Dennis Kuzminer
CSCI-UA 310-001 PS8

1. Show that the number of keys that satisfy this equation

$$\sum_{i=1}^{s}(a_{2i-1}+\lambda_{2i-1})(a_{2i}+\lambda_{2i})-(b_{2i-1}+\lambda_{2i-1})(b_{2i}+\lambda_{2i})=0$$

is at most $m^{t-1} \rightarrow$

$$\sum_{i=1}^{s}(a_{2i-1}a_{2i}+a_{2i-1}\lambda_{2i}+a_{2i}\lambda_{2i-1}+\lambda_{2i}\lambda_{2i-1})-(b_{2i-1}b_{2i}+b_{2i-1}\lambda_{2i}+b_{2i}\lambda_{2i-1}+\lambda_{2i}\lambda_{2i-1}) \rightarrow$$

$$\sum_{i=1}^{s} a_{2i-1}a_{2i}+a_{2i-1}\lambda_{2i}+a_{2i}\lambda_{2i-1}-b_{2i-1}b_{2i}-b_{2i-1}\lambda_{2i}-b_{2i}\lambda_{2i-1} \rightarrow$$

$$\sum_{i=1}^{s} a_{2i-1}a_{2i}-b_{2i-1}b_{2i}+\sum_{i=1}^{s} a_{2i-1}\lambda_{2i}+a_{2i}\lambda_{2i-1}-\sum_{i=1}^{s} b_{2i-1}\lambda_{2i}+b_{2i}\lambda_{2i-1}=0 \rightarrow$$

Rearanging $\rightarrow \sum_{i=1}^{t} a_i\lambda_i-\sum_{i=1}^{t} b_i\lambda_i=0$ (we do not need to consider the constant as there is no random variable associated with it)

$c_i=a_i-b_i \rightarrow \sum_{i=1}^{t} c_i\lambda_i \rightarrow \lambda_1=-c_1^{-1}\sum_{i=2}^{t} c_i\lambda_i \rightarrow$ There are m^{t-1} ways of choosing $\lambda_2,...,\lambda_t$, and each yields one solution. So $N=m^{t-1}$ where $N \leq |\Lambda|/m = m^{t-1}$

This shows $\mathcal{H}$ is universal.

Dennis Kuzminer
CSCI-UA 310-001 PS8

2.

Dennis Kuzminer
CSCI-UA 310-001 PS8

3. Consider the size of the collision set $h_I(a) = h_I(b)$ over the size of the seed set.
   $h_I(a) = h_I(b) \rightarrow$ How many times are $a \in I,\ b \in I$
   Collision set size = ${}_{n-2}C_{k-2} = \frac{(n-2)!}{(k-2)!(n-k)!}$
   Seed set size = ${}_nC_k = \frac{n!}{k!(n-k)!}$
   $\frac{(n-2)!}{(k-2)!(n-k)!} / \frac{n!}{k!(n-k)!} \rightarrow \frac{(n-2)!k!(n-k)!}{n!(k-2)!(n-k)!} \rightarrow \frac{k(k-1)}{n(n-1)} \rightarrow \frac{k(k-1)}{n(n-1)} \leq 1/n \rightarrow \frac{k(k-1)}{(n-1)} \leq 1 \rightarrow$
   $k(k-1) \leq n-1 \rightarrow k(k-1) \leq k^2 \leq n-1 \rightarrow k \leq \sqrt{n-1}$
   This means that because $k \leq \sqrt{n-1},\ \mathcal{H}$ is universal.

Dennis Kuzminer
CSCI-UA 310-001 PS8

4.

a. If two strings will have the same hash this means that they are (most likely) the same. The algorithm creates a "window" for each substring of length t in a and b and hashes the substrings. Given the same seed and hash function, if some $r_i == s_i$ then the hash will be the same.

b. As Karp/Rubin uses ε-universal hashing, meaning ε=(t-1)/m
$|r_i|, |s_i| = n - t + 1$
Therefore, the maximum $Pr[h_\lambda(a) = h_\lambda(b),\ a \neq b] \leq (n - t + 1)(n - t + 1)(\frac{t-1}{m}) \rightarrow (n - t + 1)^2(\frac{t-1}{m})$

c. If we store the value of all r_i in a hashmap/hashtable, we can get an expected constant lookup time. So the only work that would need to be done is to assign the key-value pairs/create a hashtable for all r_i. This means we can simply loop through s_i and lookup each value in the hashtable with expected $O(1)$ time.

Dennis Kuzminer
CSCI-UA 310-001 PS8

5.
6.