CMPSCI 711: "Really Advanced Algorithms"

Lecture 5 – Set Balancing and Routing in a Boolean Hypercube

Andrew McGregor

Last Compiled: February 12, 2009

Outline

Set Balancing

Let $A_1, \ldots, A_n$ be subsets of $[n]$ such that $|A_i| = n/2$. We want to partition $[n]$ into B and C such that

$$\max_i \big| |A_i \cap B| - |A_i \cap C| \big|$$

is minimized.

Hint: Use $\mathbb{P}\left[|X - \mathbb{E}\left[X\right]| < \delta\mu \right] \leq 2\exp(-\mathbb{E}\left[X\right]\delta^2/4)$.

Set Balancing Algorithm and Analysis

Algorithm: Consider a random partition!

Set Balancing Algorithm and Analysis

Algorithm: Consider a random partition!

Lemma

$\max_i \big||A_i \cap B| - |A_i \cap C|\big| \leq 4\sqrt{n \ln n}$ *with prob. at least* $1 - 2n^{-3}$.

Set Balancing Algorithm and Analysis

Algorithm: Consider a random partition!

Lemma

$\max_i \big||A_i \cap B| - |A_i \cap C|\big| \le 4\sqrt{n \ln n}$ *with prob. at least* $1 - 2n^{-3}$.

Proof.

- Let $X_j = 1$ if j-th element of A_i is in C and 0 otherwise.

□

Set Balancing Algorithm and Analysis

Algorithm: Consider a random partition!

Lemma

$\max_i \left| |A_i \cap B| - |A_i \cap C| \right| \leq 4\sqrt{n \ln n}$ *with prob. at least* $1 - 2n^{-3}$.

Proof.

- Let $X_j = 1$ if j-th element of A_i is in C and 0 otherwise.
- Then $X = \sum_j X_j = |A_i \cap C|$ and $\mathbb{E}[X] = n/4$

□

Set Balancing Algorithm and Analysis

Algorithm: Consider a random partition!

Lemma

$\max_i \left| |A_i \cap B| - |A_i \cap C| \right| \leq 4\sqrt{n \ln n}$ *with prob. at least* $1 - 2n^{-3}$*.*

Proof.

- Let $X_j = 1$ if j-th element of A_i is in C and 0 otherwise.
- Then $X = \sum_j X_j = |A_i \cap C|$ and $\mathbb{E}[X] = n/4$
- $\left| |A_i \cap B| - |A_i \cap C| \right| = \left| n/2 - 2|A_i \cap C| \right| = 2\left| \mathbb{E}[X] - X \right|$

□

Set Balancing Algorithm and Analysis

Algorithm: Consider a random partition!

Lemma

$\max_i \Big| |A_i \cap B| - |A_i \cap C| \Big| \leq 4\sqrt{n \ln n}$ *with prob. at least* $1 - 2n^{-3}$.

Proof.

- Let $X_j = 1$ if j-th element of A_i is in C and 0 otherwise.
- Then $X = \sum_j X_j = |A_i \cap C|$ and $\mathbb{E}[X] = n/4$
- $\Big| |A_i \cap B| - |A_i \cap C| \Big| = \Big| n/2 - 2|A_i \cap C| \Big| = 2\Big| \mathbb{E}[X] - X \Big|$
- By an application of the Chernoff bound:

$$\mathbb{P}\left[2\Big|\mathbb{E}[X] - X\Big| \geq 4\sqrt{n \ln n}\right] = \mathbb{P}\left[\Big|\mathbb{E}[X] - X\Big| \geq 8\sqrt{n^{-1} \ln n} \cdot \mathbb{E}[X]\right]$$
$$\leq 2e^{-(n/4)(64n^{-1} \ln n)/4} = 2n^{-4}$$

□

Set Balancing Algorithm and Analysis

Algorithm: Consider a random partition!

Lemma

$\max_i \big||A_i \cap B| - |A_i \cap C|\big| \leq 4\sqrt{n \ln n}$ *with prob. at least* $1 - 2n^{-3}$.

Proof.

- Let $X_j = 1$ if j-th element of A_i is in C and 0 otherwise.
- Then $X = \sum_j X_j = |A_i \cap C|$ and $\mathbb{E}[X] = n/4$
- $\big||A_i \cap B| - |A_i \cap C|\big| = \big|n/2 - 2|A_i \cap C|\big| = 2\big|\mathbb{E}[X] - X\big|$
- By an application of the Chernoff bound:

$$\mathbb{P}\left[2\big|\mathbb{E}[X] - X\big| \geq 4\sqrt{n \ln n}\right] = \mathbb{P}\left[\big|\mathbb{E}[X] - X\big| \geq 8\sqrt{n^{-1} \ln n} \cdot \mathbb{E}[X]\right]$$
$$\leq 2e^{-(n/4)(64n^{-1} \ln n)/4} = 2n^{-4}$$

- Apply union bound over all i.

□

Outline

Routing in Boolean Hypercube

Boolean hypercube:

- $N = 2^n$ nodes where each is labeled by a length $\{0,1\}^n$
- Edge between $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$ iff $\Delta(x,y) = 1$.

Routing in Boolean Hypercube

Boolean hypercube:

- $N = 2^n$ nodes where each is labeled by a length $\{0,1\}^n$
- Edge between $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$ iff $\Delta(x, y) = 1$.

Problem: Let π be a permutation of $[N]$

- For $i \in [N]$, packet v_i needs routed from i-th node to $\pi(i)$-th node.
- At each step, a packet may traverse an edge (or stay still).
- At most one packet can traverse an edge in the same step.

Routing in Boolean Hypercube

Boolean hypercube:

- $N = 2^n$ nodes where each is labeled by a length $\{0, 1\}^n$
- Edge between $x \in \{0, 1\}^n$ and $y \in \{0, 1\}^n$ iff $\Delta(x, y) = 1$.

Problem: Let π be a permutation of $[N]$

- For $i \in [N]$, packet v_i needs routed from i-th node to $\pi(i)$-th node.
- At each step, a packet may traverse an edge (or stay still).
- At most one packet can traverse an edge in the same step.

Want an "Oblivious Routing":

- Pick route ρ_i for v_i that only depends only on $\pi(i)$.
- Pick rule for deciding who gets precedence when two packets want to use same edge.

The Algorithm

The Algorithm

- For each v_i, pick random intermediate destination $\sigma(i)$
- Phase 1: Route each v_i from i-th node to $\sigma(i)$-th node
- Phase 2: Route each v_i from $\sigma(i)$-th node to $\pi(i)$-th node
- Use "first in, first out" queueing policy.

The Algorithm

- For each v_i, pick random intermediate destination $\sigma(i)$
- Phase 1: Route each v_i from i-th node to $\sigma(i)$-th node
- Phase 2: Route each v_i from $\sigma(i)$-th node to $\pi(i)$-th node
- Use "first in, first out" queueing policy.

In each phase, route using "bit-fixing":

- At each step, forward v_i to neighboring node whose label agrees with longest prefix of label of $\sigma(i)$-th node
- E.g., to get from (1011) to (0110), the route would be $(1011) \rightarrow (0011) \rightarrow (0111) \rightarrow (0110)$

The Algorithm

- For each v_i, pick random intermediate destination $\sigma(i)$
- Phase 1: Route each v_i from i-th node to $\sigma(i)$-th node
- Phase 2: Route each v_i from $\sigma(i)$-th node to $\pi(i)$-th node
- Use "first in, first out" queueing policy.

In each phase, route using "bit-fixing":

- At each step, forward v_i to neighboring node whose label agrees with longest prefix of label of $\sigma(i)$-th node
- E.g., to get from (1011) to (0110), the route would be $(1011) \to (0011) \to (0111) \to (0110)$

Once two paths diverge, they don't merge again.

The Result

Theorem

All packets get routed to final destination in $14n$ steps with probability at least $1 - 2/N$.

The Result

Theorem
All packets get routed to final destination in $14n$ steps with probability at least $1 - 2/N$.

For comparison (we won't prove this):

Theorem
For any deterministic oblivious routing algorithm, there is a permutation π that requires $\Omega(\sqrt{2^n/n})$ time.

Analysis Overview

- Let ρ_i be the path taken by v_i in Phase 1.

Analysis Overview

- Let ρ_i be the path taken by v_i in Phase 1.
- Let $H_{i,j} = 1$ if ρ_i and ρ_j intersect and 0 otherwise ($H_{i,i} = 0$).

Analysis Overview

- Let ρ_i be the path taken by v_i in Phase 1.
- Let $H_{i,j} = 1$ if ρ_i and ρ_j intersect and 0 otherwise ($H_{i,i} = 0$).
- $H_{i,1}, H_{i,2}, \ldots, H_{i,N}$ are independent Poisson trials for each i.

Analysis Overview

- Let ρ_i be the path taken by v_i in Phase 1.
- Let $H_{i,j} = 1$ if ρ_i and ρ_j intersect and 0 otherwise ($H_{i,i} = 0$).
- $H_{i,1}, H_{i,2}, \ldots, H_{i,N}$ are independent Poisson trials for each i.
- Thm: Time for v_i to get to intermediate destination is

$$\text{Length of Path} + \text{Delays} \leq n + \sum_j H_{i,j}$$

Analysis Overview

- Let ρ_i be the path taken by v_i in Phase 1.
- Let $H_{i,j} = 1$ if ρ_i and ρ_j intersect and 0 otherwise ($H_{i,i} = 0$).
- $H_{i,1}, H_{i,2}, \ldots, H_{i,N}$ are independent Poisson trials for each i.
- Thm: Time for v_i to get to intermediate destination is

$$\text{Length of Path} + \text{Delays} \leq n + \sum_j H_{i,j}$$

- Thm: With probability at least $1 - 2^{-6n}$, $\sum_j H_{i,j} \leq 6n$

Analysis Overview

- Let ρ_i be the path taken by v_i in Phase 1.
- Let $H_{i,j} = 1$ if ρ_i and ρ_j intersect and 0 otherwise ($H_{i,i} = 0$).
- $H_{i,1}, H_{i,2}, \ldots, H_{i,N}$ are independent Poisson trials for each i.
- Thm: Time for v_i to get to intermediate destination is

$$\text{Length of Path} + \text{Delays} \leq n + \sum_j H_{i,j}$$

- Thm: With probability at least $1 - 2^{-6n}$, $\sum_j H_{i,j} \leq 6n$
- By union bound, with probability at least $1 - 2^{-5n} \geq 1 - 1/N$, all packets get to intermediate destination in at most $7n$ time.

Analysis Overview

- Let ρ_i be the path taken by v_i in Phase 1.
- Let $H_{i,j} = 1$ if ρ_i and ρ_j intersect and 0 otherwise ($H_{i,i} = 0$).
- $H_{i,1}, H_{i,2}, \ldots, H_{i,N}$ are independent Poisson trials for each i.
- Thm: Time for v_i to get to intermediate destination is

$$\text{Length of Path} + \text{Delays} \leq n + \sum_j H_{i,j}$$

- Thm: With probability at least $1 - 2^{-6n}$, $\sum_j H_{i,j} \leq 6n$
- By union bound, with probability at least $1 - 2^{-5n} \geq 1 - 1/N$, all packets get to intermediate destination in at most $7n$ time.
- Analysis of Phase 2 is identical.

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (1/2)

Theorem

Total delay incurred by v_i is at most $\sum_j H_{i,j}$.

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (1/2)

Theorem

Total delay incurred by v_i is at most $\sum_j H_{i,j}$.

Proof.

- Let route of v_i be $\rho_i = (e_1, e_2, \ldots, e_k)$.

□

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (1/2)

Theorem

Total delay incurred by v_i is at most $\sum_j H_{i,j}$.

Proof.

- Let route of v_i be $\rho_i = (e_1, e_2, \ldots, e_k)$.
- Let S be the set of $\sum_j H_{i,j}$ packets (other than v_i) whose routes pass through an edge in ρ_i

□

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (1/2)

Theorem

Total delay incurred by v_i is at most $\sum_j H_{i,j}$.

Proof.

- Let route of v_i be $\rho_i = (e_1, e_2, \ldots, e_k)$.
- Let S be the set of $\sum_j H_{i,j}$ packets (other than v_i) whose routes pass through an edge in ρ_i
- Say $v \in S$ leaves ρ_i at the time step at which it traverses an edge in ρ_i for the last time.

□

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (1/2)

Theorem

Total delay incurred by v_i is at most $\sum_j H_{i,j}$.

Proof.

- Let route of v_i be $\rho_i = (e_1, e_2, \ldots, e_k)$.
- Let S be the set of $\sum_j H_{i,j}$ packets (other than v_i) whose routes pass through an edge in ρ_i
- Say $v \in S$ leaves ρ_i at the time step at which it traverses an edge in ρ_i for the last time.
- If a packet is ready to traverse edge e_j at time t, we define its lag at time t to be $t - j$. Final lag for v_i is the delay.

□

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (1/2)

Theorem

Total delay incurred by v_i is at most $\sum_j H_{i,j}$.

Proof.

- Let route of v_i be $\rho_i = (e_1, e_2, \ldots, e_k)$.
- Let S be the set of $\sum_j H_{i,j}$ packets (other than v_i) whose routes pass through an edge in ρ_i
- Say $v \in S$ leaves ρ_i at the time step at which it traverses an edge in ρ_i for the last time.
- If a packet is ready to traverse edge e_j at time t, we define its lag at time t to be $t - j$. Final lag for v_i is the delay.
- Claim: If lag of v_i when it traverses e_k is ℓ then for each $\ell' \leq \ell$ there exists a $v \in S$ that leaves ρ_i with lag ℓ'.

□

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (1/2)

Theorem

Total delay incurred by v_i is at most $\sum_j H_{i,j}$.

Proof.

- Let route of v_i be $\rho_i = (e_1, e_2, \ldots, e_k)$.
- Let S be the set of $\sum_j H_{i,j}$ packets (other than v_i) whose routes pass through an edge in ρ_i
- Say $v \in S$ leaves ρ_i at the time step at which it traverses an edge in ρ_i for the last time.
- If a packet is ready to traverse edge e_j at time t, we define its lag at time t to be $t - j$. Final lag for v_i is the delay.
- Claim: If lag of v_i when it traverses e_k is ℓ then for each $\ell' \leq \ell$ there exists a $v \in S$ that leaves ρ_i with lag ℓ'.
- Hence $|S| \leq$ (final lag of v_i) $=$ (total delay incurred by v_i).

□

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (2/2)

Claim

If lag of v_i when it traverses e_k is ℓ then for each $\ell' < \ell$ there exists a $v \in S$ that leaves ρ_i with lag ℓ'.

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (2/2)

Claim

If lag of v_i when it traverses e_k is ℓ then for each $\ell' < \ell$ there exists a $v \in S$ that leaves ρ_i with lag ℓ'.

Proof.

□

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (2/2)

Claim

If lag of v_i when it traverses e_k is ℓ then for each $\ell' < \ell$ there exists a $v \in S$ that leaves ρ_i with lag ℓ'.

Proof.

- There exists a packet in S that has lag ℓ' at some stage:
  1. Consider the step when v_i increases from lag ℓ' to $\ell' + 1$ because it cannot use some edge e
  2. At this step, the packet that does use e has lag ℓ'.

□

Analysis Part 1: Expressing delay in terms of $H_{i,j}$ (2/2)

Claim

If lag of v_i when it traverses e_k is ℓ then for each $\ell' < \ell$ there exists a $v \in S$ that leaves ρ_i with lag ℓ'.

Proof.

- There exists a packet in S that has lag ℓ' at some stage:
  1. Consider the step when v_i increases from lag ℓ' to $\ell' + 1$ because it cannot use some edge e
  2. At this step, the packet that does use e has lag ℓ'.
- Let t' be the last time at which a packet in S has lag ℓ'.
  1. All packets waiting to traverse e_j for $j = t' - \ell'$ have lag ℓ'.
  2. The packet that traverses e_j still has lag ℓ' unless it leaves ρ_i.
  3. So if it doesn't leave ρ_i, t' wasn't the last time at which a packet in S has lag ℓ'.

□

Analysis Part 2: Bounding tail probability of delay (1/2)

Let $h_i = \sum_j H_{i,j}$. By Chernoff bound, $\mathbb{P}\left[h_i \geq 6\mathbb{E}\left[h_i\right]\right] \leq 2^{-6\mathbb{E}[h_i]}$.

Analysis Part 2: Bounding tail probability of delay (1/2)

Let $h_i = \sum_j H_{i,j}$. By Chernoff bound, $\mathbb{P}\left[h_i \geq 6\mathbb{E}\left[h_i\right]\right] \leq 2^{-6\mathbb{E}[h_i]}$.

Lemma

$\mathbb{E}\left[h_i\right] \leq n/2$ and so $\mathbb{P}\left[h_i \geq 6n\right] \leq 2^{-6n}$.

Analysis Part 2: Bounding tail probability of delay (1/2)

Let $h_i = \sum_j H_{i,j}$. By Chernoff bound, $\mathbb{P}\left[h_i \geq 6\mathbb{E}\left[h_i\right]\right] \leq 2^{-6\mathbb{E}[h_i]}$.

Lemma

$\mathbb{E}\left[h_i\right] \leq n/2$ and so $\mathbb{P}\left[h_i \geq 6n\right] \leq 2^{-6n}$.

Proof.

- Let $T(e)$ be number of routes that include edge e.

□

Analysis Part 2: Bounding tail probability of delay (1/2)

Let $h_i = \sum_j H_{i,j}$. By Chernoff bound, $\mathbb{P}[h_i \geq 6\mathbb{E}[h_i]] \leq 2^{-6\mathbb{E}[h_i]}$.

Lemma

$\mathbb{E}[h_i] \leq n/2$ and so $\mathbb{P}[h_i \geq 6n] \leq 2^{-6n}$.

Proof.

- Let $T(e)$ be number of routes that include edge e.
- $h_i \leq \sum_{e \in \rho_i} T(e)$ and, by linearity of expectation:

$$\mathbb{E}[h_i] \leq \sum_{e \in \rho_i} \mathbb{E}[T(e)]$$

□

Analysis Part 2: Bounding tail probability of delay (1/2)

Let $h_i = \sum_j H_{i,j}$. By Chernoff bound, $\mathbb{P}[h_i \geq 6\mathbb{E}[h_i]] \leq 2^{-6\mathbb{E}[h_i]}$.

Lemma

$\mathbb{E}[h_i] \leq n/2$ and so $\mathbb{P}[h_i \geq 6n] \leq 2^{-6n}$.

Proof.

- Let $T(e)$ be number of routes that include edge e.
- $h_i \leq \sum_{e \in \rho_i} T(e)$ and, by linearity of expectation:

$$\mathbb{E}[h_i] \leq \sum_{e \in \rho_i} \mathbb{E}[T(e)]$$

- Claim: $\mathbb{E}[T(e)] = 1/2$.

□

Analysis Part 2: Bounding tail probability of delay (1/2)

Let $h_i = \sum_j H_{i,j}$. By Chernoff bound, $\mathbb{P}[h_i \geq 6\mathbb{E}[h_i]] \leq 2^{-6\mathbb{E}[h_i]}$.

Lemma

$\mathbb{E}[h_i] \leq n/2$ and so $\mathbb{P}[h_i \geq 6n] \leq 2^{-6n}$.

Proof.

- Let $T(e)$ be number of routes that include edge e.
- $h_i \leq \sum_{e \in \rho_i} T(e)$ and, by linearity of expectation:

$$\mathbb{E}[h_i] \leq \sum_{e \in \rho_i} \mathbb{E}[T(e)]$$

- Claim: $\mathbb{E}[T(e)] = 1/2$.
- Hence, $\mathbb{E}[h_i] \leq \sum_{e \in \rho_i} \mathbb{E}[T(e)] = |\rho_i|/2 \leq n/2$

□

Analysis Part 2: Bounding tail probability of delay (2/2)

Claim

$\mathbb{E}[T(e)] = 1/2$.

Analysis Part 2: Bounding tail probability of delay (2/2)

Claim

$\mathbb{E}[T(e)] = 1/2$.

Proof.

- Expected number of bits to be fixed is $n/2$: $\mathbb{E}[|\rho_j|] = n/2$.

□

Analysis Part 2: Bounding tail probability of delay (2/2)

Claim

$\mathbb{E}[T(e)] = 1/2$.

Proof.

- Expected number of bits to be fixed is $n/2$: $\mathbb{E}[|\rho_j|] = n/2$.
- By linearity of expectation: $\mathbb{E}\left[\sum_j |\rho_j|\right] = Nn/2$.

□

Analysis Part 2: Bounding tail probability of delay (2/2)

Claim

$\mathbb{E}[T(e)] = 1/2$.

Proof.

- Expected number of bits to be fixed is $n/2$: $\mathbb{E}[|\rho_j|] = n/2$.
- By linearity of expectation: $\mathbb{E}\left[\sum_j |\rho_j|\right] = Nn/2$.
- Each edge in a route contributes to a $T(e)$ and vice versa:

$$\sum_j |\rho_j| = \sum_e T(e)$$

□

Analysis Part 2: Bounding tail probability of delay (2/2)

Claim

$\mathbb{E}[T(e)] = 1/2$.

Proof.

- Expected number of bits to be fixed is $n/2$: $\mathbb{E}[|\rho_j|] = n/2$.
- By linearity of expectation: $\mathbb{E}\left[\sum_j |\rho_j|\right] = Nn/2$.
- Each edge in a route contributes to a $T(e)$ and vice versa:

$$\sum_j |\rho_j| = \sum_e T(e)$$

- Total number of edges is Nn and $\mathbb{E}[T(e)]$ is the same for all e.

□

Outline

Set Balancing

Routing in Boolean Hypercube

Readings

Readings

For next time, please make sure you've read:

- Chapter 4: Up to and including 4.2 (12 pages)