Obfuscated Fuzzy Hamming Distance and Conjunctions from Subset Product Problems

Steven D. Galbraith and Lukas Zobernig

TCC 2019

Introduction

Outline

- Motivation
- Preliminaries
- Secure Sketches from Error Correction Codes
- Computational Assumptions
- Our Scheme
- Obfuscation Notions
- Security
- Conjunctions
- Conclusion

Motivation

Can we:

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

Example

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

Keywords:

- Secure sketch
- Fuzzy extractor

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 $y \in \{0,1\}^n$:

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

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

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

- 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

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*
- $X = \prod_{i=1}^{n} p_i^{x_i} \mod q$,

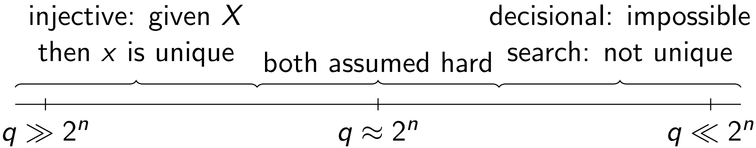
the problem is to find x.

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

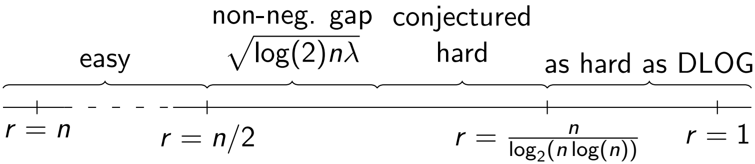
This problem is to distinguish the distribution of $MSP_{r,n,D}$ samples from uniformly random over $\mathbb{Z}_q$.

Computational Assumptions

- Search vs Decision



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



Our Scheme

Definition (Fuzzy Hamming Distance)

Let $r < n/2 \in \mathbb{N}$. Given $(p_i)_{i=1,\ldots,n}, q$ 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:

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

Obfuscation Notions

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:

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

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:

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

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. Suppose that D-MSP$_{r,n,D}$ is hard and that $\mathcal{O}_{PT}$ is a dependent auxiliary input distributional VBB point function obfuscator. Then the Hamming distance obfuscator $\mathcal{O}_H$ is a distributional VBB obfuscator.

Theorem

Let $(n(\lambda), r(\lambda))$ be parameters satisfying $r > r_f(n)$. 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 $r = |\{i \mid x_i = \star\}|$.
- We prescribe the possible error positions.

Conclusion

- New computational assumption: Modular Subset Product Problem
- Obtain fuzzy Hamming distance obfuscator for full parameter range $r \leq n/2 - \sqrt{\log(2)n\lambda}$
- Obtain conjunction obfuscator, same parameter range
- Separate security notions: VBB and input hiding obfuscation

Thank you for your attention!