathcal{S}} \left[\mathcal{S}^P(|P|) = \varphi(P) \right] \right| \leq \epsilon(\lambda).$$

Hence, a VBB obfuscated program $\mathcal{O}(P)$ does not reveal anything more than would be revealed from having **black-box** access to the program P itself.

Definition (Input Hiding Obfuscator)

For every $\mathcal{A}$, there exists ϵ, such that for every $n \in \mathbb{N}$ and for every *auxiliary input* α:

$$\Pr_{P \leftarrow \mathcal{P}_n} [P(\mathcal{A}(\alpha, \mathcal{O}(P))) = 1] \leq \epsilon(n).$$

Preliminaries

We need a *good* class of programs to obfuscate.

Definition (Evasive Program Collection)

Let $\mathcal{P} = \{\mathcal{P}_n\}_{n\in\mathbb{N}}$ be a collection of polynomial-size programs such that every $P \in \mathcal{P}_n$ is a program $P : \{0,1\}^n \to \{0,1\}$. The collection $\mathcal{P}$ is called **evasive** if there exists a negligible function ϵ such that for every $n \in \mathbb{N}$ and for every $y \in \{0,1\}^n$:

$$\Pr_{P \leftarrow \mathcal{P}_n}[P(y) = 1] \leq \epsilon(n).$$

Example ($x, y \in \{0,1\}^n$)

Hamming distance: $d_H(x,y) = \#\{i \mid x_i \neq y_i\}$
Hamming ball of radius r: $B_{H,r}(x) = \{y \mid d_H(x,y) \leq r\}$
When is **Hamming ball membership** evasive?

Preliminaries

Lemma

Let $\lambda \in \mathbb{N}$ be a security parameter and let $r, n \in \mathbb{N}$ such that

$$r \leq \frac{n}{2} - \sqrt{\log(2)n\lambda}.$$

Fix a point $x \in \{0,1\}^n$. Then the following probability is negligible

$$\Pr_{y \leftarrow \{0,1\}^n}[y \in B_{H,r}(x)] \leq \frac{1}{2^\lambda}.$$

$\Rightarrow$ Hamming ball membership of uniform $y \leftarrow \{0,1\}^n$ is **evasive** for $r \leq \frac{n}{2} - \sqrt{\log(2)n\lambda}$.

Preliminaries

Given **secret** $x \in \{0,1\}^n$ and random $h \in \{0,1\}^k$, and a random linear error correction code G. A **secure sketch** is then given by

$$s = x \oplus Gh.$$

- Given $y \in B_{H,r}(x)$:

$$s' = y \oplus s = y \oplus x \oplus Gh = e \oplus Gh$$

(where $e = y \oplus x$)

- Decoding s' reveals h (and also x).
- **Pitfalls:**
  - (G, s) can be quite large.
  - **Hard** to control r, n, k (recall $r \leq n/2 - \sqrt{\log(2)n\lambda}$).
  - Unclear **decoding**/reusability.

A Natural Problem

- **(Modular) Subset Sum Problem**: Fix a modulus $q \in \mathbb{N}$, and a set $S \subset \mathbb{Z}/q\mathbb{Z}$. Given $x = \sum_{a \in A} a \pmod q$ for a *random* subset $A \subset S$, find A.

These problems are intimately related to the **Short Integer Solution** (SIS) problem.

Short Integer Solution

Fix dimensions $m, n \in \mathbb{N}$, a modulus q, and a threshold $\beta \in \mathbb{R}$. Given m uniformly random vectors $a_i \in (\mathbb{Z}/q\mathbb{Z})^n$, forming the columns of a matrix $A \in (\mathbb{Z}/q\mathbb{Z})^{n \times m}$, find a nonzero integer vector $z \in \mathbb{Z}^m$ of norm $\|z\| \leq \beta$ such that $Az = 0$.

- The SIS problem is a *lattice* problem.
- It is believed to be post-quantum secure for appropriate parameters.
- Some more buzzwords: **CVP**, **SVP**, **LWE**, **BDD**, ...

Modular Subset Products

Think of a **multiplication version** of the subset sum problem.

Modular Subset Product Problem

- Fix $r < n/2 \in \mathbb{N}$, distinct primes $(p_i)_{i=1,\ldots,n}$, and
- a prime q such that $\prod_{i \in I} p_i < q$ for all subsets $I \subset \{1, \ldots, n\}$ of size r.
- Given an integer $X = \prod_{i=1}^{n} p_i^{x_i} \pmod{q}$ for a secret vector $x \in \{0,1\}^n$,
- the problem is to find x.

Imagine a **decisional version**, the *decisional modular subset product problem*: Distinguish between a modular subset product instance and a uniformly random element of $(\mathbb{Z}/q\mathbb{Z})^*$.
There is a relation to problems studied by Contini et al. [2] for constructing their *very smooth hash*.

Computational Assumptions

Problem (Modular Subset Product Problem, $\mathrm{MSP}_{r,n,D}$)

Let $r, n \in \mathbb{N}$, a distribution D over $\{0,1\}^n$, a secret $x \leftarrow D$, $(p_i)_{i=1,\ldots,n}$ a sequence of small primes, a prime $q \sim \prod_{r \text{ largest } p_i} p_i$. Given

- $(p_i)_{i=1,\ldots,n}$,
- q, *and*
- $X = \prod_{i=1}^{n} p_i^{x_i} \mod q$,

the problem is to find x.

Problem (Decisional MSP, $\text{D-MSP}_{r,n,D}$)

This problem is to distinguish the distribution of $MSP_{r,n,D}$ samples from uniformly random over $\mathbb{Z}_q$.

Computational Assumptions: Reduction

Conjecture

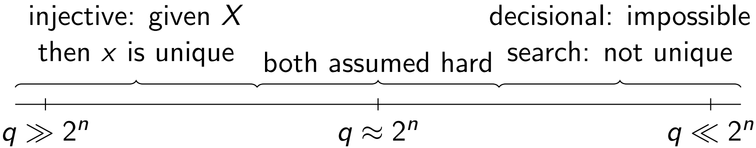
Let $r, n, (p_i)_{i=1,\dots,n}, q$ be as before, with the extra condition that $q \le 2^n$. Let D be the uniform distribution on $\{0,1\}^n$. Then the statistical distance of the distribution $\prod_{i=1}^{n} p_i^{x_i} \mod q$ over $x \leftarrow D$ and the uniform distribution on $(\mathbb{Z}/q\mathbb{Z})^*$ is negligible.

Theorem

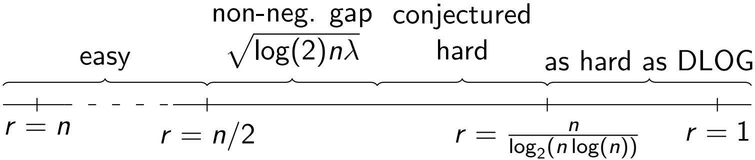
Fix $r, n \in \mathbb{N}$ such that $r < n/2$. Let q be prime such that $q \le 2^n$ and $(p_i)_{i=1,\dots,n}$ be a sequence of distinct primes such that $p_i \in [2, O(n\log(n))]$. Assume above conjecture holds and suppose $MSP_{r,n,D}$ can be solved with probability 1 in time T. Then there is an algorithm to solve the DLP in $(\mathbb{Z}/q\mathbb{Z})^*$ with expected time $\tilde{O}(nT)$.

Computational Assumptions: Summary

- Search vs Decision



- Hardness ($r \leq n/2 - \sqrt{\log(2)n\lambda}$)



Fuzzy Matching

Definition (Hamming Ball Membership)

Let $r < n/2 \in \mathbb{N}$. Given $(p_i)_{i=1,\dots,n}, q$ as in $\mathrm{MSP}_{r,n,D}$, output $X = \prod_{i=1}^{n} p_i^{x_i} \mod q$ as an **encoding** of a secret $x \in \{0,1\}^n$.

- Given $y \in B_{H,r}(x)$, compute $Y = \prod_{i=1}^{n} p_i^{y_i} \mod q$, then:

$$E = XY^{-1} \mod q = \prod_{i=1}^{n} p_i^{x_i - y_i} \mod q = \prod_{i=1}^{n} p_i^{e_i} \mod q.$$

- Recover $e \in \{-1,0,1\}^n$ from E by expanding E/q into a **continued fraction** and **factoring**.
- **Decoding fails** if $\sum_{i=1}^{n} |e_i| > r$ as then $\prod_{i=1}^{n} p_i^{|e_i|} > q$.

Example

$\exists s \in \mathbb{Z} : ED = N + sq \Rightarrow s/D$ is a convergent of E/q

$$q = 751, \qquad (p_i) = (2, 3, 5, 7, 11, 13, 17, 19)$$
$$x = (1, 0, 0, 1, 0, 1, 1, 0), \qquad X = 90$$
$$y = (0, 1, 1, 1, 1, 1, 1, 0), \qquad Y = 666$$

Continued fraction expansion of $XY^{-1}/q = 264/751$ yields convergents h_i/k_i; factor $XY^{-1}k_i \mod q$ and k_i:

- $i = 0$: $1/2 \Rightarrow 223, 2$ ϟ
- $i = 1$: $1/3 \Rightarrow 41, 3$ ϟ
- $i = 2$: $6/17 \Rightarrow 2 * 3^2, 17$ ϟ
- $i = 3$: $13/37 \Rightarrow 5, 37$ ϟ
- $i = 4$: $45/128 \Rightarrow 3, 2^7$ ϟ
- $i = 5$: $58/165 \Rightarrow 2, 3 * 5 * 11$ ✓ $\Rightarrow e = (1, 1, 1, 0, 1, 0, 0, 0)$

Security

Theorem

Let $(n(\lambda), r(\lambda))$ be a sequence of parameters for $\lambda \in \mathbb{N}$. Let $D = \{D_\lambda\}_{\lambda \in \mathbb{N}}$ be an ensemble of Hamming distance evasive distributions with auxiliary information. *Suppose that* entropic *D-MSP$_{r,n,D}$ is hard. Then the Hamming distance obfuscator $\mathcal{O}_H$ is a distributional VBB obfuscator for D in the random oracle model.*

(Note that the distribution of secrets and the computational problem in the assumptions above are **entropic** to make the VBB proof work.)

Theorem

Let $(n(\lambda), r(\lambda))$ be be a sequence of parameters for $\lambda \in \mathbb{N}$. Let $D = \{D_\lambda\}_{\lambda \in \mathbb{N}}$ be an ensemble of Hamming distance evasive distributions. Suppose that $MSP_{r,n,D}$ is hard. Then the Hamming distance obfuscator $\mathcal{O}_H$ is input hiding.

Conjunctions

- **Conjunctions** on Boolean variables $(b_i)_{i=1,\dots,n}$: $\bigwedge_{i=1}^{n}(\neg)b_i$
- Equivalent to **pattern matching with wildcards**: vector $x \in \{0,1,\star\}^n$ where $\star$ symbolises a *wildcard*.
- To encode pattern x, use the map $\sigma : \{0,1,\star\} \to \{-1,0,1\}$ that acts as $0 \mapsto -1, 1 \mapsto 1, \star \mapsto 0$. Publish then $X = \prod_{i=1}^{n} p_i^{\sigma(x_i)} \mod q$.
- Same parameters and scheme as for Hamming distance if we choose $r = |\{i \,|\, x_i = \star\}|$.
- We prescribe the possible error positions.

The Relation Lattice

- Consider parameters as before and the following group morphism:

$$\phi : \mathbb{Z}^n \to (\mathbb{Z}/q\mathbb{Z})^*,$$
$$(x_1, \ldots, x_n) \mapsto \prod_{i=1}^{n} p_i^{x_i} \pmod q.$$

- The kernel of ϕ defines the **relation lattice**

$$\Lambda = \left\{ x \in \mathbb{Z}^n \,\middle|\, \prod_{i=1}^{n} p_i^{x_i} = 1 \pmod q \right\}.$$

- This lattice has been studied by Ducas et al. [4] for constructing BDD lattices.
- Similar ideas have been considered by Brier et al. [1] to construct a number theoretic error correction code.

Post-Quantum Hardness

- Consider adversary with **quantum computer** for computing discrete logarithms.
- Given encoding $((p_i)_{i=1,\ldots,n}, q, X)$ of a secret $x \in \{0,1\}^n$.
- Transform it into a modular subset sum instance (by taking logs wrt. to some g)

$$\log_g(X) = \sum_{i=1}^{n} x_i \log_g(p_i) \pmod{q-1}.$$

- Modular subset sum problem may be classified by **density** $d = n/\log_2(q)$.

Know **polynomial time algorithms** for low-density subset sum instances where $d < 0.645$ and $d < 0.941$, respectively [3, 5] given access to a lattice oracle.

Post-Quantum Hardness

- In our case, we can give an estimate for when we **expect post-quantum security**.
- By the prime number theorem, we have $q \sim (n \log n)^r$, i.e. $d \sim n/(r \log_2(n \log n))$.
- To ensure density of $d > 1$ we require

$$r < \frac{n}{\log_2(n \log n)} = r_{\mathrm{PQ}}(n).$$

Hence we conjecture post-quantum hardness of the modular subset product problem when $r < r_{\mathrm{PQ}}(n)$, and potentially even for slightly larger values for r.

Thank you!

References

[1] Eric Brier, Jean-Sébastien Coron, Rémi Géraud, Diana Maimuţ, and David Naccache. A number-theoretic error-correcting code. In *International Conference for Information Technology and Communications*, pages 25–35. Springer, 2015.

[2] Scott Contini, Arjen K. Lenstra, and Ron Steinfeld. Vsh, an efficient and provable collision-resistant hash function. In *EUROCRYPT 2006*, pages 165–182. Springer, 2006.

[3] Matthijs J Coster, Antoine Joux, Brian A LaMacchia, Andrew M Odlyzko, Claus-Peter Schnorr, and Jacques Stern. Improved low-density subset sum algorithms. *Computational Complexity*, 2(2):111–128, 1992.

[4] Léo Ducas and Cécile Pierrot. Polynomial time bounded distance decoding near minkowski's bound in discrete logarithm lattices. *Designs, Codes and Cryptography*, 87(8):1737–1748, 2019.

[5] Jeffrey C Lagarias and Andrew M Odlyzko. Solving low-density subset sum problems. *Journal of the ACM*, 32(1):229–246, 1985.