6.857 Homework Qian Long Joseph W Presbrey Mikhail Rudoy	**Problem Set 1**	# 1-3. Detecting Pad Reuse February 17, 2014

a) Define $R(c_i, c_j)$ to be the length of the longest run of identical bits in identical positions in ciphertexts c_i and c_j.

First we will prove a lemma:

Lemma. *Given that two ciphertexts a and b of length n are properly encrypted, for any asymptotically positive polynomial $q(n)$, we have $\mathbb{P}\Big(R(a,b) > \log_2(n) + \log_2 \ln(n)\Big) < \frac{1}{q(n)}$ for all sufficiently large n.*

Proof. Since a and b are properly encripted, they are chosen uniformly at random from the space of bitstrings of length n. Then the xor of the two ciphertexts, $a \oplus b$, is also chosen uniformly at random from the space of bitstrings of length n. A bit of $a \oplus b$ is 0 if and only if the corresponding bits of a and b are the same. The value $R(a,b)$ is the length of the longest sequence of bits in a such that the corresponding sequence of bits in b is the same. In other words, $R(a,b)$ is the length of the longest run of 0s in $a \oplus b$.

Since $a \oplus b$ is chosen uniformly at random, we can pretend that it was generated by a sequence of n coinflips. Then if we identify 0s in $a \oplus b$ with coins landing heads, we see that the useful fact given in the problem applies. The given fact asserts (in this case) that for any asymptotically positive polynomial $q(n)$ and for all $n \geq N$ for some N,

$$\mathbb{P}\Big(\log_2(n) - \log_2 \ln \ln(n) \leq R(a,b) \leq log_2(n) + \log_2 \ln(n)\Big) \geq 1 - \frac{1}{q(n)}$$

Then

$$\mathbb{P}\Big(\log_2(n) - \log_2 \ln \ln(n) > R(a,b) \text{ or } R(a,b) > log_2(n) + \log_2 \ln(n)\Big) < \frac{1}{q(n)}$$

and finally

$$\mathbb{P}\Big(R(a,b) > log_2(n) + \log_2 \ln(n)\Big) < \frac{1}{q(n)}$$

for all $n \geq N$, as desired. □

Let $l(n)$ be the number of ciphertexts we are given. Let $r(n)$ be any asymptotically positive polynomial.

What we are interested is the value

$$\mathbb{P}\left(\max_{1 \leq i < j \leq l(n)} R(c_i, c_j) \leq \log_2(n) + \log_2 \ln(n)\right)$$

and in particular, we wish to show that it is greater than $1 - \frac{1}{r(n)}$ for $n \geq n_0$ for some n_0.

Alternatively, since

$$\mathbb{P}\left(\max_{1 \leq i < j \leq l(n)} R(c_i, c_j) > \log_2(n) + \log_2 \ln(n)\right) = 1 - \mathbb{P}\left(\max_{1 \leq i < j \leq l(n)} R(c_i, c_j) \leq \log_2(n) + \log_2 \ln(n)\right)$$

we must simply show that $\mathbb{P}\left(\max_{1\le i<j\le l(n)} R(c_i,c_j) > \log_2(n)+\log_2\ln(n)\right) < \frac{1}{r(n)}$ holds for $n \ge n_0$ for some n_0.

In addition, we know that

$$\mathbb{P}\left(\max_{1\le i<j\le l(n)} R(c_i,c_j) > \log_2(n)+\log_2\ln(n)\right) = \mathbb{P}\begin{pmatrix} R(c_1,c_2) > \log_2(n)+\log_2\ln(n) & \text{or} \\ R(c_1,c_3) > \log_2(n)+\log_2\ln(n) & \text{or} \\ R(c_2,c_3) > \log_2(n)+\log_2\ln(n) & \text{or} \\ \dots & \text{or} \\ R(c_{l(n)-1},c_{l(n)}) > \log_2(n)+\log_2\ln(n) & \end{pmatrix}$$

and that

$$\mathbb{P}\begin{pmatrix} R(c_1,c_2) > \log_2(n)+\log_2\ln(n) & \text{or} \\ R(c_1,c_3) > \log_2(n)+\log_2\ln(n) & \text{or} \\ R(c_2,c_3) > \log_2(n)+\log_2\ln(n) & \text{or} \\ \dots & \text{or} \\ R(c_{l(n)-1},c_{l(n)}) > \log_2(n)+\log_2\ln(n) & \end{pmatrix} \le \sum_{i=1}^{l(n)-1}\sum_{j=i+1}^{l(n)} \mathbb{P}\Big(R(c_i,c_j) > \log_2(n)+\log_2\ln(n)\Big)$$

By the lemma, $\mathbb{P}\Big(R(c_i,c_j) > \log_2(n)+\log_2\ln(n)\Big) < \frac{1}{q(n)}$ for every i, j, and asymptotically positive polynomial $q(n)$, and for every $n \ge n_0$ for some n_0 dependent only on q. Let $q(n) = \frac{1}{2}r(n)l(n)(l(n)-1)$, and let N be the associated value of n_0. Then for $n \ge N$ we see that

$$\sum_{i=1}^{l(n)-1}\sum_{j=i+1}^{l(n)} \mathbb{P}\Big(R(c_i,c_j) > \log_2(n)+\log_2\ln(n)\Big) < \frac{l(n)(l(n)-1)}{2} \times \frac{1}{q(n)} = \frac{l(n)(l(n)-1)}{2q(n)}$$

where

$$\frac{l(n)(l(n)-1)}{2q(n)} = \frac{l(n)(l(n)-1)}{2\frac{1}{2}r(n)l(n)(l(n)-1)} = \frac{1}{r(n)}$$

We can conclude that $\mathbb{P}\left(\max_{1\le i<j\le l(n)} R(c_i,c_j) > \log_2(n)+\log_2\ln(n)\right) < \frac{1}{r(n)}$ for $n \ge N$, so

$$\mathbb{P}\left(\max_{1\le i<j\le l(n)} R(c_i,c_j) \le \log_2(n)+\log_2\ln(n)\right) > 1 - \frac{1}{r(n)}$$

for $n \ge N$. Since $r(n)$ can be any asymptotically positive polynomial, we see that the length of the longest repeated bitstring $\left(\max_{1\le i<j\le l(n)} R(c_i,c_j)\right)$ is, with high probability, at most $\log_2(n) + \log_2\ln(n)$, as desired.

b)

c)