1. Arrival

(3 parts) ← Queuing model

$\sqcup / \sqcup / \sqcup$

arrival rate

general service time (e.g task size)

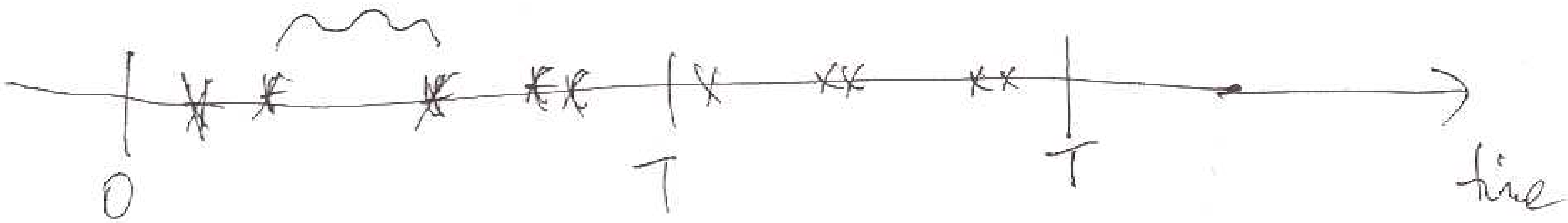
Servers

if # Servers $= \infty$, then there is <u>no</u> pb as we have infinite capacity

Arrival $\equiv$ follows Poison distribution

L6-1

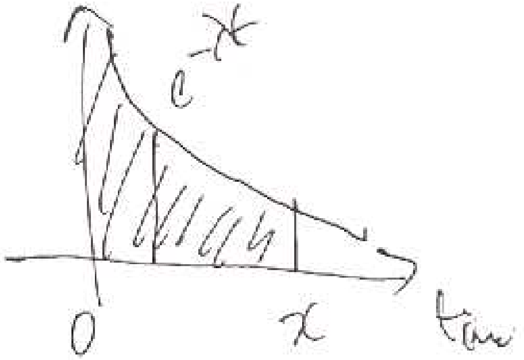
x = time between 2 events



period (15mn)

$$P(\#(T)=k) = \frac{e^{-\lambda T}(\lambda T)^k}{k}$$

$$P(X<x) = \lambda \int_0^x e^{-\lambda t}\,dt$$



Probability that the distance between 2 events is less than x

exponential Distribution

$x \uparrow \longrightarrow P(X>x) = 0$

LC-2

Arrival rate λ

$P(X \geq x) \simeq 0$

$x \uparrow$

exponential distribution of arrival rate = ~~a Poisson distribution~~

$\simeq$ " stable system " in terms of tasks arriving to the system

2 Task size distribution

how the task size varies

Exponential distribution (≃ nearly similar size for most of the tasks)

Pareto Distribution (there high variation in task sizes

LG-3

e.g. classical systems, like OS

$P(X > x)$

probability that a certain task has more x in size

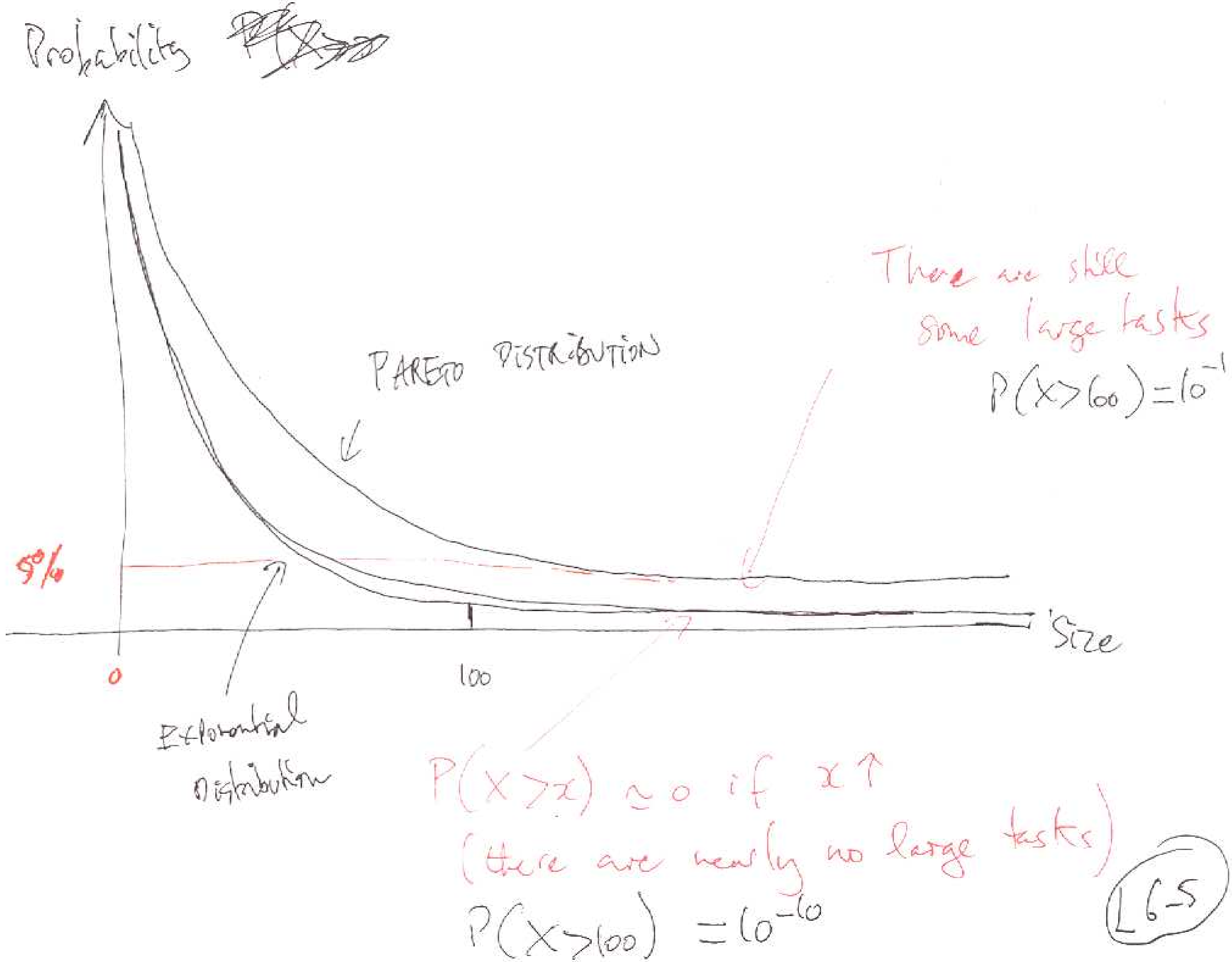
if $x \uparrow$ (big) $P(X > x) \sim 0$

e.g. Web Servers

size

10-20% of large tasks taking around 80% CPU

$P(X > x) \neq 0$


Probability
PARETO DISTRIBUTION
There are still some large tasks
P(X>100) = 10^-1
5%
0
100
Size
Exponential distribution
P(X>x) ≃ 0 if x ↑
(there are nearly no large tasks)
P(X>100) = 10^-10

L6-5

M/M/1 — exponential for both arrival/task size distribution

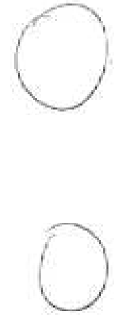
M/G/1 — exponential for arrival (distance between two events)

pareto distribution for task size

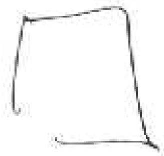
Periodical UPDATES

S1

S2



Dispatcher



Load index

S1	5
S2	10

every period T, S1 and S2 send information to the dispatcher

L6-6

$S_0 < S_1 < \cdots < S_n$

$\underbrace{S_0 \quad S_1 \quad \cdots}_{} \quad S_{i-1} \quad S_i \quad S_{i+1} \quad \cdots \quad S_n$

$i=1$

K-subset

INSTEAD OF

WORKING WITH "n" SERVERS

order them from the least loaded to most loaded.

P_i = probability a given request is assigned to S_i

$$\frac{\binom{n-i-1}{k-1}}{\binom{n-1}{k-1}}$$

\# of $(k-1)$ sets from $n-i-1$ elements

give you the length of the set

a b c

$3 = \binom{3}{2} =$ (a,b) (b,c) (a,c)

L6-7

What we want is to select ONLY K nodes out of n existing ones

already selected as an element of K-subset

$S_0 \quad S_2 \quad \cdots \quad S_{i-1} \quad S_i \quad \cdots$

$\binom{n-1}{k-1}$ = number of possible $(k-1)$ sets we can construct from $n-1$ element

$n = 5 \qquad K = 3 \qquad$ K-set = 3-set = set of 3 elements

$1 \quad 2 \quad 3 \quad 4 \quad 5$

$\binom{n-1}{K-1} = \binom{5-1}{3-1} = \binom{4}{2} = \left\{ \begin{array}{ccc} (12\,3) & (14\,3) & (15\,3) \\ (24\,3) & (25\,3) & (45\,3) \end{array} \right.$

$= 6$

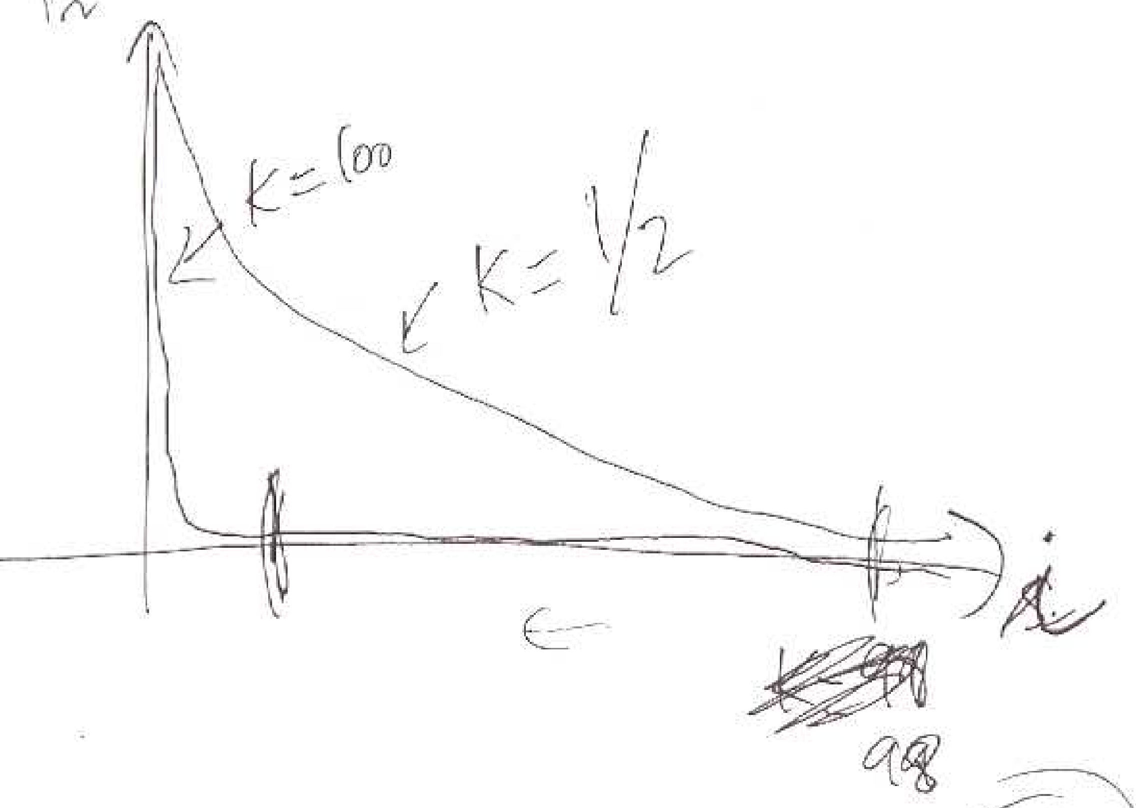
L6-8

$$P_i = \frac{\binom{99-i}{K-1}}{\binom{99}{K-1}} \qquad \frac{\binom{99-i}{99}}{\binom{99}{99}} = 1$$

$K \uparrow$ $K \approx 100$

$K \downarrow_{=2}$ (K smaller)

$$P_i = \frac{\binom{99-i}{1}}{\binom{99}{1}} = 99$$



169

$$\frac{L_{total} + arrive_T}{n} - L_i$$ ← previous load at server i

total number of tasks arriving ~~at~~ during period T

i

○ S1 L1

→ ○ S2 L2

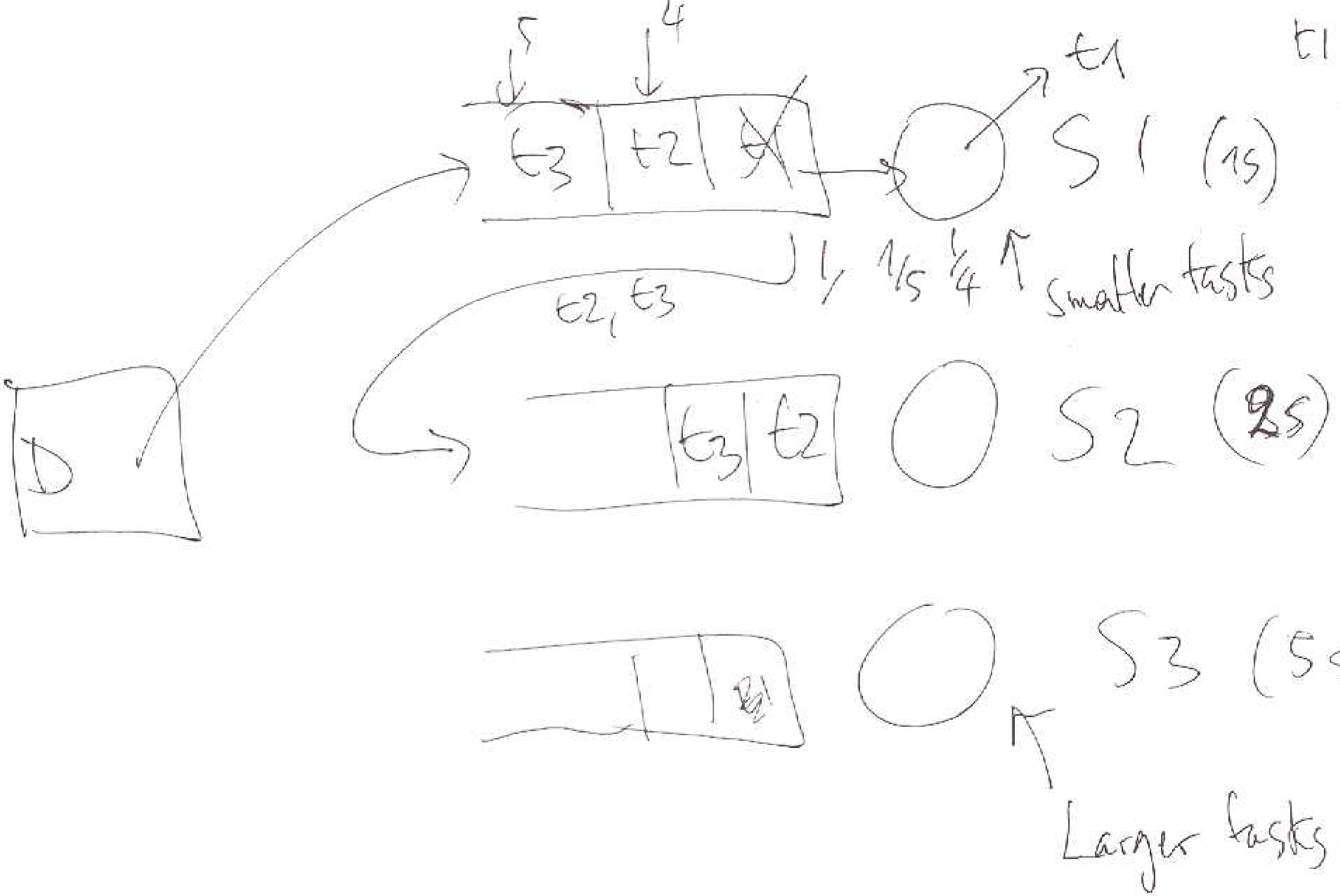
○ S3 L3

now

$$\begin{array}{ccc} S1 & S2 & S3 \\ \circ & \circ & \circ \end{array} \xrightarrow{T} \begin{array}{ccc} S1 & S2 & S3 \\ \circ & \circ & \circ \end{array}$$

$$L1 + L2 + L3 \qquad \frac{(L1 + L2 + L3) + arrive_T}{3}$$

L6-L0



$t1 < t2 < t3$

smaller task size

LG-11