F, 11/22/19

Fall’19 CSCE 629

Analysis of Algorithms

Fang Song
Texas A&M U

Lecture 33

- Randomized algorithms

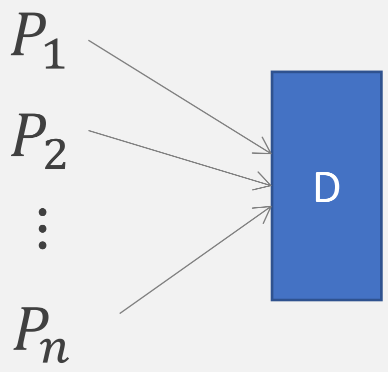
Contention resolution in a distributed system

Given: processes $P_1, \dots, P_n$,

- each process competes for access to a shared database.
- If ≥ 2 processes access the database simultaneously, all processes are locked out.

Goal: a protocol so all processes get through on a regular basis

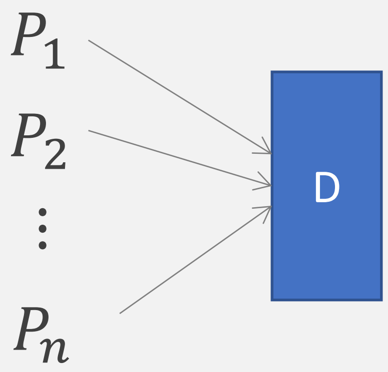
- Restriction: Processes can't communicate.



Contention resolution: randomized protocol

Protocol. Each process requests access to the database in round t with probability $p = 1/n$.

Theorem. All processes will succeed in accessing the database at least once within $O(n \ln n)$ rounds except with probability $\leq \frac{1}{n}$.



Randomized contention resolution: analysis 1

Def. $S[i, t]$ = event that process i succeeds in accessing the database in round t.

- Claim1. $\frac{1}{e \cdot n} \leq \Pr(S[i, t]) \leq \frac{1}{2n}$
- Pf. $\Pr(S[i, t]) = p(1-p)^{n-1}$

[Geometric distribution: independent Bernoulli trials]

Process i requests access None of remaining request access

$\Rightarrow \Pr(S[i, t]) = \frac{1}{n}(1 - 1/n)^{n-1} \in [\frac{1}{en}, \frac{1}{2n}]$ $[p = 1/n]$

- $(1 - 1/n)^n$ converges monotonically from $1/4$ up to $1/e$.
- $(1 - 1/n)^{n-1}$ converges monotonically from $1/2$ down to $1/e$.

Randomized contention resolution: analysis 2

- Claim2. The probability that process i fails to access the database in $e \cdot n$ rounds is at most $1/e$. After $e \cdot n\ (c \ln n)$ rounds, the probability $\leq\ n^{-c}$.

- Pf. Let $F[i, t]$ = event that process i fails to access database in rounds 1 through t.

$$\Pr(F[i,t]) = \Pr\left(\overline{S[i,1]}\right) \cdot \ldots \cdot \Pr\left(\overline{S[i,t]}\right) \leq \left(1 - \frac{1}{en}\right)^{t}$$ [Independence & Claim 1]

  - Choose $t = en$: $\Pr(F[i,t]) \leq \left(1 - \frac{1}{en}\right)^{en} \leq \frac{1}{e}$
  - Choose $t = en \cdot \mathrm{c}\ln n$: $\Pr(F[i,t]) \leq \left(\frac{1}{e}\right)^{\mathrm{c}\ln n} \leq n^{-c}$

Randomized contention resolution: analysis 3

Theorem. All processes will succeed in accessing the database *at least once* within $2en \ln n$ rounds except with probability $\leq \frac{1}{n}$.

- Pf. Let $F[t]$ = event that some process fails to access database in rounds 1 through t.

Union Bound

Let E, F be two events. Then $\Pr(E \cup F) \leq \Pr(E) + \Pr(F)$.

$$\Pr(F[t]) = \Pr(\cup_{i=1}^{n} F[i,t]) \leq \sum_{i=1}^{n} \Pr(F[i,t]) \leq n \cdot \Pr(F[1,t])$$

- Choose $t = en \cdot 2\ln n$: $\Pr(F[t]) \leq n \cdot n^{-2} = 1/n$

Recall: quick sort

- Main Idea
  - Divide array into two halves.
  - Recursively sort each half.
  - Merge two halves to make sorted whole.

with condition: $L \leq pivot \leq \ R$

trivially

- Analysis
  - Correctness
  - Running time*

$$T(n) = 2T(n/2) + O(n)$$

Cost in divide, not merge

* best-case partition

- Can you think of a worst-case scenario?

Randomized quicksort

- Pick the pivot randomly

```
Rand-QuickSort(A):
 if (array A has zero or one element)
    Return
 Pick pivot p ∈ A uniformly at random
 (L, M, R) ← PARTITION-3-WAY(A, p)      → O(n)
 Rand-QuickSort(L)                      → T(i)
 Rand-QuickSort(R)                      → T(n − i − 1)
```

Theorem. The expected number of compares to quicksort an array of n distinct elements is $O(n\log n)$.

Probability 102

- Random variable $X: \Omega \to \mathbb{N}$
  - Assign each outcome a number
  - "$X = x$" is the event $E \coloneqq \{\omega \in \Omega: X(\omega) = x\}$
  - Independent random variables:

X, Y are indep. iff. for all possible x and y, events $X = x$ and $Y = y$ are indep.

- Expectation: a weighed average
  - $\mathbb{E}[X] = \sum_{z \in Z} \Pr(X = z) \cdot z$
  - Linearity: $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ (independence NOT needed)
- Ex. $\Omega =$ roll 4 dices independently
  - Let X be the sum of 4 rolls; X_i be value of ith roll, $i = 1, \dots, 4$
  - $\mathbb{E}[X] = \mathbb{E}[X_1 + \cdots + X_4] = 4 \cdot \mathbb{E}[X_1] = 4 \times 3.5 = 14$

Randomized quicksort: analysis

Theorem. The expected number of compares to quicksort an array of n distinct elements is $O(n \log n)$.

Assume $A = \{z_1, z_2, \ldots, z_n\}, z_1, < z_2 < \cdots < z_n$

Observation: any pair z_i & z_j $(i < j)$ is compared at most once

- How many comparisons? $X :=$ total number of comparisons
  - Indicator variable: $X_{ij} := \begin{cases} 1, \text{if } z_i \text{ is compared to } z_j \\ 0, \text{otherwise} \end{cases}$

$$\Rightarrow \mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right]$$
$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[X_{ij}] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr[X_{ij} = 1]$$

Linearity

Randomized quicksort: analysis cont'd

Theorem. The expected number of compares to quicksort an array of n distinct elements is $O(n \log n)$.

$$\boxed{\mathbb{E}[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr[X_{ij} = 1]} \qquad X_{ij} := \begin{cases} 1, \text{if } z_i \text{ is compared to } z_j \\ 0, \text{otherwise} \end{cases}$$

- When two items are compared?

7

2 1 3 $\underbrace{4\ 6\ 5 \qquad\qquad 9\ 8}_{\text{No comparison across these two groups}}$ 10

- Observation: z_i & z_j compared iff. z_i or z_j was the first chosen as a pivot from $Z_{ij} = \{z_i, z_{i+1}, \dots, z_j\}$

Randomized quicksort: analysis cont'd

- Observation: z_i & z_j compared iff. z_i or z_j was the first chosen as a pivot from $Z_{ij} = \{z_i, z_{i+1}, \dots, z_j\}$

$$
\begin{aligned}
&\Pr[X_{ij} = 1] \\
&= \Pr[z_i \ \& \ z_j \text{ compared}] = \Pr[z_i \text{ or } z_j \text{ is 1st pivot chosen from } Z_{ij}] \\
&= \Pr[z_i \text{ is 1st pivot from } Z_{ij}] + \Pr[z_j \text{ is 1st pivot from } Z_{ij}] \\
&= \frac{1}{j-i+1} + \frac{1}{j-i+1} = \frac{2}{j-i+1}
\end{aligned}
$$

$$
\mathbb{E}[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \frac{2}{k+1} \le 2 \cdot \sum_{i=1}^{n-1} \boxed{\sum_{k=1}^{n} \frac{1}{k}} = O(n \cdot \log n)
$$

Harmonic series

LP relaxation for set cover

(Set cover ILP Π) Min $\sum_{i=1}^{m} x_i$
Subject to:
$$\sum_{i:u\in S_i} x_i \geq 1, \ \forall u \in U$$
$$x_i \in \{0,1\}, \qquad \forall i \in \{1, \dots, m\}$$

(Set cover LP Σ) Min $\sum_{i=1}^{m} x_i$
Subject to:
$$\sum_{i:u\in S_i} x_i \geq 1, \ \forall u \in U$$
$$0 \leq x_i \leq 1, \forall i \in \{1, \dots, m\}$$

Let x^* be an optimal soln. for LP Σ & optimal value $\mathrm{OPT} = \sum_i x_i^*$

☹ $x_i \coloneqq \lfloor x_i^* \rceil$

- Randomized rounding: set $x_i = 1$ with probability x_i^*

$$\mathbb{E}[\sum_{i=1}^{m} x_i] = \sum_{i=1}^{m} \mathbb{E}[x_i] = \sum_{i=1}^{m} x_i{}^*$$

- But is it feasible? [Further analysis on board & Panigrahi's notes]

Theorem. There is a poly-time randomized algorithm achieving $O(\log n)$ expected approximation ratio, except w. probability $O(1/n)$.