An Elementary Visit to Crypto Dark Matter for Hash Function

Shen Dong

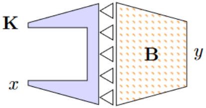
August 29, 2024

Outline

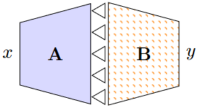
Contents

(2,3)-Construction

$m \geq n, m \geq t$, $A \in \mathbb{Z}_2^{m\times n}, B \in \mathbb{Z}_3^{t\times m}$ fixed and public, **uniformly random/full-rank/toeplitz circulant.**



(2, 3)-wPRF

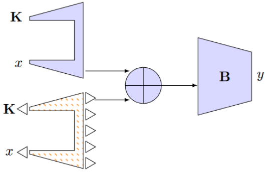


(2, 3)-OWF

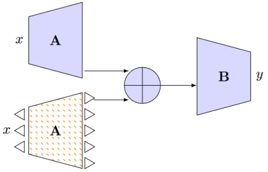
Mod-2/Mod-3 wPRF:
$\mathbb{Z}_2^{m\times n} \times \mathbb{Z}_2^n \to \mathbb{Z}_3^t$ (with a private matrix as key)
$w = Kx \mod 2, y = Bw^* \mod 3$, $*$ means converted to $\mathbb{Z}^3$.
Mod-2/Mod-3 OWF:
$\mathbb{Z}_2^n \to \mathbb{Z}_3^t$ (both matrices $A,\ B$ are fixed and public)
$w = Ax \mod 2, y = Bw^* \mod 3$.

LPN-Construction

$m \geq n, m \geq t,\ A \in \mathbb{Z}_2^{m\times n}, B \in \mathbb{Z}_3^{t\times m}$ fixed and public, **uniformly random/full-rank/toeplitz circulant.**



LPN-wPRF



LPN-PRG

LPN wPRF: $\mathbb{Z}_2^{m\times n} \times \mathbb{Z}_2^n \to \mathbb{Z}_3^t$

$u = Kx \mod 2,\ v = (K^* x^* \mod 3) \mod 2,$

$w = u \oplus v,\ y = Bw \mod 2$

LPN PRG: $\mathbb{Z}_2^n \to \mathbb{Z}_3^t$

$u = Ax \mod 2,\ v = (A^* x^* \mod 3) \mod 2,$

$w = u \oplus v,\ y = Bw \mod 2$

Constant-noise LPN

With constant noise rate $\frac{1}{3}$

$$w = [(\mathbf{A}x \bmod 2) + (\mathbf{A}x \bmod 3) \bmod 2] \bmod 2$$

- Here both matrix structured and noise structured (somehow correlated, though not found yet).
- When $t = 1$, the structure is exposed (entailing attack CCKK21. But with B, one more **compression**.
- $t = m$ trivial for Gaussian elimination, so $1 << t << m$.
- With one key for wPRF, number of samples limited to 2^40 for security.

Contents

Cryptographic Hash Function

Finished and Todo for a hash function:

- hash value looks (pseudo)-random (PRG & PRF)
- hard to invert and get the preimage for a hash (OWF)
- hard to find a second preimage for a hash
- collision-resistant

How to avoid collision? Toeplitz hash may be collision-resistant? but to what extent? what if we add module and B (acting like a compression)?

Final Parameters

Yes they can!

Public matrices: toeplitz, private key matrix: circulant (prefered full-rank, but uniformly random works).

Lemma

Let $n = 2^{n'}$ for a positive integer n' and let $\mathbf{K} \in \mathbb{Z}_2^{n \times n}$ be a circulant matrix selected uniformly at random. Then, for any $a \in \{0, \dots n\}, \Pr[\mathrm{rank}(\mathbf{K}) \leq a] = 2^{-n+a}$.

Construction	Parameters (n, m, t)	Comment
(2, 3)-OWF	$(s, 3.13s, s/\log 3)$ $(s, 3.53s, s/\log 3)$	aggressive conservative
(2, 3)-wPRF	$(2s, 2s, s/\log 3)$ $(2.5s, 2.5s, s/\log 3)$	aggressive conservative
LPN-PRG	$(s, 3s, 2s)$	
LPN-wPRF	$(2s, 2s, s)$	

Onewayness: Subset Sum

For codeword w parity-check matrix $\mathbf{P}$ shaped $(m-n) \times m$ s.t.:

$$\mathbf{A}x = w \text{ if and only if } \mathbf{P}w = \mathbf{0}$$

(2,3)-OWF: $\mathbf{P}w = \mathbf{0}$ (over $\mathbb{Z}_2$), $\mathbf{B}w = \hat{y}$ (over $\mathbb{Z}_3$).
Find set $J \subseteq [m]$ of unit vectors:

$$\left(\sum_{j \in J} \mathbf{P} e_j \bmod 2, \sum_{j \in J} \mathbf{B} e_j \bmod 3 \right) = (\mathbf{0}, \hat{y})$$

Then $\mathbf{A}x = \sum_{j \in J} e_j \bmod 2$, and solve x
For LPN-construction also: exhausive searching w then reduced to subset sum.

Inapproximability by Low-Degree Polynomials, Lemma 4.2 in BIP+18

Lemma

For $n > 0$ and $d < n/2$, let $B(n, d) = \frac{1}{2^n} \cdot \sum_{i=0}^{n/2-d-1} \binom{n}{i}$. Then, for all primes $p \neq q$, the function $\mathrm{map}_p : \{0,1\}^n \to \mathbb{Z}_q$ on n-bit inputs that maps $x \mapsto \sum_{i \in [n]} x_i (\bmod p)$ is $B(n, d)$-far from all degree-d polynomials over $\mathrm{GF}\left(q^\ell\right)$ for all $\ell \geq 1$.

Exhaustive Search

For OWF and wPRF with fewer samples try to mount the secret key x:

- **Basic attack:** guess $m-t$ bits of $w = Ax$ and solve the other t bits as variable with equations $\hat{y} = Bw$ over $\mathbb{Z}_3$. Complexity 2^{m-t}.
- **Bipartite speedup:** when w and w' doesn't have 1 in common, then $w + w' \mod 2 = w + w' \mod 3$

$$\begin{aligned} \mathbf{B}\left(w + w' \bmod 2\right) \bmod 3 &= \\ \mathbf{B}\left(w + w' \bmod 3\right) \bmod 3 &= \\ (\mathbf{B}w \bmod 3) + (\mathbf{B}w' \bmod 3) \bmod 3 & \end{aligned}$$

Partition indices: I_1 and $I_2 = [m] \backslash I_1$, $|I_1| = |I_2| = m/2$.

1. $i \in \left\{0, 1, \ldots 2^{m/2} - 1\right\}$, w_i on the indices of I_1 is i, and is 0 on the indices of I_2. $\forall i$, evaluate $\mathbf{B}w_i \bmod 3 = y_i$ and form a table $\mathcal{T}$;
2. For $j \in \left\{0, 1, \ldots 2^{m/2} - 1\right\}$, w'_j on I_2 is j 0 on I_1. $\forall j$, evaluate $\mathbf{B}w'_j \bmod 3 = y'_j$ and search $\mathcal{T}$ for the value $\hat{y} - y'_j \bmod 3$. Return matched $w = w_i + w'_j \bmod 2$.

Complexity $\mathbf{B}w = \hat{y}$ is $2^{m - \log 3 \cdot t}$, constant speedup.

Bias from Linearization

If we find $v \in \mathbb{Z}_3^m$ and $u \in \mathbb{Z}_3^t$ such that $u\mathbf{B} = v$, $|v| = \ell$
$y = \mathbf{B}w \bmod 3 \Longrightarrow uy \bmod 3 = vw \bmod 3$
Attacker obtains the value of a linear combination mod3 of ℓ entries of $w \in \{0,1\}^m$.
Assuming that w is uniformly distributed in $\mathbb{Z}_2^m$ the bias should be (by induction on ℓ or by analysis of sums of binomial coefficients):

$$\Pr\left[\sum_{i \in I} v_i w_i \bmod 3 = a\right] \in \left\{\frac{1}{3} \pm \frac{1}{2^\ell}, \frac{1}{3} \pm \frac{2}{2^\ell}\right\}$$

where I is the non-zero indices in v and for any $a \in \{0,1,2\}$.
Thus, the bias of $vw \bmod 3$ is bounded by $\frac{2}{2^\ell}$.

Linear Test Frame BCGI+22

Minimum distance of linear code:

$$d(A) = \sharp\text{the minimum weight of a vector in } A\text{'s rowspan.}$$

Minimum distance decides security of LPN in linear attacks, but NP-hard to compute.

Lemma

The subspace spanned by the rows of $\mathbf{B}$ *(A random matrix, or a random Toeplitz matrix forms a linear code) contains a vector of Hamming weight at most* ℓ *with probability at most* $2 \cdot 2^{m(H(\ell/m)-\log 3)+\ell+\log 3\cdot t}$.

Conditional Bias

Bias may increase if information about the variables w_i is known. e.g. parity $\sum_{i \in I} w_i \bmod 2$.
e.g. if $\mathbf{K}$ circulant, x even weight, then $\sum_{i \in [m]} w_i \bmod 2 = 0$
The conditional bias:

$$\left| \Pr \left[\sum_{i \in I} w_i \bmod 3 = 0 \mid \sum_{i \in I} w_i \bmod 2 = 0 \right] - 1/3 \right|$$

can be as large as about $2^{-0.21\ell}$.
therefore they choose $l \approx s/2$.

Circulant K

- Circulant matrices preserves symmetry.
- $x \in \mathbb{Z}_2^n$ 2-symmetric (two haves equal) $\Longrightarrow w = Kx$ also.
- In 2^r samples, probability to have 2-symmetric $2^{-\frac{n}{2}+r}$, parameters must chosen make this negl, or an attacker may try to guess.
- Also 4-symmetric and more, linear combination of samples symmetric, approximately 2-symmetric...

Contents

Toeplitz Matrix

$$T = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-n+1} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{pmatrix}$$

Embedded into a circulant matrix of size $2n$:

$$A = \left(\begin{array}{ccccc|ccccc} a_0 & a_{-1} & \cdots & \cdots & a_{-n+1} & 0 & a_{n-1} & \cdots & a_2 & a_1 \\ a_1 & a_0 & \ddots & & \vdots & a_{-n+1} & \ddots & \ddots & & a_2 \\ a_2 & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & & \ddots & \vdots \\ \vdots & & \ddots & a_0 & a_{-1} & a_{-2} & & \ddots & \ddots & a_{n-1} \\ a_{n-1} & \cdots & \cdots & a_1 & a_0 & a_{-1} & a_{-2} & \cdots & a_{-n+1} & 0 \\ \hline 0 & a_{n-1} & \cdots & \cdots & a_1 & a_0 & a_{-1} & \cdots & \cdots & a_{-n+1} \\ a_{-n+1} & \ddots & \ddots & & a_2 & a_1 & a_0 & \ddots & & \vdots \\ \vdots & \ddots & & \ddots & \vdots & a_2 & \ddots & \ddots & \ddots & \vdots \\ a_{-2} & & \ddots & \ddots & a_{n-1} & \vdots & & \ddots & a_0 & a_{-1} \\ a_{-1} & a_{-2} & \cdots & \cdots & 0 & a_{n-1} & \cdots & \cdots & a_1 & a_0 \end{array}\right).$$

FFT for Toeplitz Matrix

$$
\begin{aligned}
Tv &= \begin{pmatrix} I_n & 0_n \end{pmatrix} A \begin{pmatrix} v \\ 0_n \end{pmatrix} = \begin{pmatrix} I_n & 0_n \end{pmatrix} \begin{pmatrix} T & T' \\ T' & T \end{pmatrix} \begin{pmatrix} v \\ 0_n \end{pmatrix} \\
&= \begin{pmatrix} I_n & 0_n \end{pmatrix} \begin{pmatrix} Tv \\ T'v \end{pmatrix} = Tv
\end{aligned}
$$