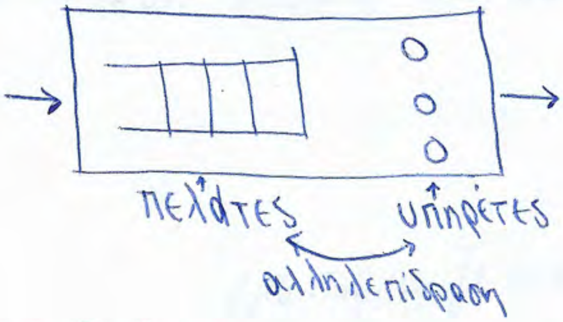
16/5/19
Σ.μ.Ε.Ε.

Μαθημα 19

Εισαγωγή στις Ουρές Αναμονής (ιστορία 110 χρόνων · 1° paper : του γνωστού Erlang!) (1909)

① Βασική ιδέα



Ουρά = Σύστημα Εξυπηρέτησης = Σύστημα Εισόδου-Εξόδου με διακριτές μονάδες και στοχαστικό χαρακτήρα

[Δεν είναι απαραίτητο να μιλάμε για ανθρώπους και ταμεία· σύστημα εξυπηρέτησης μπορεί πχ να είναι ένα αεροδρόμιο με υπηρέτες τους αεροδιαδρόμους και πελάτες τα αεροπλάνα(!)] (υπάρχει τυχαιότητα στις αφίξεις και στις εξυπηρετήσεις).

② Ερωτήματα (είτε από τη σκοπιά του διαχειριστή είτε του πελάτη)

- Πόσοι πελάτες είναι στην Ουρά κατά μέσο όρο;
- Πόσο περιμένει ένας πελάτης κατά μέσο όρο; (είτε χρόνος αναμονής είτε χρόνος συνολικής παραμονής → δεν είναι πάντα κακό ο περισσότερος χρόνος: πχ σε ένα εστιατόριο δε θέλουμε μεγάλη αναμονή, αλλά μόλις καθίσουμε δεν μας πειράζει να παραμείνουμε πολύ → εξαρτάται πάντα από το είδος της εξυπηρέτησης)
- Ποιο ποσοστό του χρόνου εξυπηρετεί ο υπηρέτης;

③ Βασικά Χαρακτηριστικά Ουρών

Ονοματολογία Kendall (ήρθε αρκετά αργότερα από το paper του Erlang)

Ονοματολογία Kendall : A / B / c / k ( )

γράμματα: A, B
αριθμοί: c, k

A: Διαδικασία αφίξεων
B: κατανομή χρόνων εξυπηρέτησης
c: πλήθος παράλληλων υπηρετών
k: χωρητικότητα του συστήματος (θέσεις εξυπηρέτησης + θέσεις αναμονής)
( ): πειθαρχία ουράς (τρόπος επιλογής πελάτη προς εξυπηρέτηση)

(1)

Τι τιμές μπορούν να πάρουν?

- **A**: Διαδικασία αφίξεων
  - → (GI): αναγεννητική διαδικασία αφίξεων (γενικοί ανεξάρτητοι ενδιάμεσοι χρόνοι αφίξεων)
    general independent
    (general: γενικοί, independent: ανεξάρτητοι)
  - → (M): Poisson διαδικασία αφίξεων (Exp, ανεξάρτητοι ενδιάμεσοι χρόνοι αφίξεων)
    memoryless /Markovian
    (αμνήμονη ιδιότητα της Exp)
  - → (D): ντετερμινιστική διαδικασία αφίξεων (σταθεροί ενδιάμεσοι χρόνοι)
    deterministic

- **B**: Κατανομή χρόνου εξυπηρέτησης
  - → (GI): ανεξάρτητοι ισόνομοι χρόνοι εξυπηρέτησης
  - → (M): Exp ισόνομοι χρόνοι εξυπηρέτησης
  - → (D): Σταθεροί χρόνοι εξυπηρέτησης

- Πειθαρχία Ουράς
  - → FCFS: First-Come-First-Served (ή FIFO: First-In-First-Out) (ο προγενέστερος χρονικά εξυπηρετείται πρώτος)
  - → LCFS: Last-Come-First-Served (ο τελευταίος εξυπηρετείται πρώτος) (ή LIFO) [αφορά κυρίως συστήματα με πελάτες όχι ανθρώπους· έχει δειχθεί ότι, αν και ίσως λίγο "άδικη", έχει συνθήκες βελτιστότητας από υπολογιστική άποψη]
  - → SIRO: Service-In-Random-Order
  - → SSTF: Shortest-Service-Time-First

Σχετικά με την ονοματολογία:
- αν $k=\infty$, παραλείπεται
- αν η πειθαρχία είναι (FCFS), παραλείπεται.

π.χ. D/M/1/4 :
- → ντετερμινιστική διαδικασία αφίξεων (π.χ. συρμοί)
- → εκθετικοί χρόνοι εξυπηρέτησης
- → ένας υπηρέτης
- → χωρητικότητα =4
- → πειθαρχία FCFS

④ Βασικές Παράμετροι εισόδου

a: Μέσος ενδιάμεσος χρόνος αφίξεων, $\lambda = \frac{1}{a}$: Ρυθμός αφίξεων

b: Μέσος χρόνος εξυπηρέτησης, $\mu = \frac{1}{b}$: Ρυθμός εξυπηρέτησης

⑤ Βασικές στοχαστικές διαδικασίες εξόδου

αφορούν τον διαχειριστή του συστήματος

$Q(t)$ = πλήθος πελατών στο σύστημα τη στιγμή t

$Q_q(t)$ = πλήθος πελατών σε αναμονή τη στιγμή t

$Q_s(t)$ = πλήθος πελατών σε εξυπηρέτηση τη στιγμή t

(το $Q(t)$ σπάει στα $Q_q(t)$ και $Q_s(t)$)

αφορούν τους πελάτες του συστήματος

S_n = χρόνος παραμονής στο σύστημα του n-οστού πελάτη

(waiting) W_n = χρόνος αναμονής στο σύστημα του n-οστού πελάτη

X_n = χρόνος εξυπηρέτησης του n-οστού πελάτη

(το S_n σπάει στα W_n και X_n)

αφορούν τους υπηρέτες του συστήματος

Z_n = n-οστός κύκλος λειτουργίας του συστήματος (διάστημα από άφιξη σε κενό σύστημα σε άφιξη σε εκ νέου κενό)

Y_n = n-οστή περίοδος συνεχούς λειτουργίας

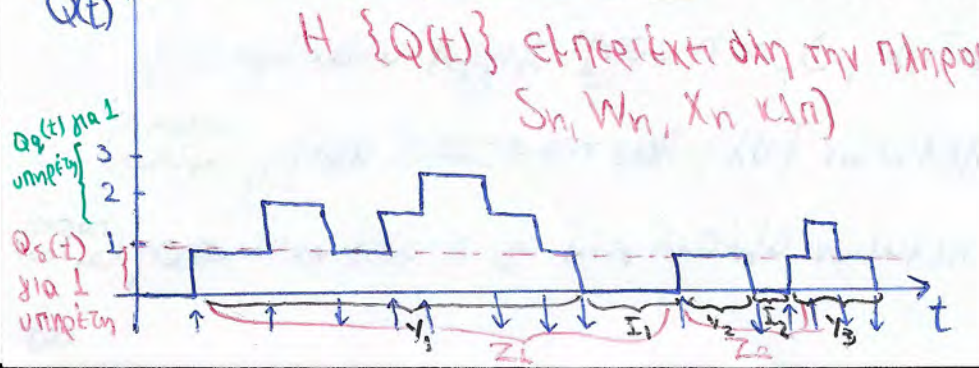
I_n = n-οστή περίοδος αργίας

(το Z_n σπάει στα Y_n και I_n)

⑥ Τυπική εξέλιξη της Q(t)

← σ.δ. συνεχούς χρόνου και διακριτού χώρου καταστάσεων

Η $\{Q(t)\}$ εμπεριέχει όλη την πληροφορία που χρειαζόμαστε (Q_q, Q_s, S_n, W_n, X_n κλπ)



⑦ Αναγεννητικότητα (το σκήμα της $\{Q(t)\}$ θυμίζει αναγεννητική σ.δ.)

Στην $GI|GI|k|c$ $\rightsquigarrow$ η $\{Q(t)\}$ είναι αναγεννητική διαδικασία και οι στιγμές αναγέννησης είναι οι ενάρξεις των κύκλων λειτουργίας

⑧ Βασικά μέτρα απόδοσης

(· πόσοι πελάτες στο σύστημα
· πόσο χρόνο παραμένει ένας πελάτης)

- $$P_j = \text{(μακροπρόθεσμο μέσο) ποσοστό χρόνου που βρίσκονται στο σύστημα } j \text{ πελάτες} = \lim_{t\to\infty} \frac{E\left[\int_0^t 1_{\{Q(u)=j\}}\,du\right]}{t} \equiv Pr[\overset{\text{οριακή κατανομή}}{Q}=j]$$

H p_j είναι η οριακή δειγματική συνάρτηση πιθανότητας της $\{Q(t)\}$
(η παραπάνω έκφραση μας παραπέμπει σε ΣΑΘ Κ (!))

- $$E[Q] = \sum_{j=0}^{\infty} j\cdot P_j = \lim_{t\to\infty} \frac{\overset{\text{με τη μέση τιμή "διώχνουμε" την τυχαιότητα}}{E}\left[\int_0^t Q(u)\,du\right]}{\underbrace{t}_{\text{πραγματικοί αριθμοί (όχι τ.μ.)}}} = \text{μακροπρόθεσμο μέσο πλήθος πελατών στο σύστημα}$$

Έχουμε τα αντίστοιχα και για τους χρόνους εξυπηρέτησης:

- $$\underbrace{F_S(x)}_{\text{οριακή δειγματική σ.κ. της } \{S_n\}} = \lim_{n\to\infty} \frac{E\left[\sum_{i=1}^{n} 1_{\{S_i\le x\}}\right]}{\underbrace{n}_{\text{πραγματικοί αριθμοί (όχι τ.μ.)}}} = \text{το ποσοστό των πελατών που μένουν στο σύστημα το πολύ } x$$

- $$E[S] = \int_0^{\infty} x\,dF_S(x) = \lim_{n\to\infty} \frac{E\left[\sum_{i=1}^{n} S_i\right]}{n} = \text{μέσος χρόνος παραμονής ενός πελάτη στο σύστημα}$$

⑨ Εμφυτευμένες κατανομές πλήθους πελατών σε στιγμές αφίξεων και αναχωρήσεων

$A_n :=$ n-οστή στιγμή άφιξης , $D_n :=$ n-οστή στιγμή αναχώρησης,

$Q_n^- = Q(A_n^-) =$ πλήθος πελατών μόλις πριν την n-οστή άφιξη (πόσους βρίσκω),

$Q_n^+ = Q(D_n^+) =$ πλήθος πελατών αμέσως μετά την n-οστή αναχώρηση (πόσους αφήνω)

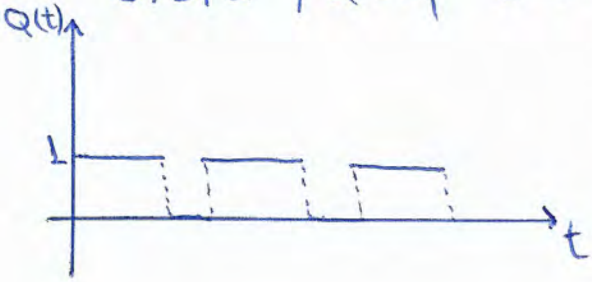
$$a_j = \begin{matrix}\text{ποσοστό των αφίξεων} \\ \text{που βλέπουν } j \text{ πελάτες}\end{matrix} = \lim_{n\to\infty} \frac{E\left[\sum_{k=1}^{n} 1_{\{Q_k^- = j\}}\right]}{n}$$

$$d_j = \begin{matrix}\text{ποσοστό των αναχωρήσεων} \\ \text{που αφήνουν } j \text{ πελάτες}\end{matrix} = \lim_{n\to\infty} \frac{E\left[\overbrace{\sum_{k=1}^{n} 1_{\{Q_k^+ = j\}}}^{\text{πόσοι από τους } n \text{ είδαν } j \text{ πελάτες αναχωρώντας}}\right]}{n}$$

Γενικά, οι p_j, a_j, d_j είναι διαφορετικές. Και οι 3 αναφέρονται στην πιθανότητα να δούμε j πελάτες, αλλά: $\begin{cases} p_j : \text{σε τυχαία χρονική στιγμή} \\ a_j : \text{σε άφιξη} \\ d_j : \text{σε αναχώρηση} \end{cases}$

⑩ Παράδειγμα 1: Γενικά, $(p_j) \neq (a_j)$

- D/D/1, $a = 1$, $b = 0.9$



$$p_j = \begin{cases} 0.1, & j=0 \\ 0.9, & j=1 \\ 0, & j\geq 2 \end{cases}, \quad a_j = \begin{cases} \underbrace{1, \ j=0}_{\substack{\text{για να μπεις πρέπει να} \\ \text{μην υπάρχει πελάτης} \\ \text{(1 υπηρέτης μόνο)}}} \\ 0, \ j\geq 1 \end{cases}$$

$\| $

d_j (όταν φεύγει ένας πελάτης, πάντα αφήνει κενό σύστημα)

- D/D/1, $a = 1$, $b = 0.1$

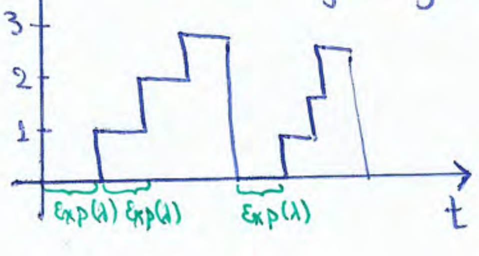
$$p_j = \begin{cases} 0.9, & j=0 \\ 0.1, & j=1 \\ 0, & j\geq 2 \end{cases}, \quad d_j = a_j = \begin{cases} 1, & j=0 \\ 0, & j\geq 1 \end{cases}$$

δηλαδή ο αφικνούμενος πελάτης βλέπει το ίδιο, παρ' όλο που τα συστήματα είναι τελείως διαφορετικά

Συμπέρασμα: η πληροφορία του αφικνούμενου πελάτη δεν μας λέει τίποτα για το σύστημα!

⑪ Παράδειγμα 2: Γενικά, $(a_j) \neq (d_j)$

Σύστημα εξυπηρέτησης: Βανάκι με k θέσεις, διαδικασία Poisson αφίξεων. Μόλις συμπληρωθούν οι θέσεις, το βανάκι αναχωρεί, $Q(t)$ = # πελατών στη στάση



$$a_j = p_j = \begin{cases} \frac{1}{k} \leftarrow \frac{1/\lambda}{k/\lambda} \ \text{(διακριτή ομοιόμορφη)}, & j = 0, 1, \ldots, k-1 \\ 0, & \text{διαφορετικά} \end{cases}, \quad d_j = \begin{cases} 1, & j=0 \\ 0, & \text{διαφορετικά} \end{cases}$$

(είναι σίγουρο ότι δεν θα δούμε πελάτες τη στιγμή της αναχώρησης)