Randomness in Computing

CS 537

LECTURE 12

Last time

- Hoeffding Bound
- Applications of Chernoff-Hoeffding Bounds
- Estimating a parameter
- Set Balancing
- Routing on the hypercube

Today

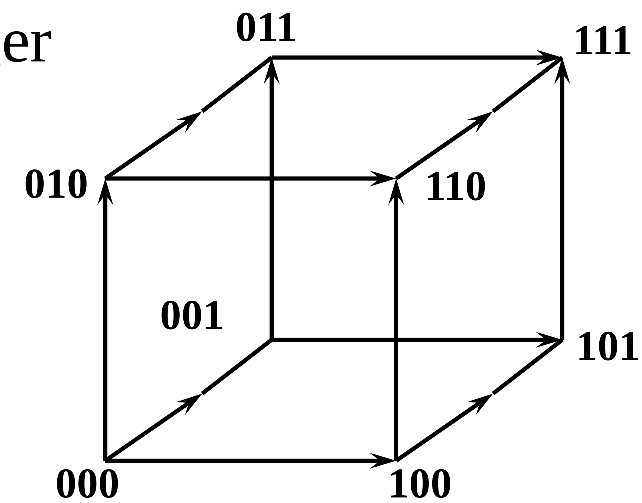
- Analysis of routing on the hypercube
- The Balls-and-Bins model

Application: Routing on Hypercube

An n-dimensional *hypercube* is a directed graph with

- $N = 2^n$ nodes, each indexed by n-bit integer
- containing the directed edge (x, y) iff
  x and y differ in exactly one bit

x	001001
y	011001



How many edges?

- **Routing.** Each node is a routing switch.

  Edges = communication channels

  An edge can carry one packet in one step.

For each depicted edge, there is also the edge in the opposite direction.

- A ***routing algorithm*** specifies a path from s to t for each pair of noes (s, t) and a ***queuing policy*** for ordering packets that are waiting for the same link (e.g., FIFO – First In First Out or FTG – Furthest to Go)

Permutation Routing Problem

- Each node is the source of one packet
- Each node is the destination of one packet

➢ E.g., on a complete graph, it can be solved in one time step.

➢ On sparse graphs?

Hypercube: N nodes $N \log N$ edges.

Bit-Fixing Routing Algorithm for the Hypercube

1. Let x be the current node and y be the destination of a packet
2. Find smallest $i \in [n]$ such that $x_i \neq y_i$
3. Traverse the edge $(x, x_1 \dots x_{i-1}\overline{x_i}x_{i+1} \dots x_n)$

How long does it take to route a packet if there are no delays?

Randomized Routing

Randomized Routing Algorithm (RRA)

0. For each packet going from x to y, pick a node z uniformly and independently at random.
1. Use Bit-Fixing to route the packet from x to z.
2. Use Bit-Fixing to route the packet from z to y.

Theorem

$\log N = \boldsymbol{n}$

For every permutation, RRA takes $O(\log N)$ steps w.p. $1 - O\left(\frac{1}{N}\right)$.

Proof:

Idea 1: Analyze Phase 1. (Phase 2 is ``symmetric'' going backwards; first think about waiting with Phase 2 until all packets are done with Phase 1.)

Idea 2: In an intermediate destination, each bit z_i is 0/1 uniformly & independently

Idea 3: # steps in Phase1 is (# bits to fix) + (waiting time in queues in Phase 1)

$\leq \quad n \quad + \quad \text{delay}$

Homework Lemma

Lemma

Let p_i be the path of some packet i in Phase 1.

Let S be the set of packets (other than i) whose routes pass through at least one edge of p_i. Then the delay of i is at most $|S|$.

Analysis of RRA

Main Lemma

Consider any packet i. It fails to reach its destination in phase 1 within $3n$ steps with probability at most $\frac{1}{N^2}$.

Proof: Let X be the number of packets (other than i) that use at least one edge from the path of i.

We would like to find an upper bound on $\mathbb{E}[X]$.

Analysis of RRA

Main Lemma

Consider any packet i. It fails to reach its destination in phase 1 within $3n$ steps with probability at most $\frac{1}{N^2}$.

Proof: Let X be the number of packets (other than i) that use at least one edge from the path of i.

Find an upper bound on $\mathbb{E}[X]$

- For any edge e, let $Y_e = \#$ routes that pass via e.
- Let $p_i = (e_1, \dots, e_K)$ be the path of packet i. Then $K \le n$.
- Then $X \le Y_{e_1} + Y_{e_2} + \cdots + Y_{e_K}$.
- By symmetry of the hypercube, $\mathbb{E}[Y_e]$ is the same for all edges.
- By linearity of expectation,

$$\mathbb{E}[X|K = k] \le \mathbb{E}\left[Y_{e_1}\right] + \cdots + \mathbb{E}\left[Y_{e_k}\right] = k \cdot \mathbb{E}[Y_e]$$

Calculating $\mathbb{E}[Y_e]$

- For any edge e, let $Y_e = \#$ routes that pass via e.
- Suppose e is an edge in dimension $d \in [n]$, that is, $e = (x_1 \dots x_n, x_1 \dots x_{d-1}\overline{x_d} \dots x_n)$
- Only packets with source $* \cdots * x_d \dots x_n$ can traverse e
- To traverse e, such a packet must have destination which happens with probability
- Thus, $\mathbb{E}[Y_e] =$

Analysis of RRA

Main Lemma

Consider any packet i. It fails to reach its destination in phase 1 within $3n$ steps with probability at most $\frac{1}{N^2}$.

Proof: Let X be the number of packets (other than i) that use at least one edge from the path of i.

Find an upper bound on $\mathbb{E}[X]$

- For any edge e, let $Y_e = \#$ routes that pass via e.
- Let $p_i = (e_1, \ldots, e_K)$ be the path of packet i. Then $K \leq n$.
- By linearity of expectation,

$$\mathbb{E}[X|K = k] \leq \mathbb{E}\left[Y_{e_1}\right] + \cdots + \mathbb{E}\left[Y_{e_k}\right] = k \cdot \mathbb{E}[Y_e] =$$

- By compact law of total expectation, $\mathbb{E}[X]$

$$= \mathbb{E}\big[\mathbb{E}[X|K]\big] \leq \mathbb{E}[K/2] =$$

Analysis of RRA

Main Lemma

Consider any packet i. It fails to reach its destination in phase 1 within $3n$ steps with probability at most $\frac{1}{N^2}$.

Proof: Let X be the number of packets (other than i) that use at least one edge from the path of i.

- So far: $\mathbb{E}[X] \leq \frac{n}{4}$ and travel time for packet i is $\leq n + X$
- By Chernoff bound, $\Pr[X \geq 2n] =$

Analysis of RRA

Main Lemma

Consider any packet i. It fails to reach its destination in phase 1 within $3n$ steps with probability at most $\frac{1}{N^2}$.

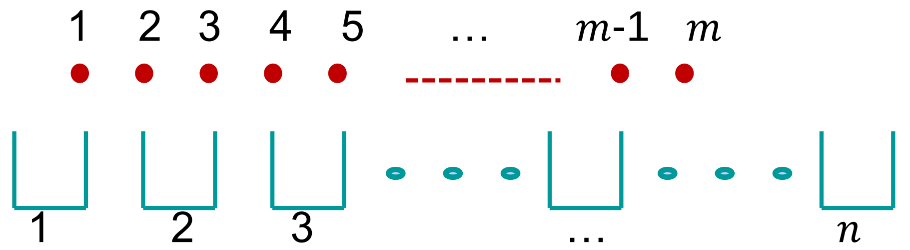
Using the Main Lemma to complete the analysis:

- By a union bound over N packets, the probability that at least one packet fails to reach its destination in phase 1 within $3n$ steps is

The Balls-and-Bins Model

- m balls thrown into n bins

Each ball falls into a uniformly random bin (u.i.r.)



Q1. Is it more likely that there is a collision or not?

(Birthday Paradox)

Q2. How many balls are in the fullest bin? (Maximum load)

Q3. How many bins are empty?

Q4. What does the distribution of the balls in the bins look like?

Birthday Paradox

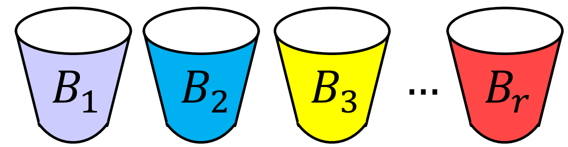
- $n = 365$ bins (days)
- What is the probability that all m people have different birthdays?
- For which m is the probability of collision more than 1/2?

Let E_i for $i \in [m]$ be the event that ball i falls into an empty bin.

Pr[no collision]

Bucket Sort

- Given: n integers from range $[r]$.
- If $r \leq n$, we can sort in $O(n)$ time
  - Use possible values as buckets
  - Keep a linked list for each bucket.
  - Make a pass over the list and put each element in the right bucket
  - Concatenate the lists.

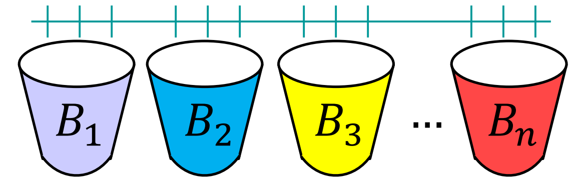


- What if $r > n$? (Suppose for simplicity that n divides r.)

> **Theorem**
>
> If n integers are chosen u.i.r. from range $[r]$, they can be sorted in expected time $O(n)$.

- Expectation is over randomness in choice of integers: Bucket Sort is deterministic.

Bucket Sort



- Idea: Break the range $[r]$ into n buckets.

The expected # of elements in each bucket is 1.

We can easily sort all buckets (say, using Insertion Sort)

Algorithm. Input: integers $a_1, \dots, a_n$

1. Make linked lists for buckets $B_1, \dots, B_n$.
2. For each $i \in [n]$, let $j = \left\lceil \frac{a_i \cdot n}{r} \right\rceil$ and add a_i to B_j.
3. Sort all buckets using Insertion Sort.
4. Output the concatenation of $B_1, \dots, B_n$

- Steps 1,2, and 4 can be implemented to run in $O(n)$ time.

Lemma

Step 3 runs in expected time $O(n)$.

Bucket Sort: Analysis

Lemma

Step 3 (sorting the buckets) runs in expected time $O(n)$.

Proof: Buckets are bins, elements are balls.

- Let $X_j =$ # of elements that land in bucket B_j, for $j \in [n]$.
- Time to sort B_j is: $\leq c \cdot X_j^2$ for some constant c
- Expected run time of Step 3:

$$\leq \mathbb{E}\left[\sum_{j\in[n]} cX_j^2\right] \overset{\text{by linearity of expectation}}{=} c \cdot \sum_{j\in[n]} \mathbb{E}\left[X_j^2\right] \overset{\text{by symmetry}}{=} cn \cdot \mathbb{E}\left[X_1^2\right]$$

Bucket Sort: Analysis

Lemma

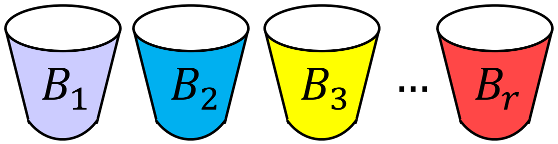
Step 3 (sorting the buckets) runs in expected time $O(n)$.

Proof: Buckets are bins, elements are balls in Balls-in-the-Bins.

- Let RV $X_j =$ # of elements that land in bucket B_j, for $j \in [n]$
- Expected run time of Step 3: $\leq cn \cdot \mathbb{E}\left[X_1^2\right]$
- $X_1 \sim$

Bucket Sort: Conclusion

- Given: n integers from range $[r]$.



Theorem

If n integers are chosen uniformly and independently from range $\{1, \ldots, r\}$, they can be sorted in expected time $O(n)$.