

# PARAMS

- $k_i$  instance performance indicator
- $k_s$  average service performance indicator
- $z_i = k_i / k_s$  instance performance ratio
- $s$  instance shutdown threshold
- $I_o$  set of instances
- $n = |I_o|$

# OBJ FUNC

$$\min \sum_{i \in I_o} \underbrace{\frac{1}{z_i}}_{\text{Se } i \text{ è buona}} \underbrace{w_i - z_i a_i}_{\text{Se l'istanza resta a caso conviene, scelto per } z_i}$$

Se  $i$  è buona  
 $z_i > 1$   
quindi conviene  
avere  $w_i$  alto

Se l'istanza  
resta a caso  
conviene,  
scelto per  
 $z_i$

# CONSTRAINT

$$1 \quad w_i \leq q_i$$

$$2 \quad w_i \geq \frac{s}{n} q_i$$

$$3 \quad \sum_{i \in I_0} w_i = 1$$

$$4 \quad w_i \leq \frac{k_i}{k_j} w_j + (1 - q_j) \quad k_i \geq k_j$$

$$5 \quad w_i \leq z_i p_i + z_i \sum_{i \neq j} p_j (1 - q_j)$$

① To impose that an instance to shutdown has  $W_i = 0$

② To impose that an active instance should have a  $W_i$  higher than a certain threshold  $S$ .

$S$  is the ratio between the # of requests processed by the instance and the # of requests that an instance should process in an ideal case (when the load is distributed equally)

L'idea è "seguire le istanze che si allontanano molto dal caso ideale. Se è così, è perché hanno performance molto basse".

Quindi ( $r_i \rightarrow$  # richieste istanza  $i$ )

$$S = \frac{r_i}{\sum r_i / n} \rightarrow \text{caso ideale}$$

Ma  $S$  non va calcolato, va visto come percentuale. Tipo,  $S = 0,8$  indica che l'istanza  $i$  è lontana dal caso ideale del 20%

Ma  $\frac{r_i}{\sum r_i} = W_i$ , quindi il  $W_i$  segue  $\bar{e} = S \cdot n$

Quindi vogliamo  $W_i \geq S/n$

③ The weights should have unitary sum

④ We want that instance weights are proportional to their performances, i.e. that  $W_i = \frac{k_i}{k_j} W_j$ .

However, if  $W_j = 0$ , it doesn't mean that  $W_i$  must be  $= 0$  if  $k_i > k_j$

Hence, we want

$W_i \leq \frac{k_i}{k_j} W_j + (1 - \alpha_j)$  if  $k_i \geq k_j$ , so that if  $\alpha_j = 0$ ,  
 $W_i$  can be for sure  $\leq 1$

⑤ We want that a weight should not grow too much with respect to its previous weight if there are no instances to shutdown. If there are, constraint ④ will take care of splitting the load fairly