Correcting Public and Private Errors

Lukas Zobernig

SRC 2019

Introduction

Working on **Cryptography** $\Rightarrow$ interested in **Error Correction**

Outline

- Passwords
- Research Question(s)
- Hamming Distance
- Correcting Public Errors
- Correcting Private Errors
- Applications
- Better Error Correction Codes?

Passwords

- Store them in **cleartext** $\Rightarrow$ insecure
- Solution: store them **hashed**

Definition (Cryptographic Hash Function, One-Way Function)

Map $H : \{0,1\}^n \rightarrow \{0,1\}^m$ (typically $m \ll n$) which is **pre-image resistant**: given $h \in \{0,1\}^m$ it is difficult to find $x \in \{0,1\}^n$ such that $H(x) = h$.

Research Question(s)

Can we:

- securely **encode** and **match** fingerprints ...
- ... as well as other biometric features (iris scans, DNA, etc.)?
- **hash** passwords in a way that allows for (small) errors?

Example

$$
\begin{aligned}
x &= (P, a, S, w, o, r, d, 1, 2, 3), \\
y &= (P, A, s, w, o, r, d, 0, 2, 3)
\end{aligned}
$$

Hamming Distance

Without loss of generality, work over the **binary** set $\mathbb{F}_2 = \{0, 1\}$.

Definition (Hamming Distance)

Let $n \in \mathbb{N}$ and $x, y \in \{0, 1\}^n$ be two binary vectors. The **Hamming distance** between x and y is then given by

$$d_H(x, y) = \#\{i \mid x_i \neq y_i\}.$$

Example

$$\begin{aligned} x &= (1, {\color{red}0}, {\color{red}1}, 1, 1, 0, 0, {\color{red}1}), \\ y &= (1, {\color{red}1}, {\color{red}0}, 1, 1, 0, 0, {\color{red}0}) \\ \Rightarrow d_H(x, y) &= 3 \end{aligned}$$

Hamming Distance

The **Hamming ball** $B_{H,r}(x)$ of radius r around a vector x:

$$B_{H,r}(x) = \{y \mid d_H(x, y) \leq r\}$$

Example

$$B_{H,1}(000) = \{000, 100, 010, 001\}$$

$$B_{H,2}(000) = \{000, 100, 010, 001, 110, 101, 011\}$$

Correcting Public Errors

Definition ((Linear) Error Correction Code)

A linear $[n, k, d]$ error correction code is

- a **generator matrix** $G \in \mathbb{F}_2^{n \times k}$, together with
- a polynomial time **decoding algorithm**

such that the minimal Hamming distance between **codewords** is d.

Example

A k-length input $x \in \mathbb{F}_2^k$ is mapped to an n-length codeword $c = Gx \in \mathbb{F}_2^n$. Two distinct codewords $c_1 = Gx_1$ and $c_2 = Gx_2$ have Hamming distance at least d: $d_H(c_1, c_2) \geq d$.

Correcting Private Errors

In cryptography, we reduce to **(computationally) hard problems**.

Definition (**M**odular **S**ubset **P**roduct Problem, MSP)

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

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

the problem is to find x.

Example

$$\begin{aligned} (p_i)_{i=1,\dots,6} &= (2,3,5,7,11,13), \\ q &= 389, \\ X &= 2^{1} 3^{1} 5^{0} 7^{0} 11^{1} 13^{1} \mod 389 = 858 \mod 389 \equiv 80 \end{aligned}$$

Correcting Private Errors

Definition (Fuzzy Hamming Distance, [Galbraith, Z., 2019])

Let $r < n/2 \in \mathbb{N}$. Given $(p_i)_{i=1,\ldots,n}, q$ as in (MSP), output X 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 $\prod_{i=1}^n p_i^{e_i} > q$.

Applications

- Securely **encoding** and **matching** fingerprints ...
- ... as well as other biometric features (iris scans, DNA, etc.).
- Password **hashing** that allows for errors.

Example

$$x = (P, a, S, w, o, r, d, 1, 2, 3),$$
$$y = (P, A, s, w, o, r, d, 0, 2, 3)$$

Better Error Correction Codes?

Take a step back and consider (MSP) again,

$$\varphi : \mathbb{Z}^n \to (\mathbb{Z}/q\mathbb{Z})^\times,$$
$$x \mapsto \prod_{i=1}^{n} p_i^{x_i} \mod q.$$

Algebra tells us that $\Lambda = \ker \varphi$ is a **lattice** [Ducas, Pierrot, 2018].

- Given $x = v + e$ for some $v \in \Lambda$ and a bounded (short) error vector $e \in \mathbb{Z}^n$, finding v is another hard problem (**B**ounded **D**istance **D**ecoding, BDD).
- Future research: find Λ such that **encoding and error size are optimal** and **BDD is easy**.

Thank you for your attention!