Tardos fingerprinting code: Fast and accurate estimation of the probability of being innocent

Options to achieve P_{fp}

- use general threshold when facing a collusion of size c
- assume that scores of innocents follow a Gaussian

$$\tau = \sqrt{2\sigma_{\text{inn}}^2} \cdot \text{erfc}^{-1}\left(\frac{2P_{\text{fp}}}{n}\right)$$

- establish threshold τ for a given couple $(\mathbf{y}, \mathbf{p})$ and a known number of users n

Monte Carlo simulation

Idea: use Monte Carlo simulation to estimate τ s.t.

$$\mathbb{P}(s(\mathbf{x}_{\text{inn}}, \mathbf{y}, \mathbf{p}) > \tau) = n^{-1} P_{\text{fp}}$$

for which it is necessary to create many possible codewords

Number of possible codewords in the order of

$$2^{m \cdot \mathbb{E}_{P \sim f_T}[h(p)]} \gg n$$

$\rightarrow$ many codewords have not used when pirated copy is found

But: $n^{-1} P_{\text{fp}}$ (single decoder) and $\binom{n}{t}^{-1} \cdot P_{\text{fp}}$ (joint decoder) are very small

Rare event simulation to the rescue

- Monte Carlo simulation not the only way to go
- Statisticians have plenty of other estimators: a field called rare event simulation
- "Importance Branching":
  - Generate random codewords,
  - modify randomly and select those with highest scores
- Properties of estimator are known: confidence interval, bias, variance