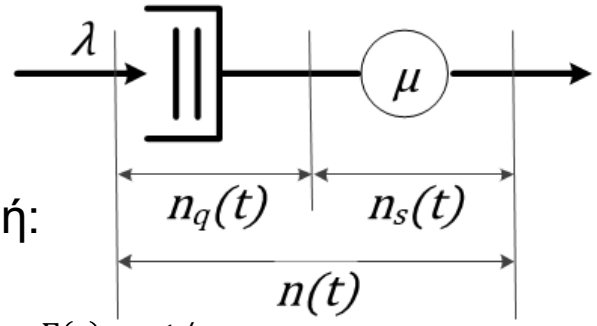
ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ
Queuing Systems
Παράμετροι Ουρών Αναμονής

Βασίλης Μάγκλαρης
maglaris@netmode.ntua.gr

13/3/2019

ΠΑΡΑΜΕΤΡΟΙ (1/3)

$$\gamma = \lambda(1 - \mathrm{P}\{\text{blocking}\}) \leq \lambda, \quad \gamma < \mu$$



- **Ένταση φορτίου (traffic intensity)**
  Σε περίπτωση 1 ουράς, 1 εξυπηρετητή:

$$\rho \triangleq \frac{\text{Μέσος Χρόνος εξυπηρέτησης}}{\text{Μέσος Χρόνος μεταξύ αφίξεων}} = \frac{\mathrm{E}(s)}{\mathrm{E}(a)} = \frac{1/\mu}{1/\lambda} = \lambda \mathrm{E}(s) = \lambda/\mu$$ ***(Erlangs)***

Ένα ***Erlang*** αντιπροσωπεύει το φόρτο κυκλοφορίας που εξυπηρετείται από έναν εξυπηρετητή που ασχολείται το 100% του χρόνου (π.χ. 1 call-minute per minute). Ένας εξυπηρετητής ασχολείται για 30 λεπτά σε μια περίοδο μιας ώρας → μεταφέρει 0.5 Erlangs κυκλοφοριακή ένταση

- **Διεκπεραίωση πελατών – Ρυθμαπόδοση (Throughput)** γ *(πελάτες/sec)*
  Σε περίπτωση 1 ουράς, 1 εξυπηρετητή:

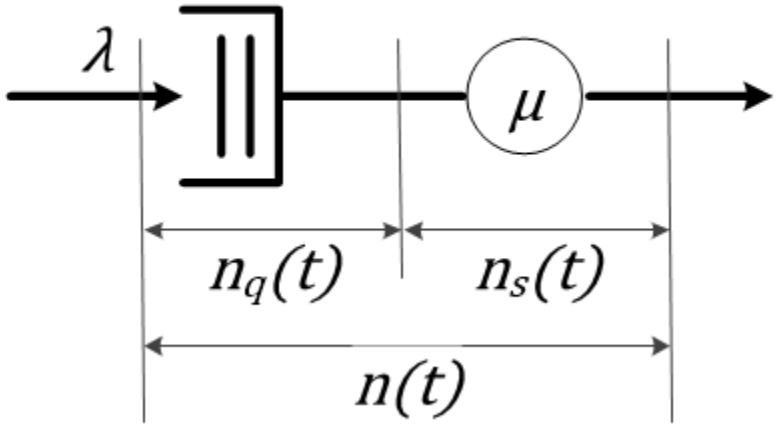
$$\gamma = \lambda(1 - \mathrm{P}\{\text{blocking}\}) \leq \lambda, \qquad \gamma < \mu$$

όπου $\mathrm{P}\{\text{blocking}\}$ είναι η πιθανότητα να χαθεί ένας πελάτης επειδή βρήκε το σύστημα πλήρες

  - Σε τηλεφωνικά δίκτυα: βαθμός υπηρεσίας, **Grade of Service** - **GoS**
  - Σε δίκτυα δεδομένων: ποιότητα υπηρεσίας, **Quality of Service** - **QoS**

ΠΑΡΑΜΕΤΡΟΙ (2/3)

$$\gamma = \lambda(1 - \mathrm{P}\{\text{blocking}\}) \leq \lambda, \quad \gamma < \mu$$



- **Μέσος ρυθμός απωλειών, ποσοστό απωλειών, πιθανότητα απώλειας πελάτη**
  - Σε περίπτωση 1 ουράς, 1 εξυπηρετητή

    Μέσος ρυθμός απωλειών: $\lambda - \gamma$

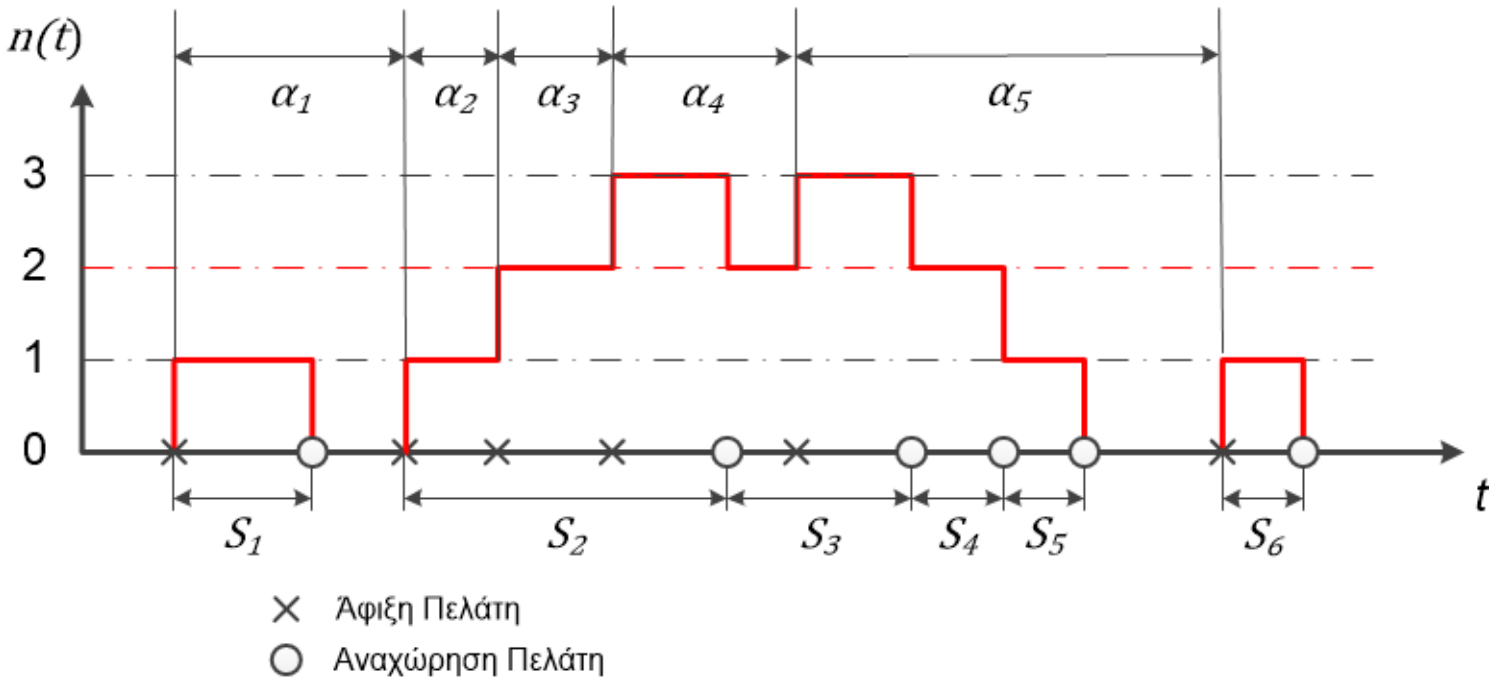
    Ποσοστό απωλειών: $\frac{\lambda - \gamma}{\lambda} = \mathrm{P}\{\text{blocking}\}$

- **Βαθμός χρησιμοποίησης εξυπηρετητή (server utilization)**
  - Σε περίπτωση 1 ουράς, 1 εξυπηρετητή

$$u \triangleq \gamma/\mu \leq \lambda/\mu = \rho$$

ΠΑΡΑΜΕΤΡΟΙ (3/3)

Εξέλιξη Αριθμού Πελατών στο Σύστημα



- **Αριθμός πελατών (κατάσταση)**

  $n(t)$, στοχαστική ανέλιξη – χρονοσειρά (stochastic process, time series)

- **Μέσος αριθμός πελατών** $\mathrm{E}\{n(t)\}$
- **Μέσος χρόνος καθυστέρησης (average time delay)**

  Μέσος χρόνος αναμονής (waiting time) + Μέσος χρόνος εξυπηρέτησης

$$\mathrm{E}(T) = \mathrm{E}(W) + \mathrm{E}(s)$$

ΚΑΤΑΣΤΑΣΗ ΣΥΣΤΗΜΑΤΟΣ

- $n(t) = 0,1,2,\dots,K$: Τυχαία μεταβλητή που ορίζει την **κατάσταση** του Συστήματος Αναμονής την χρονική στιγμή t . Η τυχαία συνάρτηση $n(t)$ αποτελεί **στοχαστική ανέλιξη** (διαδικασία) διακριτής κατάστασης με μεταβάσεις καταστάσεων σε συνεχή χρόνο (**discrete state, continuous time stochastic process**)

  $n(t) = n_q(t) + n_s(t) \leq K$ όπου:

  K η μέγιστη χωρητικότητα συστήματος

  $n_q(t) = 0,1,2,\dots,K-1$ ο αριθμός πελατών σε αναμονή

  $n_s(t) = 0,1$ ο αριθμός πελατών στην εξυπηρέτηση

- $\mathrm{P}_k(t) \triangleq \mathrm{P}\{n(t) = k\}$: Η πιθανότητα παρουσίας k πελατών (σε αναμονή και εξυπηρέτηση) τη χρονική στιγμή t

ΙΣΟΡΡΟΠΙΑ, ΕΡΓΟΔΙΚΕΣ ΚΑΤΑΣΤΑΣΕΙΣ ΣΥΣΤΗΜΑΤΟΣ

- **ΟΡΙΣΜΟΣ ΙΣΟΡΡΟΠΙΑΣ**: Αν μια στοχαστική ανέλιξη $n(t)$ **ισορροπήσει** μετά από παρέλευση μεγάλου χρονικού διαστήματος t, το μεταβατικό φαινόμενο παρέρχεται και το σύστημα παλινδρομεί τυχαία ανάμεσα σε απείρως επισκέψιμες (**γνησίως επαναληπτικές**, **positive recurrent**) καταστάσεις $n(t) = k.$ Οι $\mathrm{P}_k(t)$ συγκλίνουν σε σταθερές τιμές $P_k > 0$ ανεξάρτητες της αρχικής κατάστασης $n(0)$

- **ΠΡΟΣΟΧΗ:** *Οι στοχαστικές ανελίξεις δεν ισορροπούν υποχρεωτικά, μόνο κάτω από ειδικές συνθήκες όπως αυτές των καλοσχεδιασμένων συστημάτων αναμονής*

- Οι απείρως επισκέψιμες καταστάσεις $n(t) = k$ συστήματος σε **ισορροπία** αποκαλούνται **εργοδικές καταστάσεις**

- Σύστημα σε ισορροπία: $\lim_{t\to\infty} \mathrm{P}_k(t) = P_k > 0$ (**εργοδικές οριακές πιθανότητες**)

  $P_k = \lim_{\mathrm{T}\to\infty} \frac{\mathrm{T}_k}{\mathrm{T}} > 0$ όπου T_k ο συνολικός χρόνος στη κατάσταση $n(t) = k$ στη διάρκεια T **μιας** παρατήρησης (εξέλιξης) της ανέλιξης $n(t)$

- Εργοδικοί Μέσοι Όροι των $n(t), n_q(t), n_s(t)$ συστήματος σε ισορροπία:

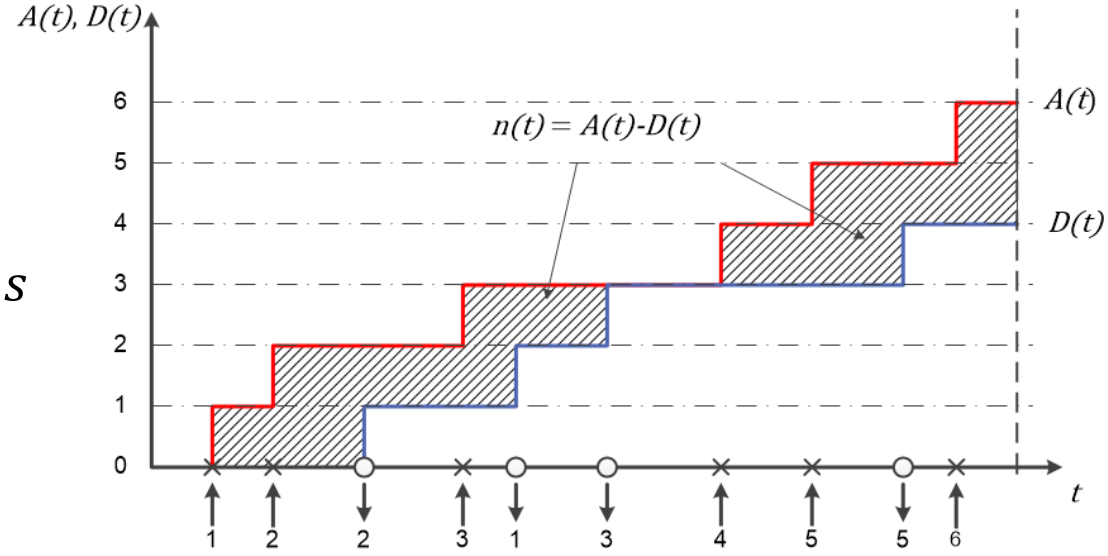
$$\mathrm{E}\{n(t)\} = \mathrm{E}\{n_q(t)\} + \mathrm{E}\{n_s(t)\}, \forall t$$

- Εργοδικοί Μέσοι Χρόνοι Καθυστέρησης (αναμονή + εξυπηρέτηση) συστήματος σε ισορροπία:

$$T = W + s, \mathrm{E}(T) = \mathrm{E}(W) + \mathrm{E}(s)$$

ΤΥΠΟΣ Little

(Σύστημα σε Ισορροπία)

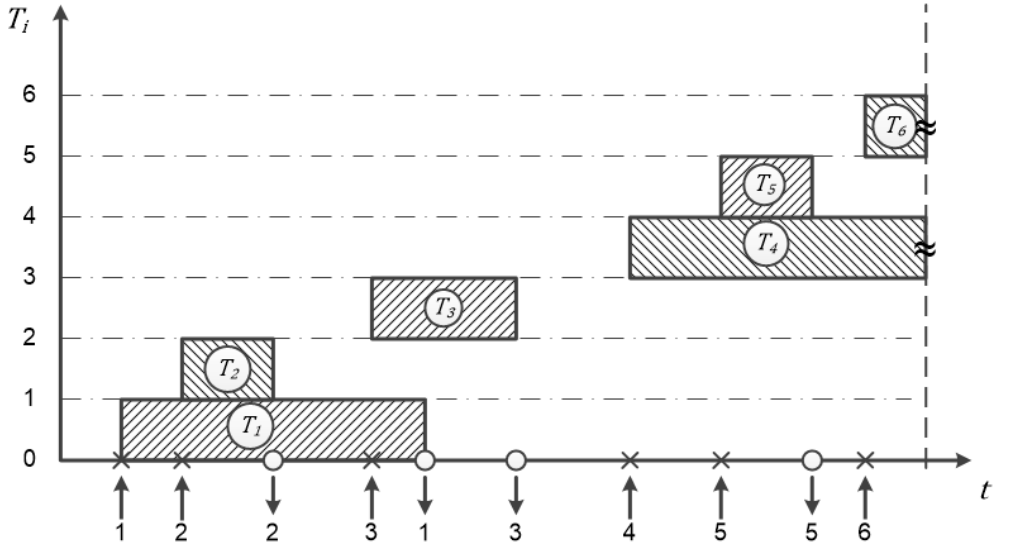


Χρόνος καθυστέρησης: $T = W + s$

Τύπος Little:

$$\mathrm{E}(T) = \frac{\mathrm{E}\{n(t)\}}{\gamma} = \mathrm{E}(W) + \mathrm{E}(s)$$

$$= \frac{\mathrm{E}\{n_q(t)\}}{\gamma} + \frac{\mathrm{E}\{n_s(t)\}}{\gamma}$$



Για ουρά με **ένα** εξυπηρετητή:

$$\mathrm{E}\{n_s(t)\} = \gamma \mathrm{E}(s) = \frac{\gamma}{\mu}$$

$$= 0 \cdot \mathrm{P}\{n(t) = 0\} + \mathrm{P}\{n(t) > 0\}$$

$$= \mathrm{P}\{n(t) > 0\} =$$

(ο βαθμός χρησιμοποίησης του εξυπηρετητή $u = \frac{\gamma}{\mu} = \mathrm{P}\{n(t) > 0\}$)

ΚΑΤΑΤΑΞΗ ΟΥΡΩΝ ΑΝΑΜΟΝΗΣ

- ***A/S/N/K***
  - *A* : Τύπος διαδικασίας εισόδου πελατών
  - *S* : Τύπος τυχαίας μεταβλητής χρόνου εξυπηρέτησης
  - *N*: Αριθμός εξυπηρετητών
  - *K* : Χωρητικότητα συστήματος (μέγιστος αριθμός πελατών στην αναμονή + εξυπηρέτηση)
- *Παραδείγματα*
  - ***M/M/1****:* Αφίξεις Poisson (*Markov, Memoryless*), ανεξάρτητοι χρόνοι εξυπηρέτησης εκθετικοί (*Markov*), 1 εξυπηρετητής, άπειρη χωρητικότητα συστήματος (*μηδενικές απώλειες ή αστάθεια*)
  - ***M/D/1:*** Αφίξεις Poisson (*Markov, Memoryless*), ανεξάρτητοι χρόνοι εξυπηρέτησης σταθεροί (*Deterministic*), 1 εξυπηρετητής, άπειρη χωρητικότητα συστήματος
  - ***M/G/1/4:*** Αφίξεις Poisson (*Markov, Memoryless*), ανεξάρτητοι χρόνοι εξυπηρέτησης γενικής κατανομής (*General*), 1 εξυπηρετητής, χωρητικότητα συστήματος 4 πελάτες
  - ***M/M/4/8****:* Αφίξεις Poisson (*Markov, Memoryless*), ανεξάρτητοι χρόνοι εξυπηρέτησης εκθετικοί (*Markov*), 4 εξυπηρετητές, χωρητικότητα συστήματος 8 πελάτες: *Μοντέλο κέντρου κλήσεων (call center) με 4 χειριστές – τηλεφωνητές & μέχρι 4 κλήσεις στην αναμονή*