LOAD BALANCING

1. m servers
2. n jobs $(0..n-1)$, $n \gg m$

FAIR LOAD: $\frac{m}{n}$ jobs per server

GOAL
- Transparency (open source)
- Persistence (job i always go to a small group of servers)
- fair load $\frac{n}{m}$

(round robin creates DoS)

Randomly choose $h \in \mathcal{H}$ universal, so that

job $i \rightsquigarrow$ server $h(i)$

$X_j = \left[\text{Load for machine } j\right] = \left[\#\text{ jobs assigned to } j\right]$

$$X_{ji} = \begin{cases} 1 & \text{if job } i \text{ is assigned to server } j = h(i) \\ 0 & \text{o.w.} \end{cases} \qquad pr = \frac{1}{m}$$

$$X_j = \sum_{i=0}^{n-1} X_{ji} \Rightarrow E\left[X_j\right] = \sum_{i=0}^{n-1} E\left[X_{ji}\right] = \sum_{i=0}^{n-1} \underbrace{Pr\left[X_{ji}=1\right]}_{\frac{1}{m}} = \frac{n}{m} \quad \text{FAIR LOAD}$$

what about the <u>max</u> load?

CONCETRATION BOUNDS : how close to expectation

- Markov's Inequality (MI)

$$\Pr[X \geq a] \leq \frac{E[X]}{a}$$ (here, we need $a = E[X]$)

- Chebyshev's inequality (CI)

$\sigma^2 = E[X^2] - E[X]^2$ variance

$$\Pr[\,|X - E[X]| \geq k\sigma\,] \leq \frac{1}{k^2}$$

$$\Pr\left[\,|X - E[X]| \geq k\sigma\,\right] \leq \frac{1}{k^2}$$

$$X_j = \sum_{i=0}^{n-1} X_{ji}\,, \quad X_{ji}^2 = X_{ji} \in \{0,1\}$$

$$\sigma^2 = E[X_j^2] - E[X_j]^2$$

SEE NEXT SLIDE

$$\sum_{i=0}^{n-1}\Big(\underbrace{E[X_{ji}^2]}_{\substack{= E[X_{ji}] \\ = \frac{1}{m}}} - \underbrace{E[X_{ji}]^2}_{=\left(\frac{1}{m}\right)^2}\Big) = n\left(\frac{1}{m} - \frac{1}{m^2}\right)$$

$$\sigma = \sqrt{n\left(\frac{1}{m} - \frac{1}{m^2}\right)}$$

$$k = \sqrt{2m}$$

$$\Pr\left[\,\left|X_j - \frac{n}{m}\right| \geq \sqrt{2m}\,\sigma\,\right] \leq \frac{1}{2m}$$

$$\sqrt{2m}\cdot\sigma = \sqrt{2\cancel{m}\cdot n\left(\frac{1}{\cancel{m}} - \cancel{\frac{1}{m^2}}\right)} \sim \sqrt{2n}$$

$$\left.\right\} \Pr\left[\,\left|X_j - \frac{n}{m}\right| \geq \sqrt{2n}\,\right] \leq \frac{1}{2m} \quad (*)$$

server j

NOTE

1) $\Pr[X_{ji}=1 \wedge X_{ji'}=1] = \Pr[X_{ji}=1] \cdot \Pr[X_{ji'}=1]$

$= \Pr[h(i)=j] \cdot \Pr[h(i')=j]$

as $h \in H$ is 2-way independent

2) $X+Y+Z$ pairwise independent $\Rightarrow$

$\sigma^2(x+y+z) = \sigma^2(x) + \sigma^2(y) + \sigma^2(y)$

$1)+2) \Rightarrow$
$$\sigma^2(X_j) = \sum_{i=0}^{n-1} \sigma^2(X_{ji})$$
see previous slide

UNION BOUND

$$\Pr\left[\max_j \left|X_j - \frac{n}{m}\right| < \sqrt{2n}\right] = 1 - \Pr\left[\exists j : \left|X_j - \frac{n}{m}\right| \geq \sqrt{2n}\right]$$

$$= 1 - \sum_{j=0}^{m-1} \underbrace{CI}_{\Pr \leq \frac{1}{2m}} \geq \frac{1}{2}$$

$$\leq \frac{1}{2}$$

In words, the max load is at most $\frac{n}{m} + \sqrt{2n}$ with pr $\geq \frac{1}{2}$

- Chernoff's bounds (CB)

  set of i.i.d variables $y_i \in [0,1]$, $Y = \sum_{i=1}^{n} y_i$, $\mu = E[Y]$

  independent identically distributed

  - $Pr[Y > \mu + \lambda] \leq e^{-\frac{\lambda^2}{2\mu+\lambda}}$

    our slack parameter

  - $Pr[Y < \mu - \lambda] \leq e^{-\frac{\lambda^2}{3\mu}}$

$$\Pr[Y > \mu + \lambda] \le e^{-\frac{\lambda^2}{2\mu+\lambda}}$$

hp. $n = m$: $\mu = E[X_j] = \frac{n}{m} = 1$

$Y = X_j$

$\lambda = 6 \ln n$

$$(*)\ \Pr\left[X_j > \frac{n}{m} + 6 \ln n\right] \le e^{-\frac{(6\ln n)^2}{2\frac{n}{m} + 6 \ln n}} \le e^{-3 \ln n} = \frac{1}{n^3} \quad \text{w.h.p}$$

$$-\frac{(6 \ln n)^2}{\underbrace{2 \cdot 1 + 6 \ln n}} \le -\frac{(\cancel{6} \ln n)^{\cancel{2}}}{\cancel{2} \cdot \cancel{6 \ln n}} = -3 \ln n$$

$$\Pr\left[\max_j X_j \le \frac{n}{m} + \boxed{6 \ln n}\right] = 1 - \sum_{j=1}^{m} (*) \ge 1 - n \cdot \frac{1}{n^3} = 1 - \frac{1}{n^2}$$