

15-451 Assignment 08

Karan Sikka

ksikka@cmu.edu

Collaborated with Dave Cummings and Sandeep Rao.

Recitation: A

November 12, 2014

1: Streaming Medians

(a)

(b)

(c)

(d)

2: Counting Substrings

3: LDIS

(a) Keep a running max length of the duplicate initial string.

For i from 1 to $|s|/2$, compute the hash of $s_{0,i}$ and $s_{i,2*i}$.

If they are equal, update the max length of the duplicate initial string to be i .

Finally return the max length of the duplicate initial strings.

Note that computing the hash at each iteration takes constant time (insert formula here).

Next we will show the probability of a false positive is less than $1/2$.

Let the length of the string be n . Suppose for any fixed locations i and $2i$ the probability of an incorrect match is δ . Then by a union bound over $n/2$ locations Then the probability of any incorrect match is at most $n/2 * \delta$ We want $n/2 * \delta < 1/2$ or equivalently $\delta < 1/n$.

We make an incorrect match when the hash a of one substring is equal to the hash b of another, modulo some random prime q . Formally, this occurs when $q|a - b$. $a - b$ has at most p distinct prime divisors since it's a p -bit number and each prime divisor is at least 2. If $q|a - b$, then q must have been one of the p prime divisors of $a - b$.

We want to choose a prime such that the chance that it's one of the p prime divisors is less than $1/n$. Choose K large enough so that there are pn primes between 2 and K .

If there are $\pi(x)$ primes between one and x , then $\pi(x) \sim \frac{x}{\ln(x)}$. Thus we want to choose K such that $\pi(K) \geq \frac{7}{8} \frac{K}{\ln(K)} > pn$.

Setting $K = 5p \ln(pn)$ achieves this result.

(b)

aaaacc

prefix traversal order:

a => yes
a a => yes
a a a => no
a a a a => no
a a a a c => no
a a a a c c => no

Keep running max length duplicate prefix

1. traverse a, then aa, then aaa...
2. each time, check if the following condition is true:

check if there exists another suffix prefix starting at $2 \cdot \text{len}(\text{prefix})$ (assuming string is 1-indexed)

$\text{prefix}.*$
 $\text{prefix}.*\text{prefix}.*$

3. If the condition is true, update the max-length duplicate prefix to $\text{len}(\text{prefix})$

TODO polish the above.