Watermarking II

A different correlated sampling algorithm

+ will admit robustness
− requires more entropy from the distribution

problem with the previous method:
need a contiguous errorless sequence.

new method: - start from a pseudorandom string

$r = G(s)$, $s \sim U_m$, $G: \{0,1\}^m \to \{0,1\}^n$ ($m < n$)

- for each bit i, let p_i be the sampling probability

(Recall: $p_i = p(x_i = 1 \mid x_1, \ldots, x_{i-1})$)

define

$p_i \geq 1/2$:
$p_{i,\phi} = 1$
$p_{i,0} = 2p_i - 1 \ (\geq 0)$

$p_i < 1/2$:
$p_{i,0} = 0$
$p_{i,1} = 2p_i \ (< 1)$

- Sample $x_i \sim \text{Binomial}(p_i, r_i)$ and output it.

Claim: if r is random, then $x_i \sim \text{Binomial}(p_i)$

assume $p_i \geq 1/2$. $\Pr[x_i = 1] = \frac{1}{2}(p_{i,1} + p_{i,0}) = p_i$

same for $p_i < 1/2$. $\Pr[x_i = 1] = \frac{1}{2} \cdot 2p_i = p_i$

Claim: $\Pr[x_i = r_i] = \begin{cases} 3/2 - p_i & p_i \geq 1/2 \\ 1/2 + p_i & p_i < 1/2 \end{cases}$

$p_i = 1/2 + \varepsilon$ (i.e. $= 1-\varepsilon$) $\qquad p_i = 1/2 - \varepsilon$ (i.e. $\leq 1-\varepsilon$)

assume $p_i \geq 1/2$. $\Pr[x_i = r_i] = \frac{1}{2}(1 + 1 - (2p_i - 1))$

$= \frac{3}{2} - p_i$

$p_i < 1/2$. $\Pr[x_i = r_i] = \frac{1}{2}(2p_i + 1)$

$= \frac{1}{2} + p_i$

// if $p_i = 1/2$. $\Pr[x_i = r_i] = 1$ – perfect correlation!

~~Claim~~ ~~$\mathbb{E}[\Delta(x,r)]$~~

if $p_i = 1/2 \pm \varepsilon_i$

$\Pr[x_i = r_i] = 1 - \varepsilon_i$

Claim: $\mathbb{E}[\Delta(x,r)] = \sum_{i=1}^{n} \varepsilon_i$

~~Claim: If~~

Claim (exercise)

if $H_\infty(\pi) \geq cn$, then $\sum_{i=1}^{n} \varepsilon_i \leq c'n$.

(for $c > 0$) (for $c' < 1/2$)

(Use concavity of the binary entropy fn. for a quantitative bound.)

Detection Is $r = G(s) + \text{error}$?

pseudorandom encryption ? not noise-tolerant

error correcting code ? not pseudorandom

} Want: pseudorandom error-correcting codes

Def: (PRC) a setup "Setup"
a Generator "G" & a detector "Detect" st

$$\begin{cases} \text{Setup} \to (pk, sk) \\ G_{pk}(s) \to r \quad (m \to n) \\ \text{Detect}_{sk}(r) \to 0/1 \end{cases}$$

Completeness:

$$\Pr\left[\text{Detect}_{sk}\left(G_{pk}(s) + \underbrace{\text{error}}_{\text{bounded}}\right) = 1\right] \geq 1 - 2^{-\lambda}$$

more generally

$$\Pr\left[\text{Detect}_{sk}\left(\mathcal{E}_{ch}\left(\mathcal{E}_{embed}\left(G_{pk}(s)\right)\right)\right) = 1\right] \geq 1 - 2^{-\lambda}.$$

Soundness:

for any fixed string x,

$$\text{Detect}_{sk}(x) = 0 \quad \text{w.p. } 1 - 2^{-\lambda}$$

pseudorandomness:

$$G_{pk}(s) \approx U_n.$$

(PRC $\Rightarrow$ watermark. robust to bounded # substitutions)

Can we construct PRCs?

Learning Parity with Noise

- Given many random "equations" in an n-dimensional secret vector $s \in \mathbb{F}_2^n$, can you recover s?

$$\{a_i, \langle a_i, s\rangle\} \xrightarrow{?} s$$

Easy: e.g. Gaussian Elimination

- How about noisy equations?

$$(a_i, \langle a_i, s\rangle + e_i) \xrightarrow{?} s$$

$e_i \sim \text{Binomial}(p)$

Learning Parity with noise. Believed to be very hard.

- In n dim, best alg: $2^{O(n/\log n)}$ time [Blum-Kalai-Wasserman 2003]
- Alternatively, largest dimension for which we have a poly-time (poly(n)) algorithm $\leq \log n \cdot \log\log n$.

($\because 2^{(\log n \log\log n)/\log(\log n \cdot \log\log n)} \simeq 2^{\log n} = \text{poly}(n)$)

We will use ↓

- dim $\log^2 n$: best alg $\sim 2^{\log^2 n/\log\log n} = \text{quasi-poly}(n)$.

- Matrix view:

$$\left(\begin{array}{c}\#\text{eqns} \\ m\end{array}\boxed{A}\ \overset{n}{},\ \boxed{A}\boxed{s}+\boxed{e}\right) \approx_c \left(\boxed{A},\ \boxed{u}\right)$$

A, $As+e$. (uniform random)

PRG from LPN: first try

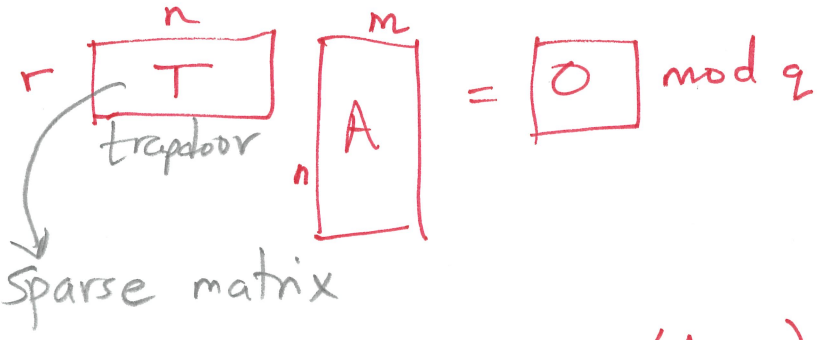
Encode(A): output $As+e$, where $s \leftarrow \mathbb{F}_2^n$, $e \sim \text{Binomial}(p)$ (for some p)

arbitrarily small constant

~~pseudorandom~~ ✓

how to detect?!

Idea: sample A together with a trapdoor — will help us detect



~~t~~-sparse matrix

- Given T, compute $T(As+e) = TAs + T \cdot e$
  $= T \cdot e \pmod 2$

each row of T is ~~t~~-sparse

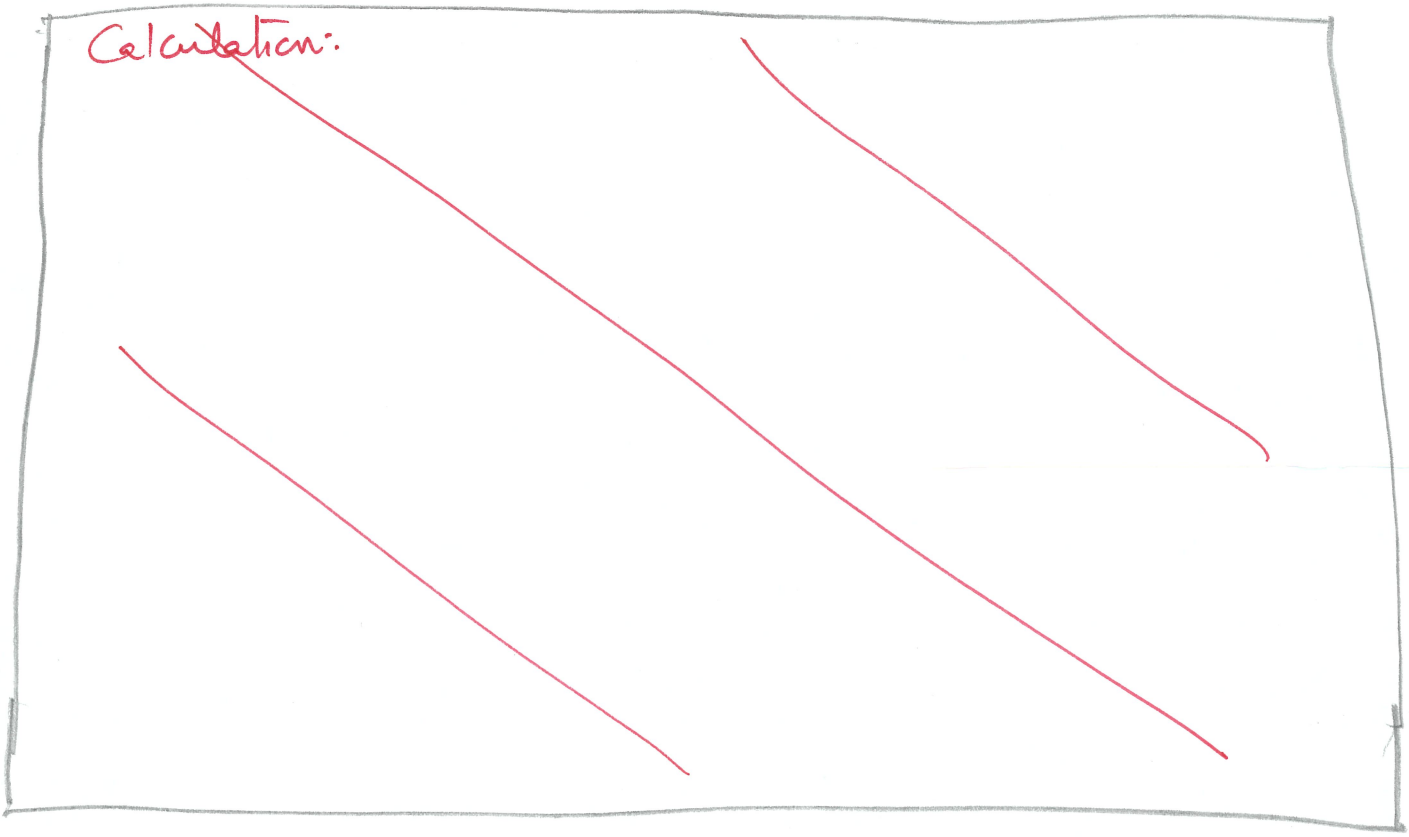
$\Rightarrow \langle t, e \rangle$ is ~~Binomial$\left(\frac{1}{2} - \frac{1}{2}p^t\right)$~~ $\text{Binomial}\left(\frac{1}{2} - \left(\frac{1}{2}-p\right)^t\right)$

$\Rightarrow$ for $p = O(1)$, need ~~t~~ $= O(\log n)$.

for a random vector u, $\langle t, u \rangle$ is unbiased.

- Detector: count the number (out of r) tests that result in 0

Calculation:



Second try:

How to sample A & T together:

1) pick T to be a random t-sparse matrix
2) pick each column of A to be a random solution to $T \cdot a = 0 \mod 2$

- If $n = O(\log^2 m)$, A is stat. close to random
- For larger n, conjectured to be pseudorandom.

Given this, here is a PRC construction

$pk = A \in \mathbb{F}_2^{n \times m}$, $z \in \mathbb{F}_2^m$ $(m \sim \log^2 n)$

$sk = T \in \mathbb{F}_2^{r \times m}$ $r \sim n^\alpha$ $(0 < \alpha < 1)$

$G_1(s)$: output $As \oplus e$ $\quad e \sim Bin(\rho)$

$(As \oplus e \oplus z)$

Detect(x): if $\gg \frac{1}{2} + \frac{\lambda}{\sqrt{r}}$ fraction of the r entries of $T(x \oplus z)$ are 0, output PRC else Random.

pseudorandomness ✓ by subexp. LPN

✓ soundness (with the ~~changes~~ addition of z):

For any fixed y, $T(y \oplus z)$ will have $\approx \frac{1}{2} r$ ones

$\Pr[\text{has} \gg \frac{r}{2} + \lambda\sqrt{r} \text{ zeroes}] \le 2^{-\lambda}$

Completeness + robustness:

$T(As \oplus e \oplus e' \oplus z \oplus z) = T(\underbrace{e \oplus e'}_{\text{worst case Hamming} = \rho + \rho' < \frac{1}{2}})$

$$\#\text{zeroes} \approx \left(\frac{1}{2} - \underbrace{\left(\frac{1}{2} - (\rho + \rho')\right)}_{\text{bias } \varepsilon > 0}^{\tau}\right) r$$

set $\tau = O(\log n)$ such that

$\left(\frac{1}{2} - (\rho + \rho')\right)^\tau \gtrsim n^{-\varepsilon}$

$\#\text{zeroes} \approx \frac{r}{2} + \underbrace{r n^{-\varepsilon}}_{\text{want: } \ge \lambda\sqrt{r}}$ $\Rightarrow$ #zeroes bounded away from $\frac{r}{2}$.

$\Rightarrow r \approx \lambda^2 n^{2\varepsilon}$