Προχωρημένη Κατανεμημένη Υπολογιστική

ΗΥ623

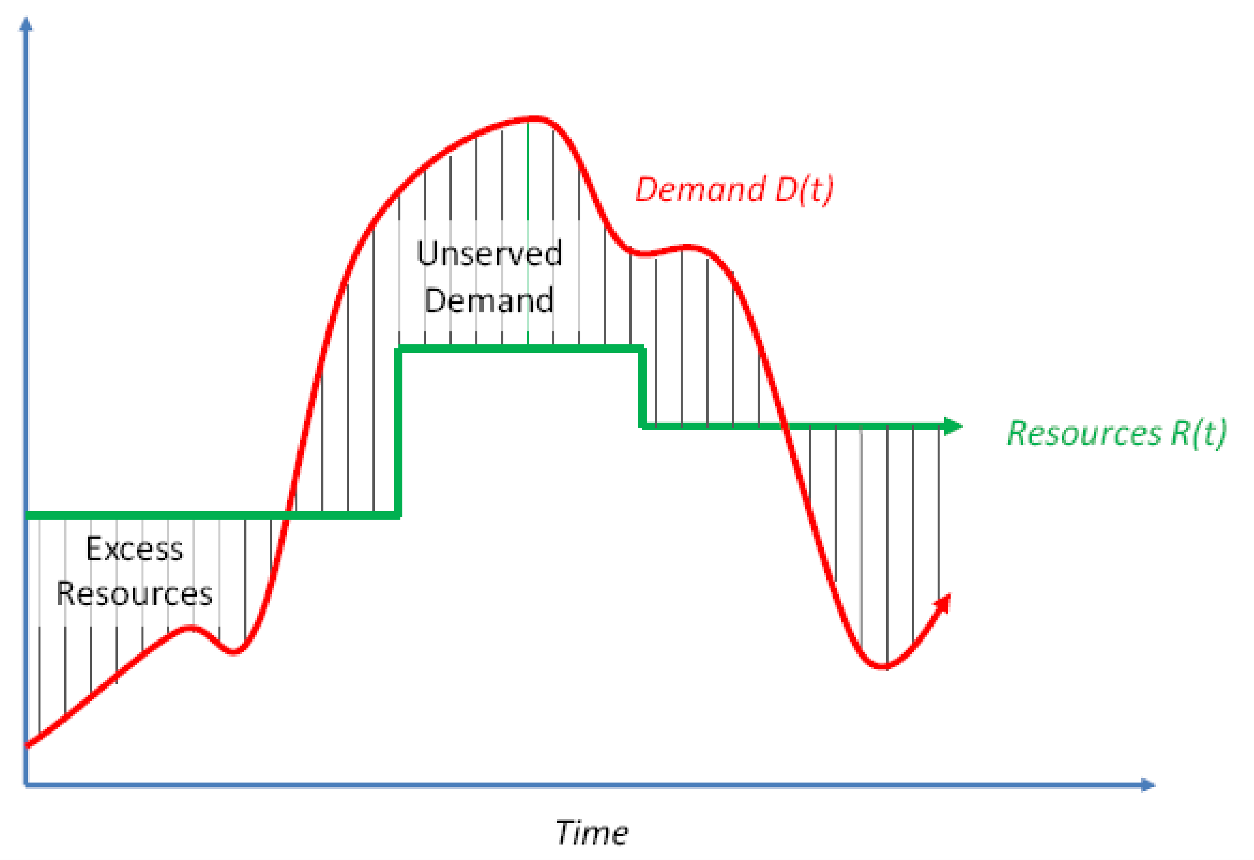
Διδάσκων –
Δημήτριος Κατσαρός

@ Τμ. ΗΜΜΥ
Πανεπιστήμιο Θεσσαλίας

Διάλεξη 9η

The cost of elasticity

Imperfect capacity



Costs & perfect capacity

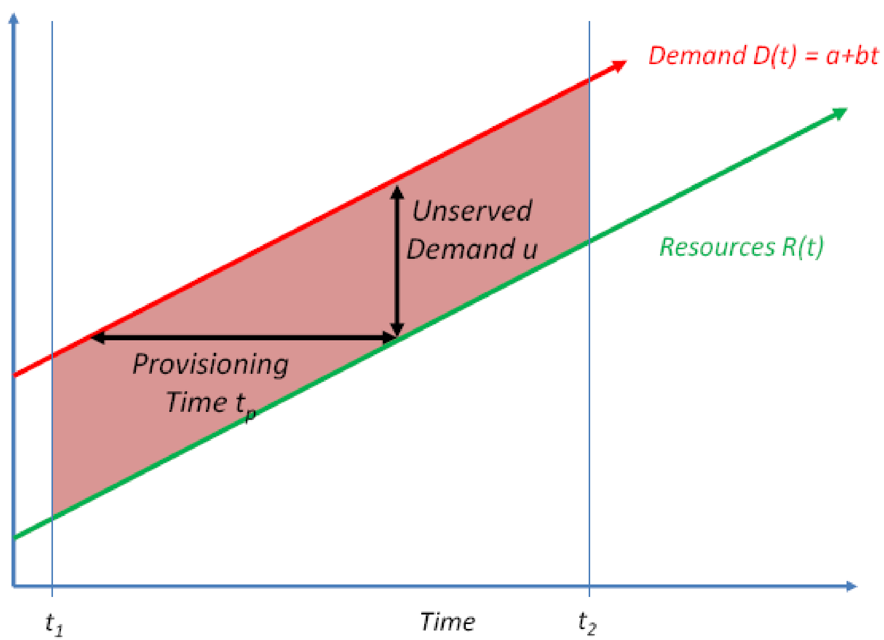
- The cost of r unemployed resources for a duration of time t is: $\mathbf{c_r \times r \times t}$
- The cost of r insufficient resources is: $\mathbf{c_d \times r \times t}$

- The **Loss function**:

$$L = \int_{t_1}^{t_2} [D(t) - R(t)] \times c_d \, dt \mid D(t) > R(t) \; + \int_{t_1}^{t_2} [R(t) - D(t)] \times c_r \, dt \mid R(t) > D(t)$$

- A **perfect capacity** strategy R, where R=D(t) ∀t, has a loss of 0
  - Trivially solved for **constant demand**

Linearly Increasing Demand, No Forecasting, Continuous Monitoring, Non-Zero Provisioning Interval



Linearly Increasing Demand with Continuous Monitoring And Non-Zero Provisioning Interval

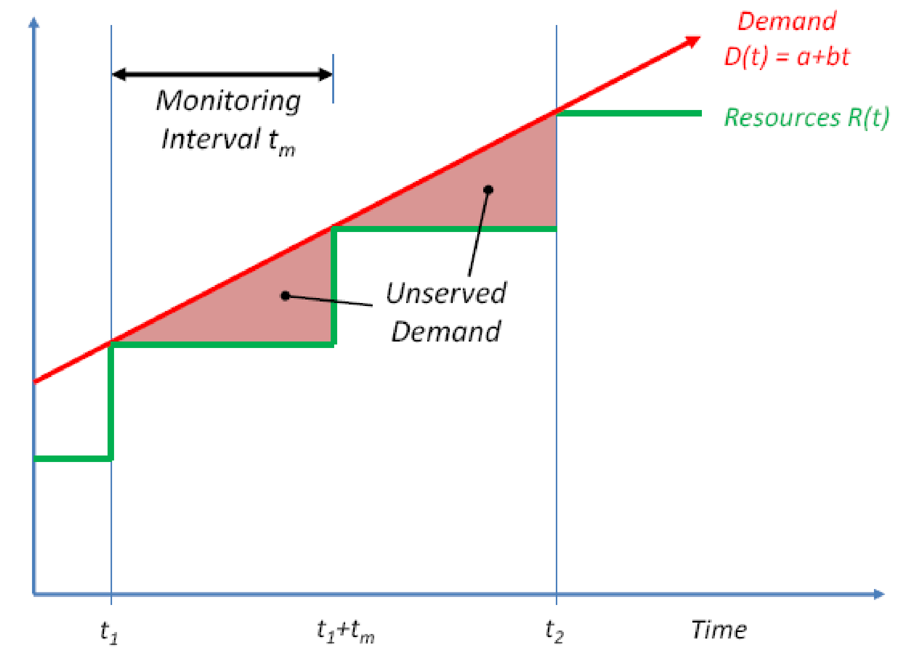
the loss L is proportional to the provisioning delay t_p

$$L = \int_{t_1}^{t_2} [D(t) - R(t)] \times c_d \, dt \mid D(t) > R(t) \; + \; \int_{t_1}^{t_2} [R(t) - D(t)] \times c_r \, dt \mid R(t) > D(t)$$

since demand is always greater than resources simplifies to

$$L = \int_{t_1}^{t_2} [D(t) - R(t)] \times c_d \, dt \longrightarrow L = \int_{t_1}^{t_2} b \times t_p \times c_d \, dt = (t_2 - t_1) \times b \times t_p \times c_d$$

Linearly Increasing Demand, No Forecasting, Periodic Monitoring, On-Demand Provisioning



Linearly Increasing Demand with Periodic Monitoring And On-Demand Provisioning

In an **on-demand environment, the loss is proportional to the monitoring interval**.

If the monitoring interval drops to zero, that is, there is continuous monitoring with on-demand provisioning, the loss drops to zero as well.

If the slope drops to zero, then the loss is zero as we showed in the flat demand case earlier, regardless of the monitoring interval

$$L = \int_{t_1}^{t_2} [D(t) - R(t)] \times c_d \, dt \mid D(t) > R(t) \; + \; \int_{t_1}^{t_2} [R(t) - D(t)] \times c_r \, dt \mid R(t) > D(t)$$

Since R(t) is strictly not greater than D(t) this reduces to:

$$L = \int_{t_1}^{t_2} [D(t) - R(t)] \times c_d \, dt \mid D(t) > R(t)$$

$$L = k \times ½\, b \times t_m^2 \times c_d = \frac{(t_2 - t_1)}{t_m} \times \frac{1}{2} b \times t_m^2 \times c_d = \frac{(t_2 - t_1)}{2} \times b \times t_m \times c_d$$

Exercise 1

- If the demands grows exponentially, the cloud performs continuous monitoring, but no forecasting, and it needs t_p time to offer the respective resources, then show that the cost grows "unbounded". c_d is the cost of underprovisioning.

- Solution
  - We need to estimate $\int_0^\infty (e^t - e^{t-t_p})dt$
  - This evaluates to: $(e^t\text{-}1)(1 - \frac{1}{e^{t_p}})c_d$
    - Only when $t_p \rightarrow 0$, the above equation is "somewhat" bounded

Exercise 2

- If the demands is random between 0 and a max value P, calculate the optimal level of offerred resources so as to minimize the total cost (underprovisioning and overprovisioning). c_d is the cost of underprovisioning, and c_r is the cost of overprovisioning
- Solution
  - Let us set this level equal to r
  - Then, totalCost= $\frac{r}{P} \times (r - \frac{r}{2}) \times c_r + \frac{P-r}{P} \times (\frac{P+r}{2} - r) \times c_d$
  - min: $\frac{\partial totalCost}{\partial r} = 0 \ldots \rightarrow r = \frac{c_d}{c_d+c_r} \times P$
    - When c_d=c_r=c, then r = $\frac{P}{2}$