Streaming Algorithms

Input is a stream

SITUATION?

Router

internet traffic

1) #{of packets}

2) #{of IP-addresses}

Each packet (IP-address destination message)

128-bit address

Sees 240 packets in a day
Streaming Alg:

Input: A stream \( S_1, S_2, \ldots, S_n \)

of each \( S_i \in \{1, \ldots, N\} \)

Goal: Compute \( f(S_1, \ldots, S_n) \)

\[ \text{statistic} \]

Restriction: \( 1) \) Memory available to Alg \( \ll n,N \)

\[ = \text{poly} \left( \log n, \log N \right) \]

\( 2) \) See the input only once in the order \( S_1, \ldots, S_n \)
Streaming

- Read a book with a little sheet to take notes

- Compute word statistics

- Observe traffic
Input: A stream $\cdots$

$\big[\text{read a book with words}\big]$

Sol: 

Alg: Sampling - "keep random sample"
300 voters who vote 0 or 1

estimate # of voters who are 1

- sample k voters $X_i \in \{0, 1\}
- Output $\frac{1}{k} \sum X_i$

With k-samples, the estimate $\frac{1}{k} \sum X_i$

is correct within accuracy $\epsilon$

with probability $\frac{1-\delta}{2}$

1) $k \cdot n$  2) $k$  3) $k \cdot \log n$  4) $k \cdot \sqrt{n}$
Suppose $S$, $\ldots$, $S_n$ is a stream of $\{0,1\}$

let $X_1, \ldots, X_n \in$ be uniformly random samples from $S_1, \ldots, S_n$

$X_q \in S_q$ (uniformly random)

Output: $\frac{1}{k} \sum X_i$

$\mathbb{E} X_i = \{\text{fraction of } 1\text{s in } S_i\}$

estimate answer $= P$

$\Pr \left[ \left| \frac{1}{k} \sum X_i - p \right| > \epsilon \right] \leq 2e^{-2\epsilon^2 k}$
RESERVOIR SAMPLING

Input: Stream $s_1, \ldots, s_n$

Goal: A uniformly random element from stream.

"Correct choice"

reservoir = $s_1$

for $i = 2$ to $\ldots$

choose random number $r \in \{1, \ldots, i\}$

If ($r = 1$) reservoir $\leftarrow s_i$

else Ignore $s_i$

Output reservoir

\( \text{reservoir}[1\ldots t] \in \text{S}[1\ldots t] \)
**Inductive Hypothesis**

At iteration $i$,

$$Pr\left[\text{reservoir} = 8j\right] = \frac{1}{i}$$

for $j = 1 \ldots i$

At iteration $i + 1$

$$Pr\left[\text{reservoir} = 8j\right] = \left(\text{reservoir} = 8j \text{ at iteration } n^t\right) = \frac{1}{i} \cdot \frac{1}{i+1}$$

AND

$$\left(8j \text{ NOT kicked out}\right) = \left(1 - \frac{1}{i+1}\right)$$