Exercise Set 5

Raja Kantheti

May 4, 2025

Fingerprinting

Deliverable 1: Theoretical False Positive Rate

We analyze how often a fingerprinting scheme might falsely claim that two unequal binary strings are equal, assuming an adversary chooses $x = 0$ and $y = K$, where K is the product of as many primes as possible between n and n^2, such that $K < 2^n - 1$.

Let P be the set of all primes between n and n^2. Let $P' \subset P$ be the subset of primes that can be multiplied together without exceeding $2^n - 1$.

- Let $N = |P'|$, the number of primes used in K.
- Let $D = |P|$, the number of primes between n and n^2.
- Then, the theoretical false positive rate is:

$$\text{FPR}_{\text{theoretical}} = \frac{N}{D}$$

This represents the probability that $K \bmod p = 0$ for a randomly chosen prime $p \in (n, n^2)$, which would cause a false positive in the protocol.

Empirical vs Theoretical Results (Plot 1)

Figure 1 shows the comparison between the theoretical and empirical false positive rates for n in the range $[6, 300]$.

- The **theoretical curve** is computed based on theabove equation.

- The **empirical curve** is derived from simulation: for each n, we construct K and run the fingerprinting scheme over many random primes, tracking the fraction of times $K \bmod p = 0$.
- Both curves drop rapidly as n increases, confirming the effectiveness of the fingerprinting scheme for larger values of n.
- The empirical values fluctuate slightly around the theoretical curve due to randomness, but they generally align closely.

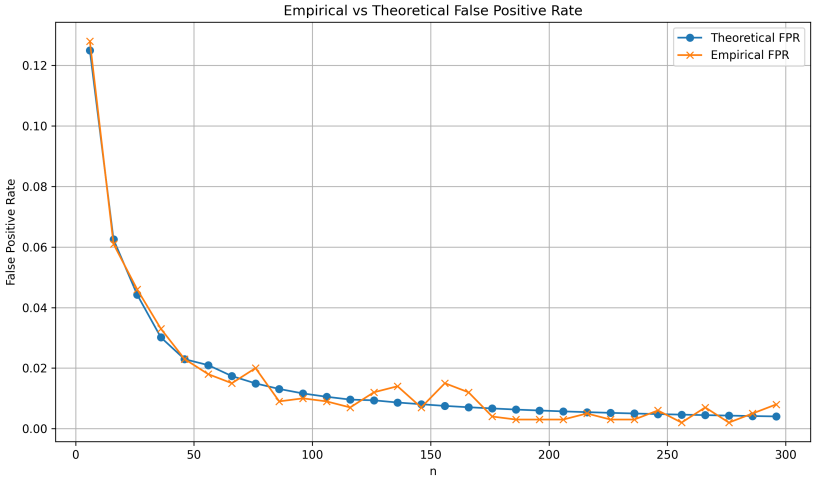

Figure 1: Empirical vs Theoretical False Positive Rate for $n = 6$ to 300

Upper Bound Analysis (Plot 2)

Figure 2 presents the upper bound on the false positive rate, computed for n ranging from 10 to 10^{100} and shown on a log-log scale.

We use the following approximations:

- Overestimate the numerator: $N = \left\lfloor \frac{n \log 2}{\log n} \right\rfloor$
- Underestimate the denominator: $D = \left\lfloor \frac{n^2}{\ln(n^2)} - \frac{n}{\ln n} \right\rfloor$

- Thus, the upper bound is:

$$\text{FPR}_{\text{upper}} = \frac{N}{D}$$

- The log-log plot confirms that the upper bound decreases faster than any polynomial in n.
- Even for extremely large n (up to 10^{100}), the false positive rate remains mathematically negligible.
- This supports the conjecture in the assignment: the false positive rate is generally less than $\frac{1}{n}$.

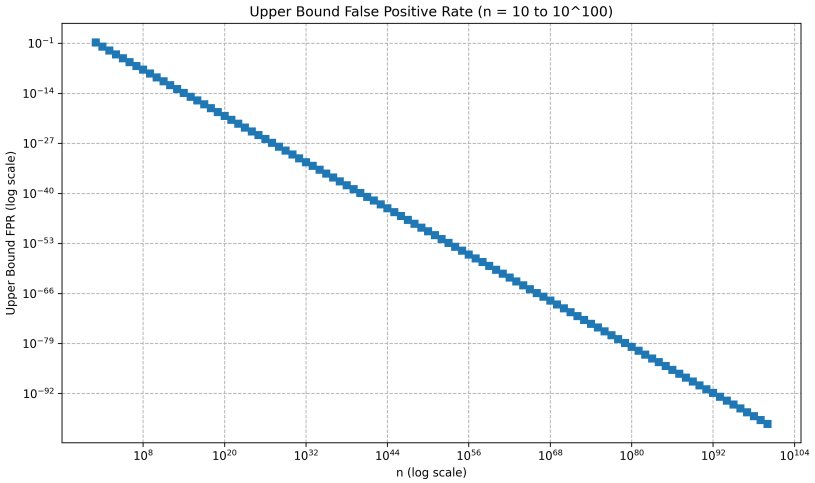

Figure 2: Upper Bound on False Positive Rate (Log-Log Scale) for $n = 10$ to 10^{100}