Balls into Bins

Randomized Load Balancing

Outline

- Load balancing
- Balls-into-Bins
  - Modeling
  - Markov
  - Chebyshev
  - Chernoff

Outline

- Load balancing
- Balls-into-Bins
  - Modeling
  - Markov
  - Chebyshev
  - Chernoff

Learning objectives:
You are able to
- describe the load balancing problem
- state the expected load per server if the loads are assigned uniformly at random
- apply Markov's inequality to bound the probability of a high load on a fixed server
- apply a Chernoff bound to bound the probability of a high load on any server

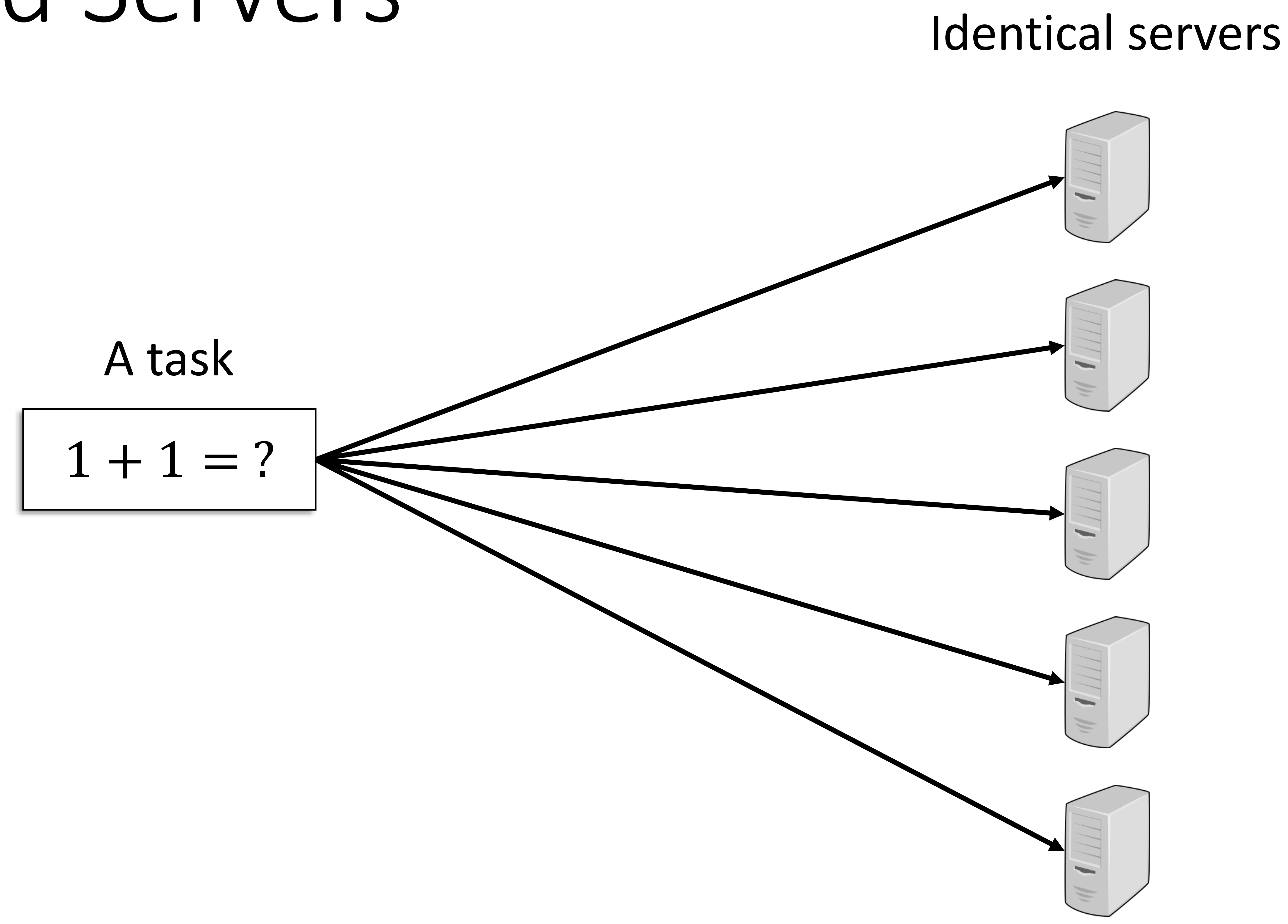
Tasks and Servers

Identical servers

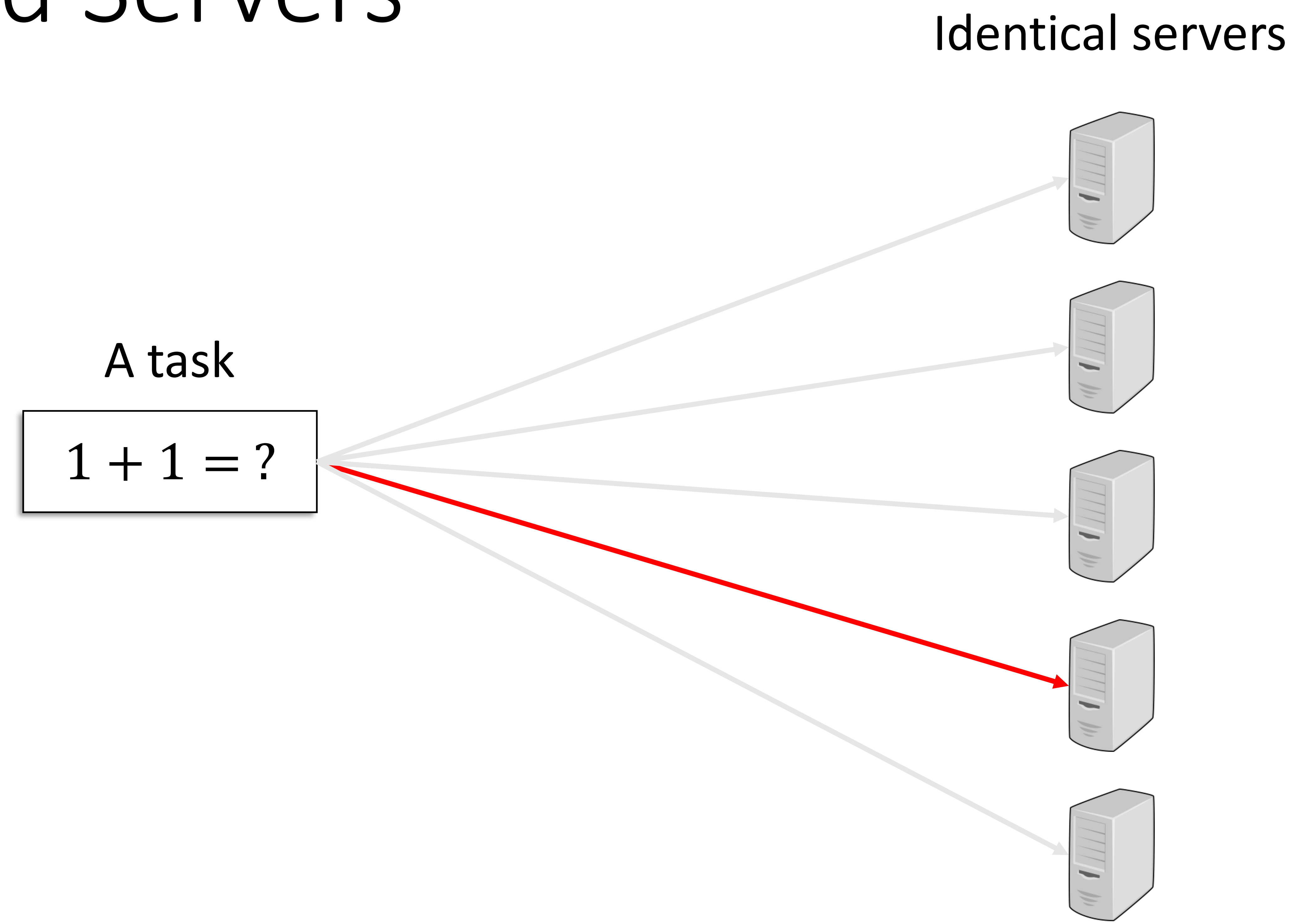
A task

$$1 + 1 = ?$$

Tasks and Servers



Tasks and Servers

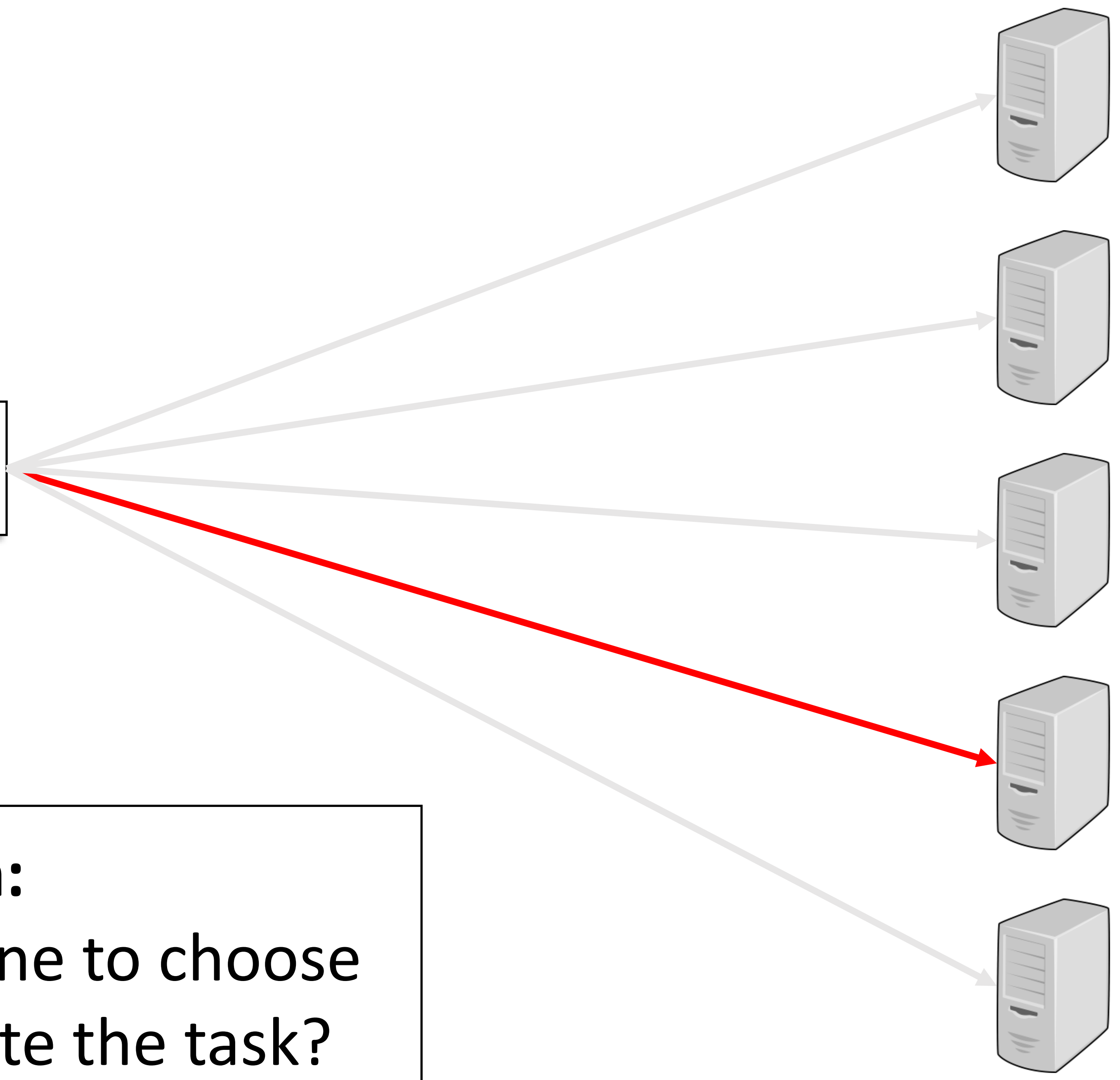


Tasks and Servers

Identical servers

A task

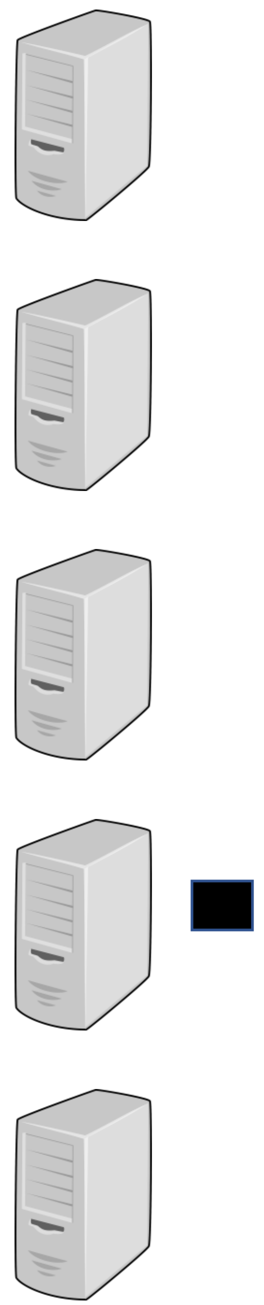
$1 + 1 = ?$



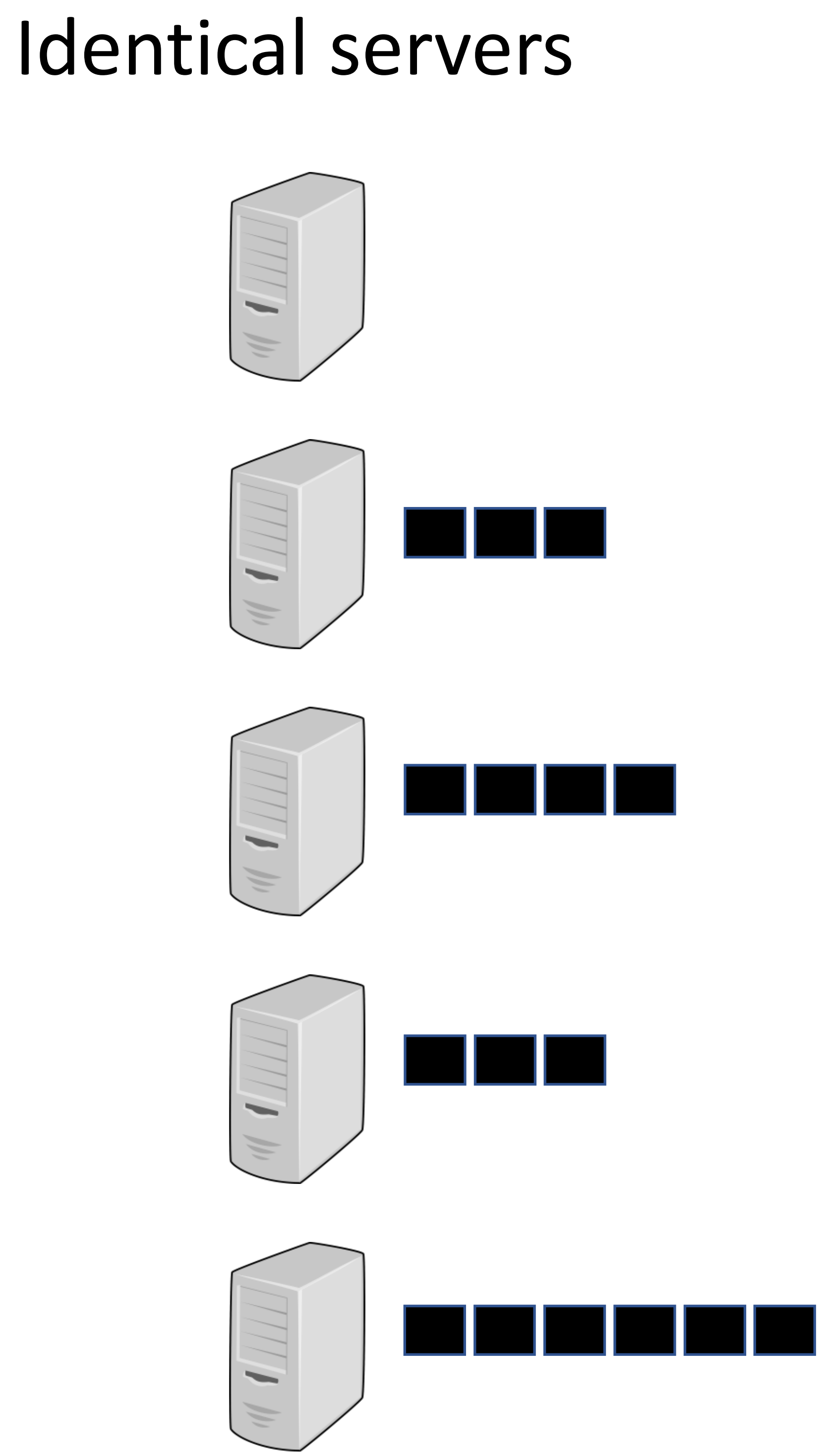
Problem:
Which one to choose to execute the task?

Tasks and Servers

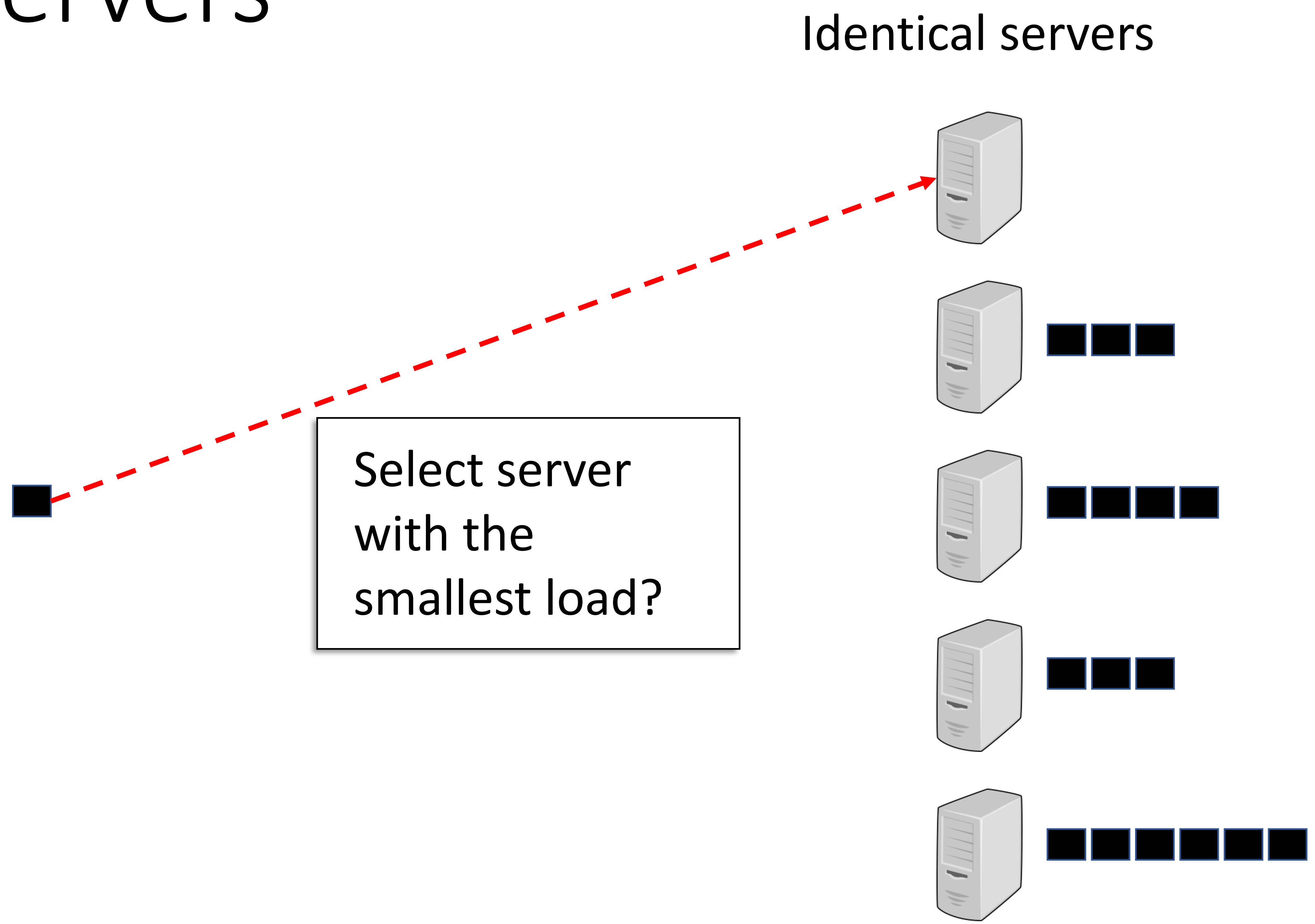
Identical servers



Tasks and Servers



Tasks and Servers



Tasks and Servers

Iterative process:
We have n servers and m tasks.
In iteration i, a task is given and we need to choose a server for it.

Goal:
In each iteration, the current task is assigned to a server.
Minimize the maximum load.

Crucial: Task t in iteration i does not know what happened in the past.

Tasks and Servers

Iterative process:
We have n servers and m tasks.
In iteration i, a task is given and we need to choose a server for it.

Goal:
In each iteration, the current task is assigned to a server.
Minimize the maximum load.

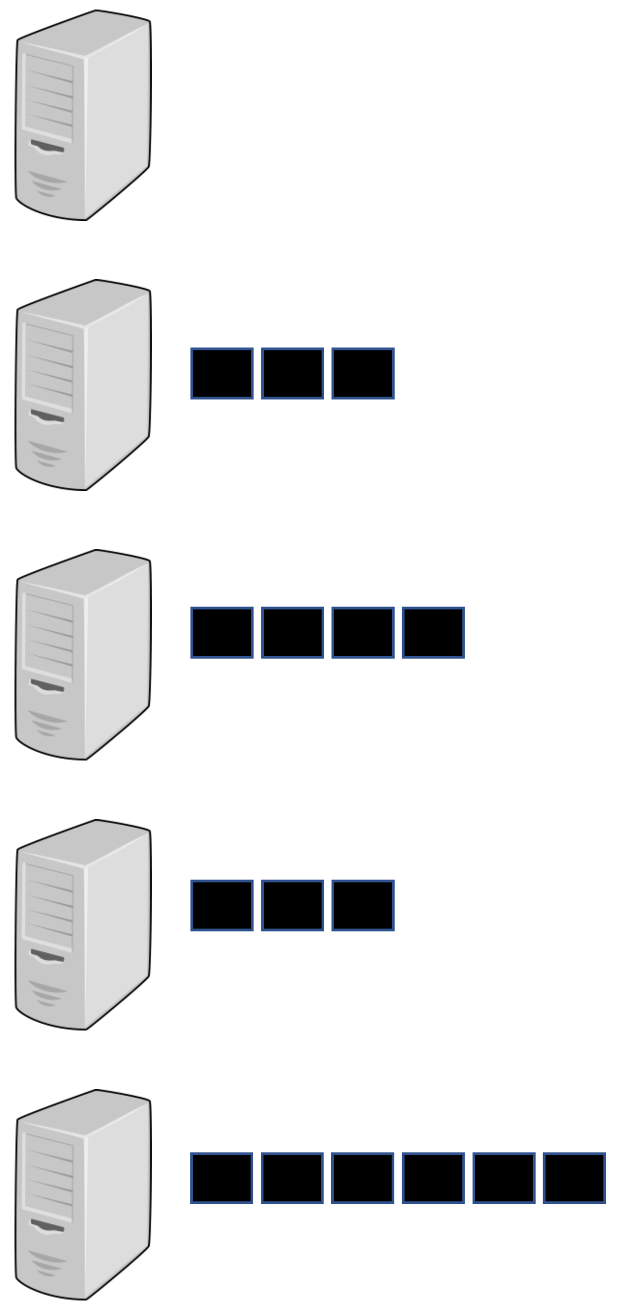
Crucial: Task t in iteration i does not know what happened in the past.

Remark:
A typical setting in online algorithms

Tasks and Servers

Greedy:
Query every server for its load.
Choose minimum.

Either takes $\Omega(n)$ time per task or needs to know the state of the system.

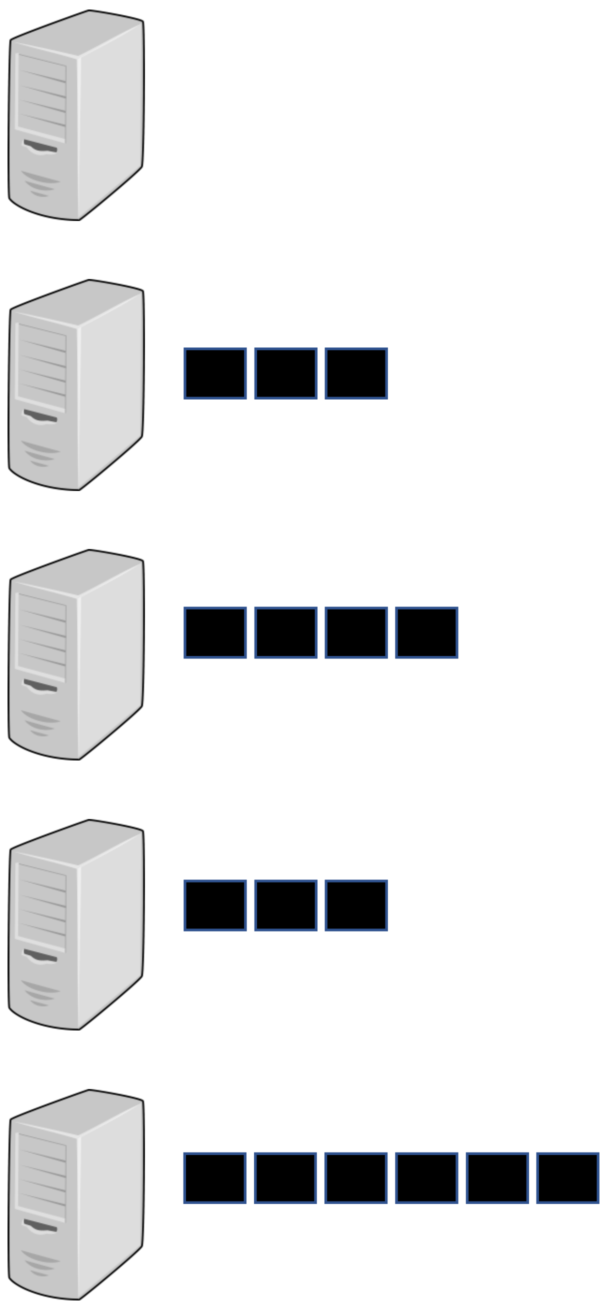


Tasks and Servers

Greedy:
Query every server for its load.
Choose minimum.

Either takes $\Omega(n)$ time per task or needs to know the state of the system.

Uniformly at Random:
Stateless and fast!



Outline

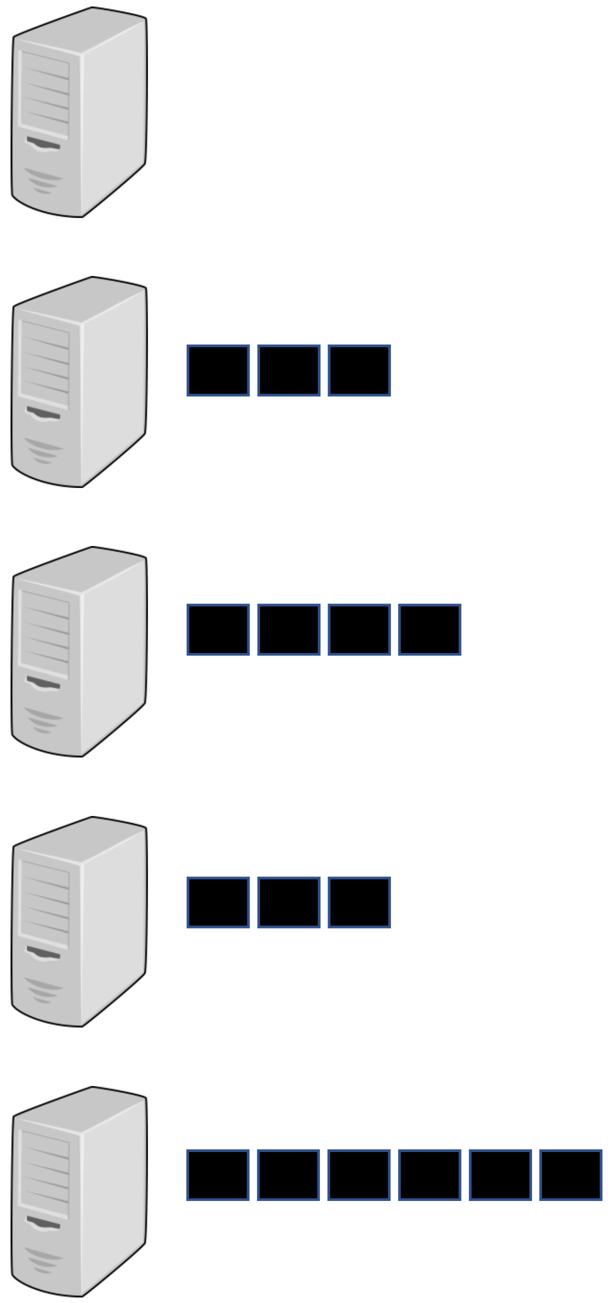
- Load balancing

- Balls-into-Bins
  - Modeling
  - Markov
  - Chebyshev
  - Chernoff

Tasks and Servers

Uniformly at Random:
Stateless and fast!

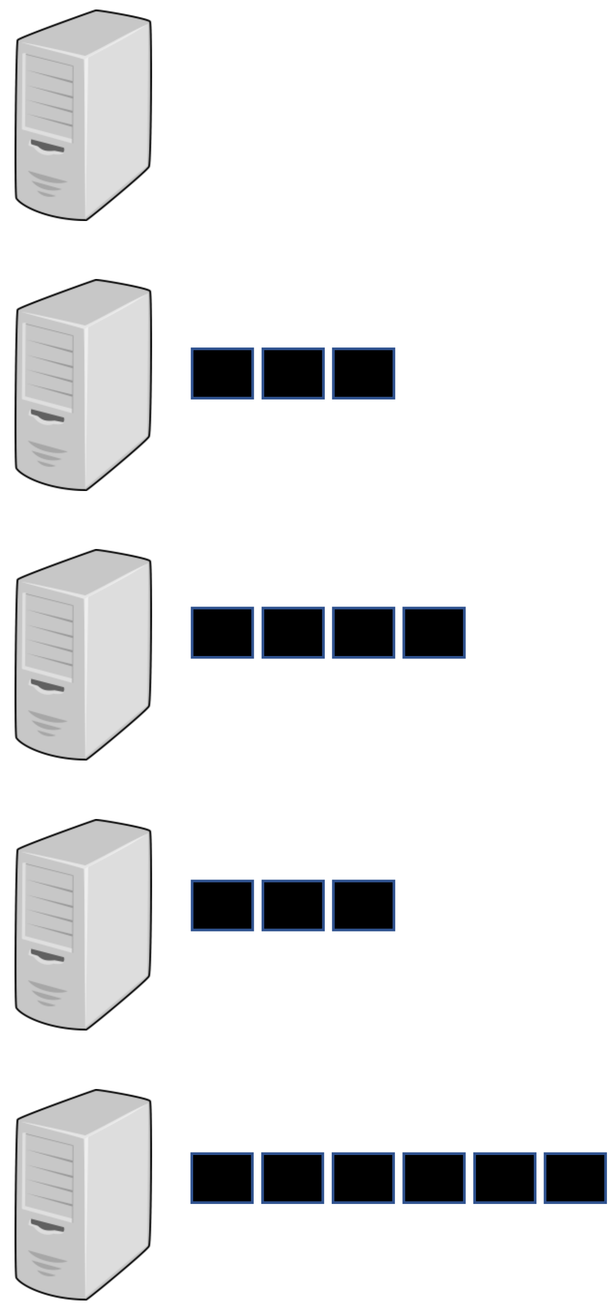
How even is the load?



Tasks and Servers

Markov Inequality:

$$P(X \geq \alpha) \leq \frac{\mu}{\alpha}$$



Tasks and Servers

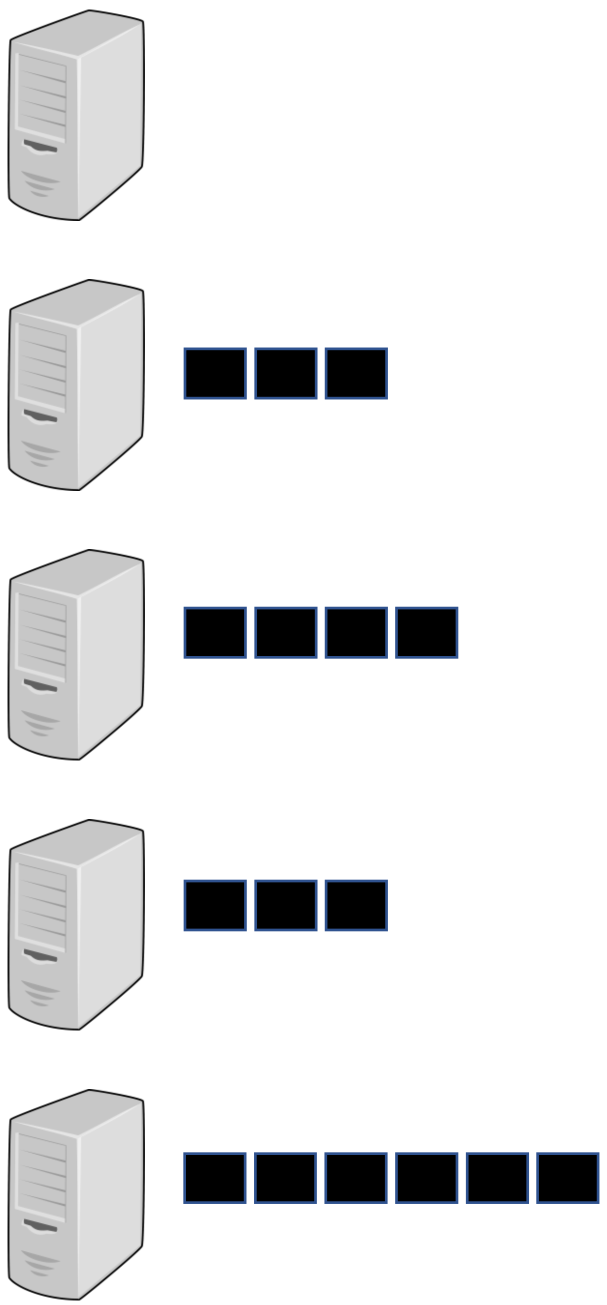
Markov Inequality:

$$P(X \geq \alpha) \leq \frac{\mu}{\alpha}$$

Random variable X counts the number of tasks on a server.

Single Server:

Expectation is $\mu = \Sigma_m \frac{1}{n} = \frac{m}{n}$



Tasks and Servers

Markov Inequality:

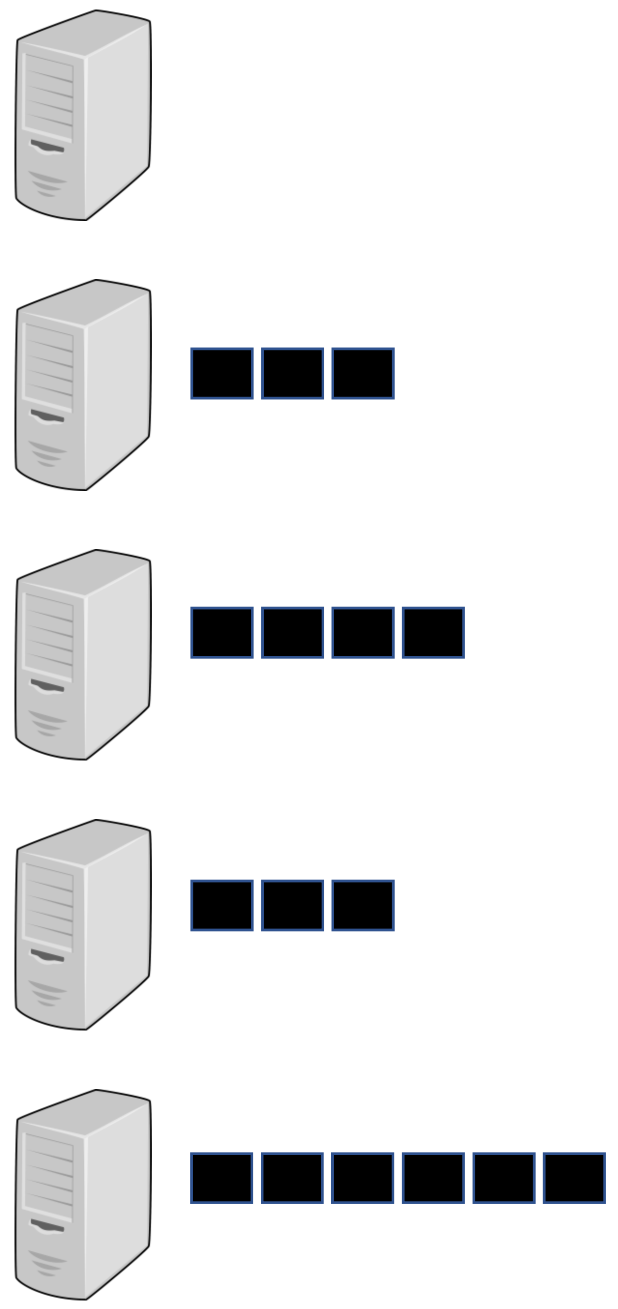
$$P(X \geq \alpha) \leq \frac{\mu}{\alpha}$$

Random variable X counts the number of tasks on a server.

Single Server:

Expectation is $\mu = \Sigma_m \frac{1}{n} = \frac{m}{n}$

Choose one server per task.



Tasks and Servers

Markov Inequality:

$$P(X \geq \alpha) \leq \frac{\mu}{\alpha}$$

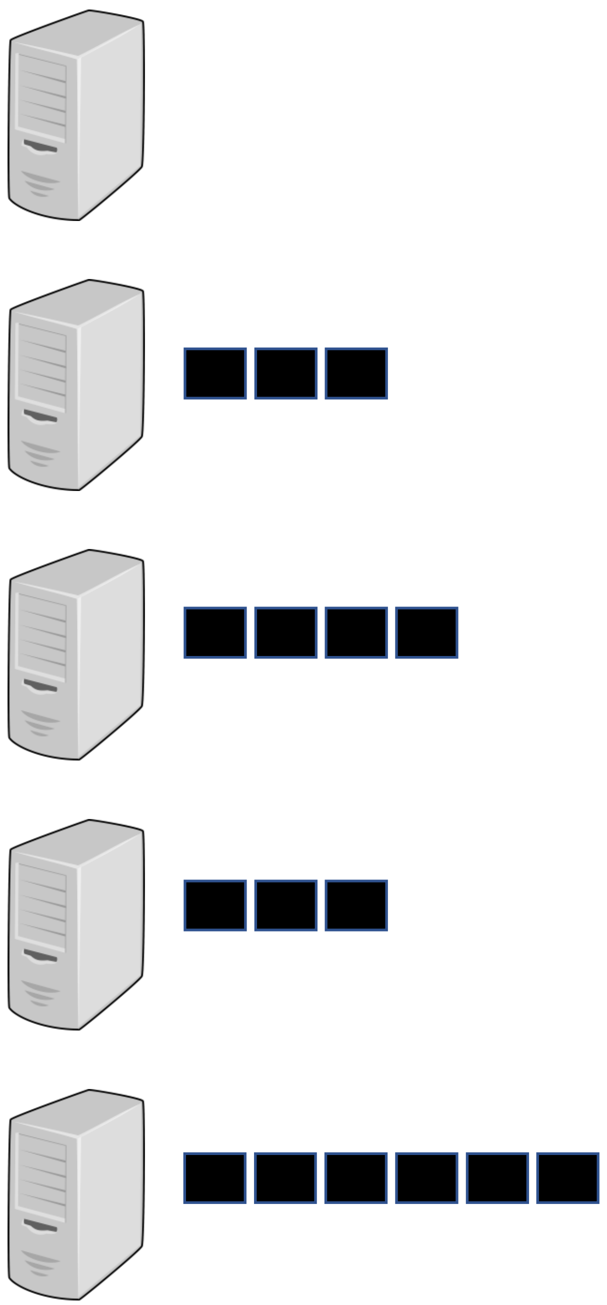
Random variable X counts the number of tasks on a server.

Single Server:

Expectation is $\mu = \Sigma_m \frac{1}{n} = \frac{m}{n}$

Sum over all tasks.

Choose one server per task.



Tasks and Servers

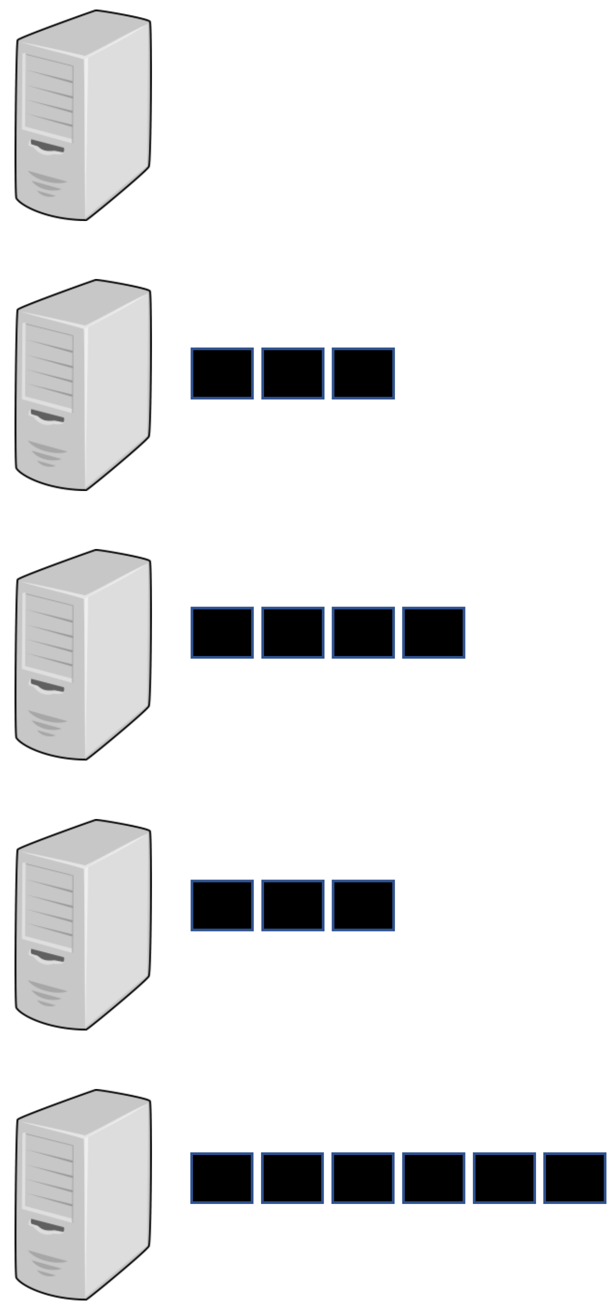
Markov Inequality:

$$P(X \geq \alpha) \leq \frac{\mu}{\alpha}$$

Single Server:

Expectation is $\mu = \Sigma_m \frac{1}{n} = \frac{m}{n}$

$$P\left(X \geq \frac{10m}{n}\right) \leq \frac{1}{10}$$



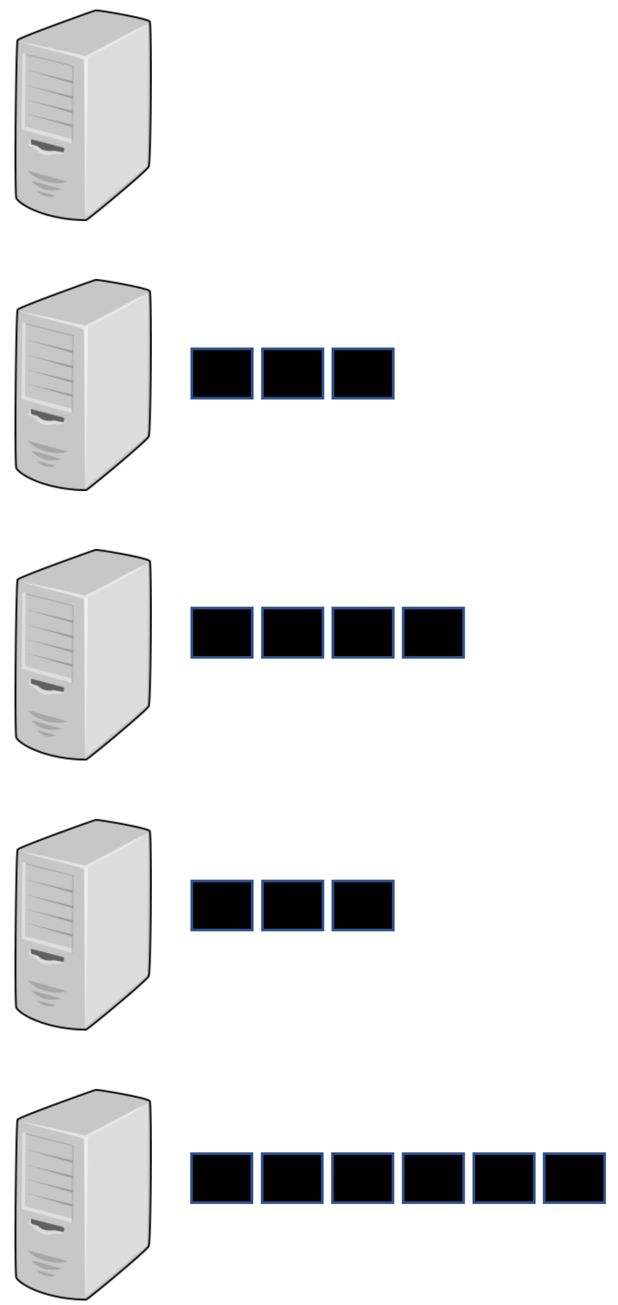
Tasks and Servers

Markov Inequality:

$$P(X \geq \alpha) \leq \frac{\mu}{\alpha}$$

$$P\left(X \geq \frac{10m}{n}\right) \leq \frac{1}{10}$$

The load of a single server is not likely to be high. Does not say much about all servers.



Tasks and Servers

Chebyshev:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Standard deviation

Tasks and Servers

Chebyshev:

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

One ball to fixed bin

$$\mu = \frac{1}{n}$$

Single Server:

Let $X_{i,j}$ be the random variable for ball i hits bin j.

$$\sigma_{i,j}^2 = E\left[\left(X_{i,j}-\mu\right)^2\right]$$

$$= \frac{1}{n}\cdot\left(1-\frac{2}{n}+\frac{1}{n^2}\right)+\left(1-\frac{1}{n}\right)\cdot\left(0-\frac{1}{n}\right)^2 = \frac{1}{n}-\frac{1}{n^2}$$

Tasks and Servers

Chebyshev:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Single Server:

Let $X_{i,j}$ be the random variable for ball i hits bin j.

$$\sigma_{i,j}^2 = E\left[\left(X_{i,j} - \mu\right)^2\right] = \frac{1}{n} - \frac{1}{n^2} < \frac{1}{n}$$

Fact:

$$\sigma_i^2 = \sum_j E\left[\left(X_{i,j} - \mu\right)^2\right] = m \cdot E\left[\left(X_{i,j} - \mu\right)^2\right]$$

Tasks and Servers

Single Server:
Let $X_{i,j}$ be the random variable for ball i hits bin j.

$$\sigma_{i,j}^2 = E\left[\left(X_{i,j} - \mu\right)^2\right] \leq \frac{1}{n}$$

Fact:

$$\sum_j E\left[\left(X_{i,j} - \mu\right)^2\right] = m \cdot E\left[\left(X_{i,j} - \mu\right)^2\right]$$

Chebyshev:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Let $X_i = \Sigma_j X_{i,j}$. Then

$$\sqrt{\frac{m}{n}} \geq \sigma_i$$

Tasks and Servers

Chebyshev:

$$P\left(\left|X_i - \frac{m}{n}\right| \geq k \cdot \sqrt{\frac{m}{n}}\right) \leq P\left(\left|X_i - \frac{m}{n}\right| \geq k \cdot \sigma\right)$$

Chebyshev:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$\mu = \frac{m}{n}, \sqrt{\frac{m}{n}} \geq \sigma$$

Tasks and Servers

Chebyshev:

$$P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2n} \cdot \sqrt{\frac{m}{n}}\right) \leq P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2n} \cdot \sigma\right)$$

Chebyshev:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$\mu = \frac{m}{n}, \sqrt{\frac{m}{n}} \geq \sigma$$

Tasks and Servers

Chebyshev:

$$P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2n} \cdot \sqrt{\frac{m}{n}}\right) \leq P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2n} \cdot \sigma\right)$$

$$\leq \frac{1}{\sqrt{2n}^2} = \frac{1}{2n}$$

Chebyshev:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$\mu = \frac{m}{n}, \sqrt{\frac{m}{n}} \geq \sigma$$

Tasks and Servers

Chebyshev:

$$P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2n} \cdot \sqrt{\frac{m}{n}}\right) \leq P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2n} \cdot \sigma\right)$$

$$\leq \frac{1}{\sqrt{2n}^2} = \frac{1}{2n}$$

Chebyshev:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$\mu = \frac{m}{n}, \sqrt{\frac{m}{n}} \geq \sigma$$

Chebyshev:

$$P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2n} \cdot \sqrt{\frac{m}{n}}\right) = P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2m}\right) \leq \frac{1}{2n}$$

Tasks and Servers

Chebyshev:

$$P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2m}\right) \leq \frac{1}{2n}$$

Union bound:

The probability that the load of any server is more than $\frac{m}{n} + \sqrt{2m} = \left(1 + \frac{\sqrt{2}n}{\sqrt{m}}\right) \cdot \frac{m}{n}$ is at most

$$\sum_{i=1,\ldots,n} P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2m}\right) \leq \frac{n}{2n} = \frac{1}{2}$$

Tasks and Servers

Chebyshev:

$$P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2m}\right) \leq \frac{1}{2n}$$

Remark:
Not great for small m.
Becomes better as m grows.

Union bound:
The probability that the load of any server is more than $\frac{m}{n} + \sqrt{2m} = \left(1 + \frac{\sqrt{2}n}{\sqrt{m}}\right) \cdot \frac{m}{n}$ is at most

$$\sum_{i=1,\ldots,n} P\left(\left|X_i - \frac{m}{n}\right| \geq \sqrt{2m}\right) \leq \frac{n}{2n} = \frac{1}{2}$$

Tasks and Servers

Chernoff:

$$P(X \geq (1+\delta) \cdot \mu) \leq 2^{\left(-\frac{\mu\delta^2}{3}\right)}$$

$$X_i = \Sigma_{\mathrm{j}}\, X_{i,j}$$

$$\delta > 0$$

Tasks and Servers

Chernoff:

$$P(X \geq (1+\delta)\cdot\mu) \leq 2^{\left(-\frac{\mu\delta^2}{3}\right)}$$

$$X_i = \Sigma_{\mathrm{j}}\, X_{i,j}$$

Balls are independent. Suppose $m = n$:

$$E[X_i] = \mu = 1$$

Tasks and Servers

Chernoff:

$$P(X \geq (1+\delta)\cdot\mu) \leq 2^{\left(-\frac{\mu\delta^2}{3}\right)}$$

$X_i = \Sigma_{\mathrm{j}}\, X_{i,j}$

Balls are independent. Suppose $m = n$:

$$E[X_i] = \mu = 1$$

Single bin. Set $\delta = 3\sqrt{\log n}$

$$P(X_i \geq (1+\delta)\cdot\mu) \leq 2^{\left(-\frac{\mu\delta^2}{3}\right)} \leq 2^{-3\cdot\sqrt{\log n}^2} = n^{-3}$$

Tasks and Servers

Chernoff:

$$P(X \geq (1+\delta)\cdot\mu) \leq 2^{\left(-\frac{\mu\delta^2}{3}\right)}$$

$$P\left(X_i \geq \left(1 + 3\cdot\sqrt{\log n}\right)\cdot\mu\right) \leq n^{-3}$$

Union bound (all bins)**:**

At most $3\cdot\sqrt{\log n}$ balls w.p. $1 - n^{-2}$

Tasks and Servers

Since μ is in the exponent, this bound also gets much stronger when $m \gg n \log n$

Chernoff:

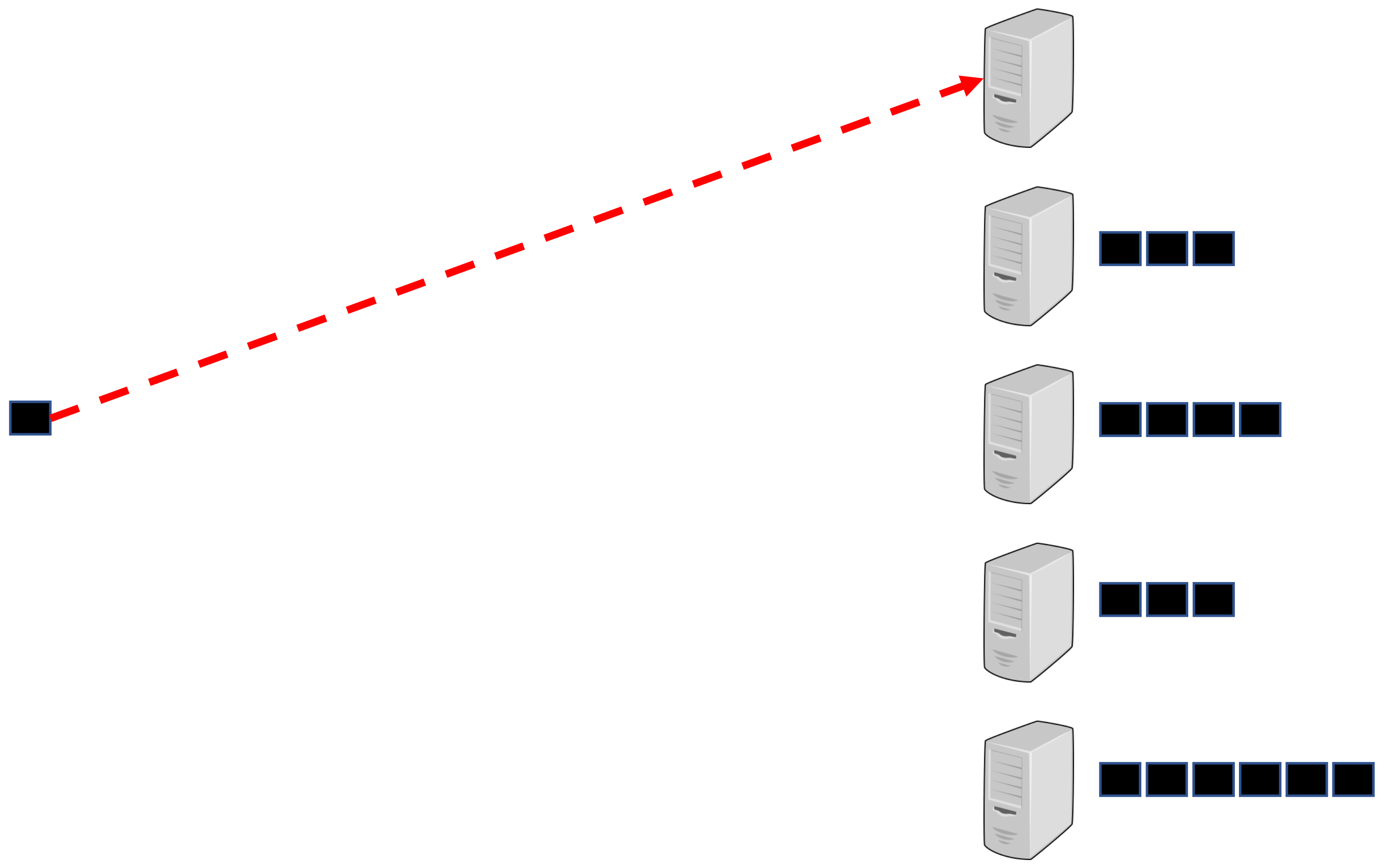
$$P(X \geq (1+\delta) \cdot \mu) \leq 2^{\left(-\frac{\mu\delta^2}{3}\right)}$$

$$P\left(X_i \geq \left(1 + 3 \cdot \sqrt{\log n}\right) \cdot \mu\right) \leq n^{-3}$$

Union bound (all bins):

At most $3 \cdot \sqrt{\log n}$ balls w.p. $1 - n^{-2}$

Wrap up



Balls to Bins

For the case of $m = n$

1. Markov: One bin is unlikely to have many balls

2. Chebyshev: No bin is likely to have $\omega(\sqrt{n})$ balls.

3. Chernoff: W.h.p. max #balls is $O\left(\sqrt{\log n}\right)$