

Ανάλυση της Απόδοσης Πληροφοριακών Συστημάτων

**Διαδικασίες Markov Συνεχούς Χρόνου
Μοντέλα Γεννήσεων - Θανάτων**

Γιάννης Γαροφαλάκης

Συμβολισμός συστημάτων αναμονής

■ $A/B/m$

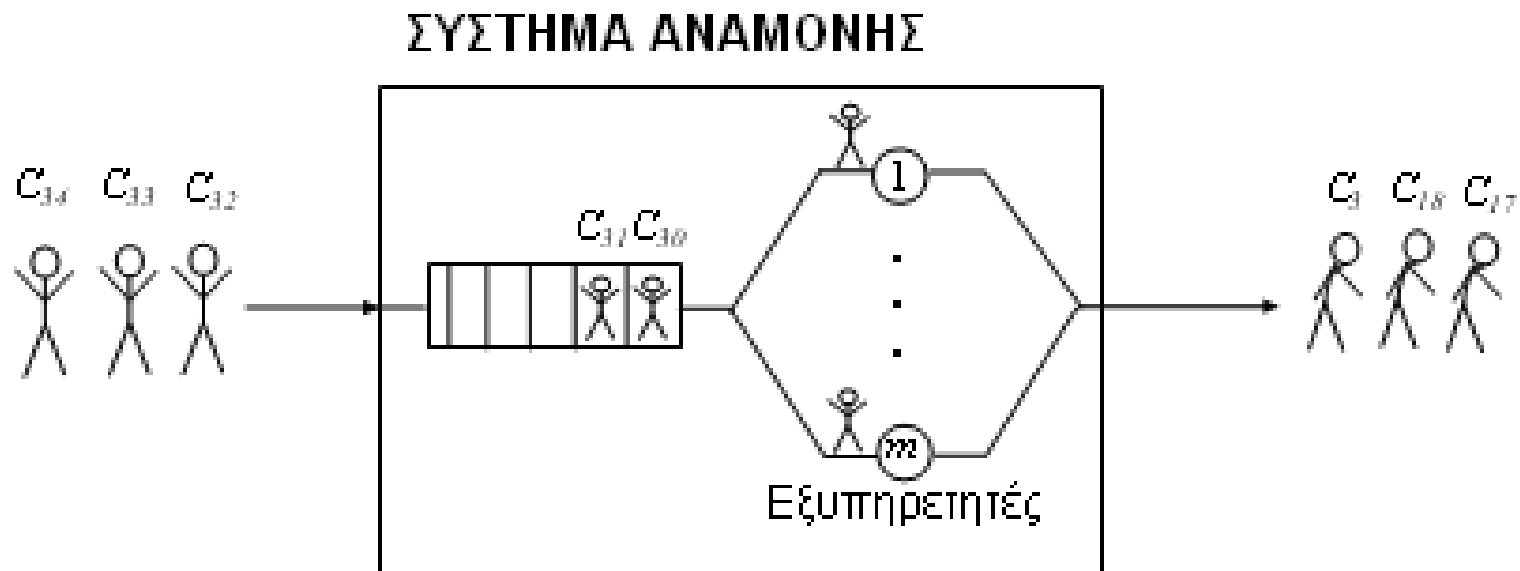
- A, B : Συναρτήσεις κατανομής πιθανότητας $X.A$ και $X.E$ αντίστοιχα. Εκφράζονται ως
 - M (για την εκθετική κατανομή).
 - D (για τη ντετερμινιστική [σταθερή] κατανομή).
 - Er (για την κατανομή Erlang r -βαθμίδων).
 - G (για ΟΠΟΙΑΔΗΠΟΤΕ ΚΑΤΑΝΟΜΗ)
- m : αριθμός εξυπηρετητών

■ $A/B/m/K/M$

- K : η χωρητικότητα του συστήματος
- M : το μέγεθος του πληθυσμού των πελατών
όταν αυτά είναι διαφορετικά από ∞

■ Παράδειγμα: **$D/M/2//200$**

Αναπαράσταση συστήματος αναμονής



- $A(t), B(x)$: αυθαίρετα
- m εξυπηρετητές
- Αριθμούμε τους πελάτες με το δείκτη n και ορίζουμε C_n τον n -οστό πελάτη που εισέρχεται στο σύστημα

Συμβολισμοί βασικών μεγεθών

- **Χρόνοι μεταξύ διαδοχικών αφίξεων:**

$\tau_n \equiv$ χρονική στιγμή άφιξης του πελάτη C_n

$t_n \equiv$ χρόνος μεταξύ των αφίξεων των C_{n-1}, C_n

$$= \tau_n - \tau_{n-1} \text{ για } n \geq 2 \quad (t_1 = \tau_1)$$

$\text{Prob}[t_n \leq t] = A(t)$, δηλαδή το $A(t)$ είναι ανεξάρτητο του n

\bar{t} μέσος χρόνος μεταξύ διαδοχικών αφίξεων

Ρυθμός αφίξεων (arrival rate) των πελατών: $\lambda = \frac{1}{\bar{t}}$

- **Χρόνοι εξυπηρέτησης:**

$x_n \equiv$ χρόνος εξυπηρέτησης του C_n

$$\text{Prob}[x_n \leq x] = B(x)$$

\bar{x} : μέσος χρόνος εξυπηρέτησης

Ρυθμός εξυπηρέτησης (service rate) των πελατών : $\mu = \frac{1}{\bar{x}}$

Συμβολισμοί βασικών μεγεθών (2)

- **Χρόνος αναμονής ενός πελάτη στην ουρά:**

$w_n \equiv$ χρόνος αναμονής (στην ουρά) του C_n .

$W = \bar{w}$ μέσος χρόνος αναμονής

- **Συνολικός χρόνος ενός πελάτη στο σύστημα (χρόνος απόκρισης):**

$s_n \equiv$ χρόνος συστήματος (ουρά + εξυπηρέτηση) του C_n

$$= w_n + x_n$$

$T = W + \bar{x}$ μέσος χρόνος συστήματος ($T \equiv \bar{s}$)

Αλυσίδες Markov συνεχούς χρόνου (1)

Τα απλούστερα συστήματα: $M/M/m/K$

- *Εκθετικά κατανομημένοι χρόνοι μεταξύ διαδοχικών αφίξεων (Χ.Α.)*

$$A(t) = 1 - e^{-\lambda t}, \quad t \geq 0$$

- *Εκθετικά κατανομημένοι χρόνοι εξυπηρέτησης (Χ.Ε.)*

$$B(x) = 1 - e^{-\mu x}, \quad x \geq 0$$

Αλυσίδες Markov συνεχούς χρόνου (2)

- *Ιδιότητα της αμνησίας*: «ο χρόνος ως το επόμενο γεγονός, είναι ανεξάρτητος από το χρόνο που έχει περάσει από το τελευταίο γεγονός».

- ΑΦΙΞΕΙΣ:

Αν έχει περάσει χρόνος t_0 από την τελευταία άφιξη (του C_{n-1})

$$Prob[t_n \leq t + t_0 \mid t_n > t_0] = Prob[t_n \leq t]$$

- ΑΝΑΧΩΡΗΣΕΙΣ:

Αν έχει περάσει χρόνος x_0 εξυπηρέτησης του πελάτη C_n

$$Prob[x_n \leq x + x_0 \mid x_n > x_0] = Prob[x_n \leq x]$$

Αλυσίδες Markov συνεχούς χρόνου (3)

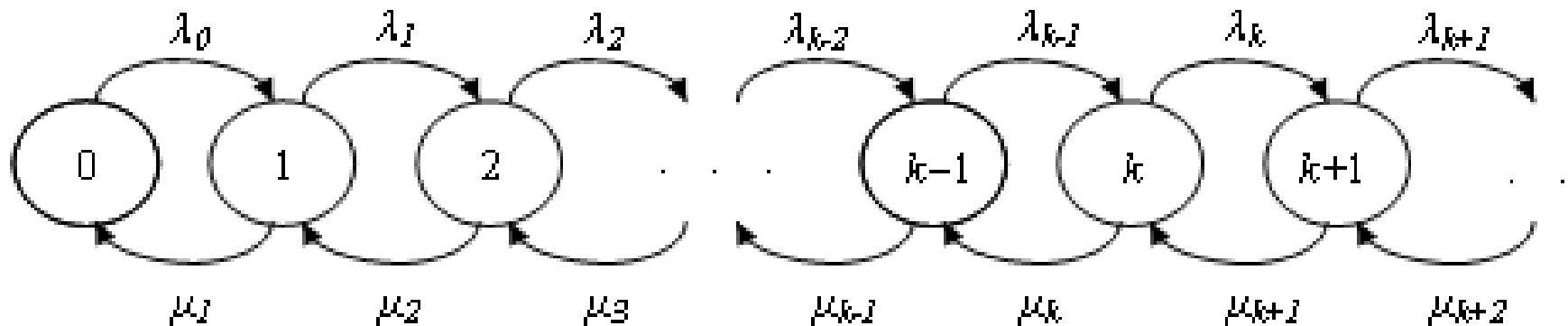
- $P_k(t) = Prob[k \text{ πελάτες στο σύστημα τη χρονική στιγμή } t]$
για $0 \leq k \leq K, \quad t \geq 0$
- $p_k = \lim_{t \rightarrow \infty} P_k(t) = Prob[k \text{ πελάτες στο σύστημα κάποια χρονική στιγμή στο μέλλον}]$
- *Κατανομή μόνιμης κατάστασης.*
- Υπάρχει, αν το σύστημα είναι σταθερό ($0 \leq \rho < 1$)

Νόμος ισορροπίας της ροής πιθανότητας

Στη μόνιμη κατάσταση, ο «ρυθμός ροής πιθανότητας» μιας αλυσίδας Markov από κάθε κατάσταση, είναι ίσος με το «ρυθμό ροής πιθανότητας» προς την κατάσταση.

Αλυσίδες Markov Γεννήσεων – Θανάτων (1)

- Αν το σύστημα βρίσκεται στην κατάσταση j , τότε στην επόμενη αλλαγή κατάστασης θα βρεθεί σε μια από τις καταστάσεις $j-1$ ή $j+1$.
- λ_k : ρυθμός αφίξεων όταν υπάρχουν k πελάτες στο σύστημα
- μ_k : ρυθμός εξυπηρέτησης όταν υπάρχουν k πελάτες στο σύστημα



Διάγραμμα καταστάσεων-ρυθμών μεταβάσεων

Αλυσίδες Markov Γεννήσεων – Θανάτων (2)

- {Ρυθμός ροής πιθανότητας **από** την κατάσταση k } =

$$p_k \cdot (\lambda_k + \mu_k)$$

- {Ρυθμός ροής πιθανότητας **προς** την κατάσταση k } =

$$p_{k-1} \cdot \lambda_{k-1} + p_{k+1} \cdot \mu_{k+1}$$

- Με βάση το νόμο ισορροπίας ροής

- Για $k \geq 1$ $p_k \cdot (\lambda_k + \mu_k) = p_{k-1} \cdot \lambda_{k-1} + p_{k+1} \cdot \mu_{k+1}$ (1)

- Για $k = 0$ $p_0 \cdot \lambda_0 = p_1 \cdot \mu_1$ (2)

Αλυσίδες Markov Γεννήσεων – Θανάτων (3)

- Ισχύει πάντα ότι
$$\sum_{k=0}^{\infty} p_k = 1 \quad (3)$$

- Λύνοντας τις εξισώσεις (1), (2), (3), παίρνουμε:

$$p_k = p_0 \cdot \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}$$

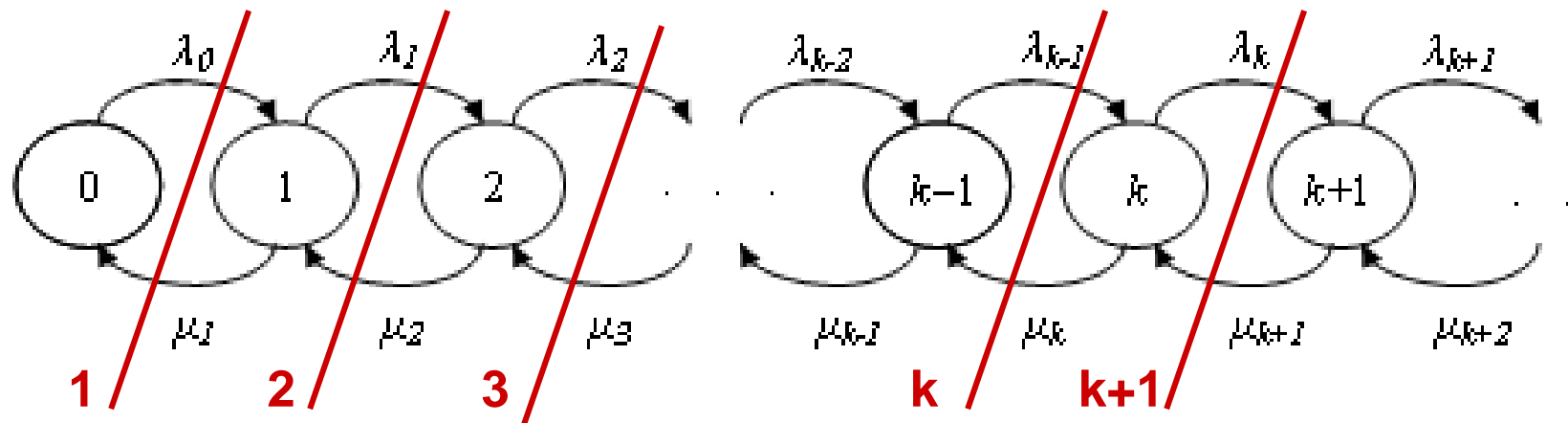
$$p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}} \quad (4)$$

- Η παραπάνω λύση υπάρχει (δηλαδή, υπάρχει μόνιμη κατάσταση), αν $p_0 > 0$, δηλαδή αν ο παρονομαστής της σχέσης (4) είναι μικρότερος από ∞ . Για να ισχύει το τελευταίο, θα πρέπει η ακολουθία λ_k / μ_k να συγκλίνει, δηλαδή θα πρέπει να υπάρχει κάποιο k_0 τέτοιο ώστε:
$$\frac{\lambda_k}{\mu_k} < 1 \quad \text{για όλα τα } k \geq k_0$$

Αλυσίδες Markov Γεννήσεων – Θανάτων (4)

ΕΝΑΛΛΑΚΤΙΚΑ

Ο Νόμος διατήρησης της ροής εφαρμόζεται και σε κάθε «σύνορο» της αλυσίδας Markov:



1: $\rho_0 \cdot \lambda_0 = \rho_1 \cdot \mu_1$

2: $\rho_1 \cdot \lambda_1 = \rho_2 \cdot \mu_2$

:

k: $\rho_{k-1} \cdot \lambda_{k-1} = \rho_k \cdot \mu_k$

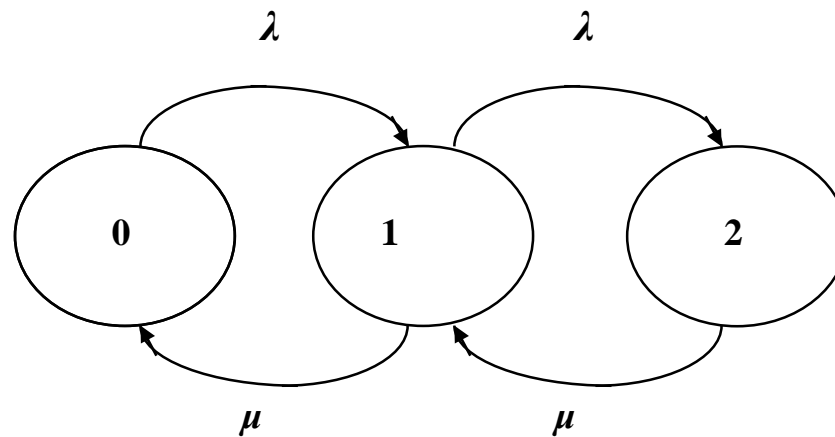
Ίδια Αποτελέσματα

Αλυσίδες Markov Γεννήσεων – Θανάτων (5)

ΠΑΡΑΔΕΙΓΜΑ

Μας δίνεται μια αλυσίδα Markov γεννήσεων – θανάτων, η οποία έχει μόνο τρεις καταστάσεις $\{0, 1, 2\}$, ενώ ισχύει:

$$\lambda_k = \lambda \text{ για } k = 0, 1 \quad \text{και} \quad \mu_k = \mu \text{ για } k = 1, 2$$



Διάγραμμα Καταστάσεων – Ρυθμών Μεταβάσεων

Αλυσίδες Markov Γεννήσεων – Θανάτων (6)

- Για την κατάσταση 0: $p_0 \cdot \lambda = p_1 \cdot \mu$
- Για την κατάσταση 1: $p_1 \cdot (\lambda + \mu) = p_0 \cdot \lambda + p_2 \cdot \mu$
- Για την κατάσταση 2: $p_2 \cdot \mu = p_1 \cdot \lambda$

Από τις παραπάνω 3 σχέσεις, μόνο οι 2 είναι ανεξάρτητες. Χρησιμοποιούμε την $p_0 + p_1 + p_2 = 1$ με 2 από τις παραπάνω, και παίρνουμε την τελική λύση:

$$p_0 = \frac{1}{1 + \lambda/\mu + (\lambda/\mu)^2}$$

$$p_1 = \frac{\lambda/\mu}{1 + \lambda/\mu + (\lambda/\mu)^2}$$

$$p_2 = \frac{(\lambda/\mu)^2}{1 + \lambda/\mu + (\lambda/\mu)^2}$$

- Η αλυσίδα αυτή αντιστοιχεί στο σύστημα **M/M/1/2**. Γιατί;
- Στο σύστημα αυτό επιτρέπεται $\lambda/\mu \geq 1$. Γιατί;

Διαδικασίες Poisson

- Ειδική περίπτωση Γεννήσεων-Θανάτων (μόνο αφίξεις)

- $\mu_k = 0$ για όλα τα k

- $\lambda_k = \lambda$ για όλα τα k

- Δεν είναι εργοδικό σύστημα. Όλες οι καταστάσεις μεταβατικές.

- Έστω το σύστημα ξεκινά τη στιγμή $t = 0$, άδειο. Δηλαδή:

$$P_k(0) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

- Τη χρονική στιγμή t : $P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$ για $k \geq 0, t \geq 0$

- Κατανομή Poisson

- Μέση Τιμή και Διακύμανση (αριθμού αφίξεων στο $[0, t]$), ίσα με λt . (αναμενόμενο).

- Δηλαδή, στο M/M/1, η διαδικασία μόνο των αφίξεων, είναι Poisson

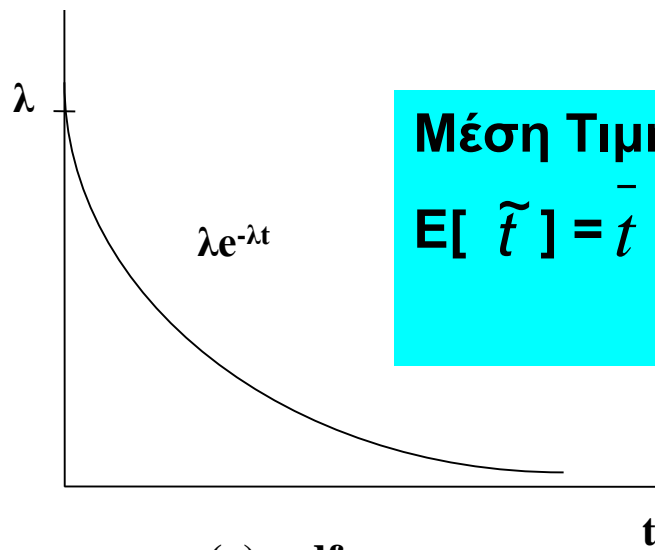
Poisson αφίξεις → Εκθετικοί χρόνοι μεταξύ αφίξεων

- \tilde{t} = ΤΜ για το χρόνο μεταξύ αφίξεων, με PDF $A(t)$ και pdf $\alpha(t)$

Poisson

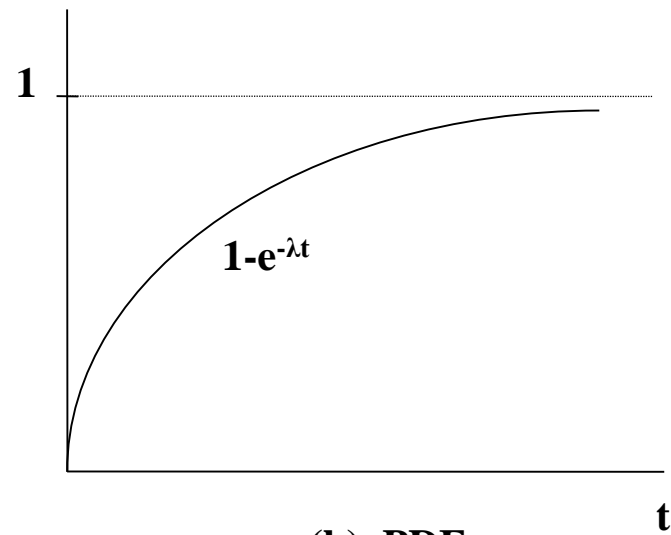
$$A(t) = 1 - P[\tilde{t} > t] = 1 - \overbrace{P_0(t)}^{\text{Poisson}} = 1 - e^{-\lambda t}, \quad t \geq 0 \quad (\text{PDF Εκθετικής})$$

Παράγωγος ως προς t : $\alpha(t) = \lambda e^{-\lambda t}, \quad t \geq 0 \quad (\text{pdf Εκθετικής})$



(a) pdf

Μέση Τιμή:
 $E[\tilde{t}] = \bar{t} = \frac{1}{\lambda}$



(b) PDF

Η εκθετική κατανομή

Ιδιότητα Αμνησίας της Εκθετικής Κατανομής

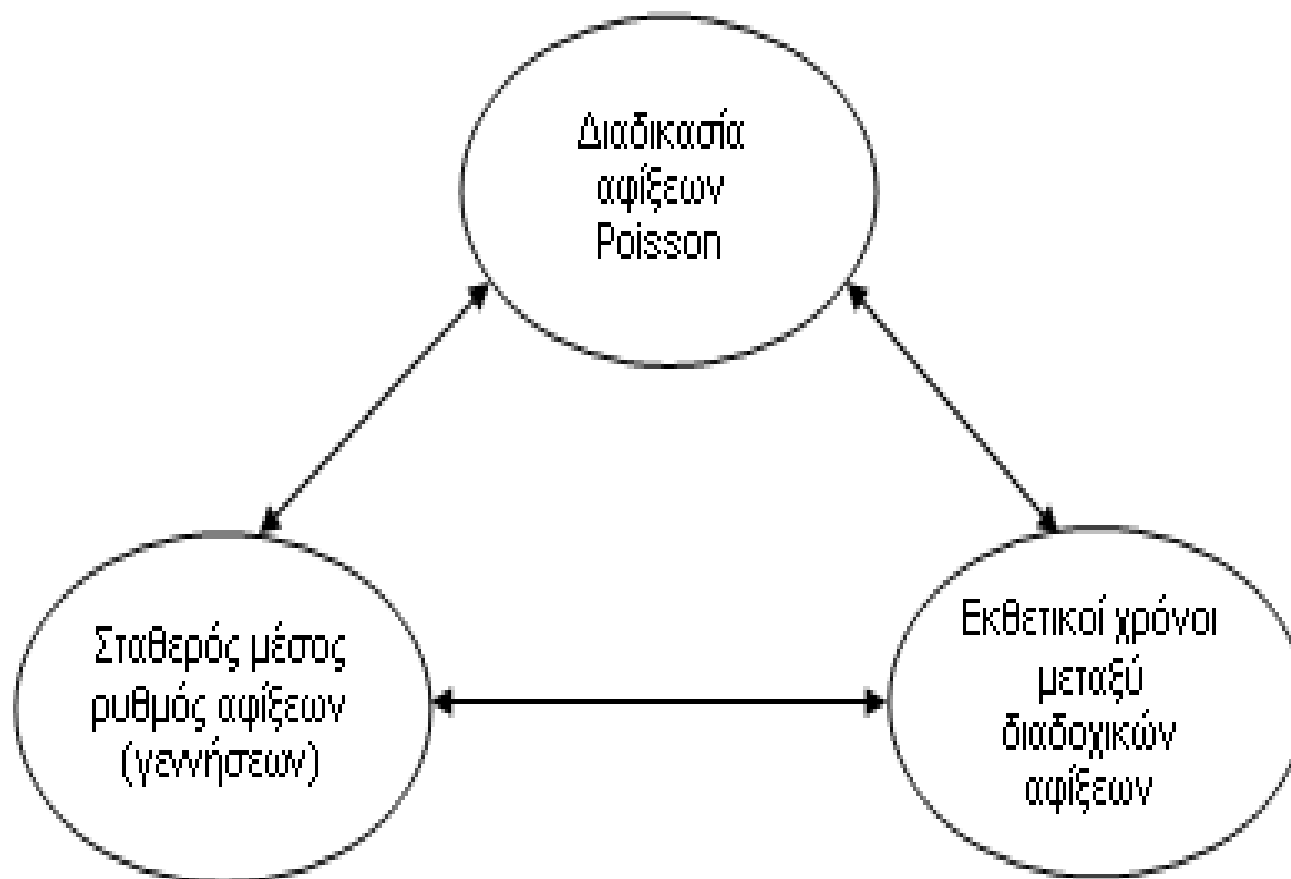
Έστω ότι γίνεται μια άφιξη τη χρονική στιγμή 0 . Τώρα, έστω ότι πέρασαν t_0 δευτερόλεπτα κατά τη διάρκεια των οποίων δεν έγινε άφιξη. Αν αυτή τη στιγμή t_0 ρωτήσουμε «ποια είναι η πιθανότητα η επόμενη άφιξη να γίνει μετά από t δευτερόλεπτα από τώρα», η απάντηση θα είναι:

$$P[\tilde{t} \leq t + t_0 \mid \tilde{t} > t_0] = \frac{P[t_0 < \tilde{t} \leq t + t_0]}{P[\tilde{t} > t_0]} = \frac{P[\tilde{t} \leq t + t_0] - P[\tilde{t} \leq t_0]}{P[\tilde{t} > t_0]}$$

$$\Leftrightarrow P[\tilde{t} \leq t + t_0 \mid \tilde{t} > t_0] = \frac{1 - e^{-\lambda(t+t_0)} - (1 - e^{-\lambda t_0})}{1 - (1 - e^{-\lambda t_0})} \Leftrightarrow$$

$$P[\tilde{t} \leq t + t_0 \mid \tilde{t} > t_0] = 1 - e^{-\lambda t} \Leftrightarrow P[\tilde{t} \leq t + t_0 \mid \tilde{t} > t_0] = P[\tilde{t} \leq t]$$

Σχέσεις



Το κλασικό Σύστημα Αναμονής

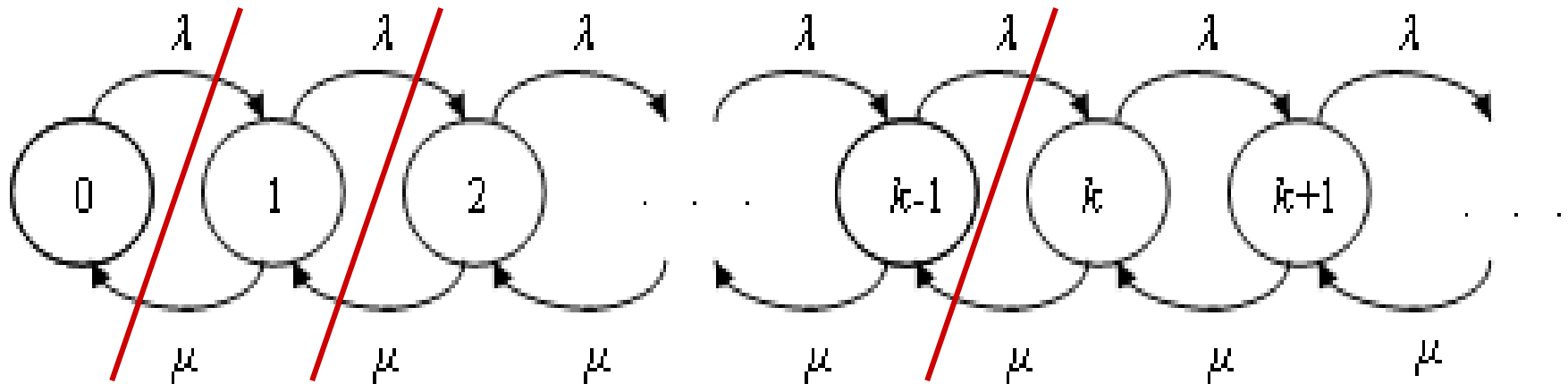
M/M/1

- *Εκθετική* κατανομή της διαδικασίας των χρόνων μεταξύ διαδοχικών αφίξεων
- *Εκθετική* κατανομή των χρόνων εξυπηρέτησης
- Ένας εξυπηρετητής
- Άπειρο μήκος ουράς
- Αλυσίδα Markov Γεννήσεων – Θανάτων
- Με τη συνηθισμένη παραδοχή ότι οι *ρυθμοί* αφίξεων και εξυπηρέτησης δεν εξαρτώνται από την κατάσταση του συστήματος (αριθμός παρόντων πελατών), ισχύει:

$$\lambda_k = \lambda \quad \text{για} \quad k = 0, 1, 2, \dots$$

$$\mu_k = \mu \quad \text{για} \quad k = 1, 2, 3, \dots$$

Το Σύστημα Αναμονής $M/M/1$ (συνέχεια)



Διάγραμμα καταστάσεων - ρυθμών μεταβάσεων για το σύστημα $M/M/1$.

Για τη λύση:

$$p_0 \cdot \lambda = p_1 \cdot \mu$$

$$p_1 \cdot \lambda = p_2 \cdot \mu$$

:

$$p_{k-1} \cdot \lambda = p_k \cdot \mu$$

:

και $\sum_{k=0}^{\infty} p_k = 1$

Λύση συστήματος $M/M/1$

- Χρησιμοποίηση (G/G/1):

$$\rho = \lambda \cdot \bar{x} = \frac{\lambda}{\mu}$$

- Συνθήκη σταθερότητας: $0 < \rho = \frac{\lambda}{\mu} < 1$

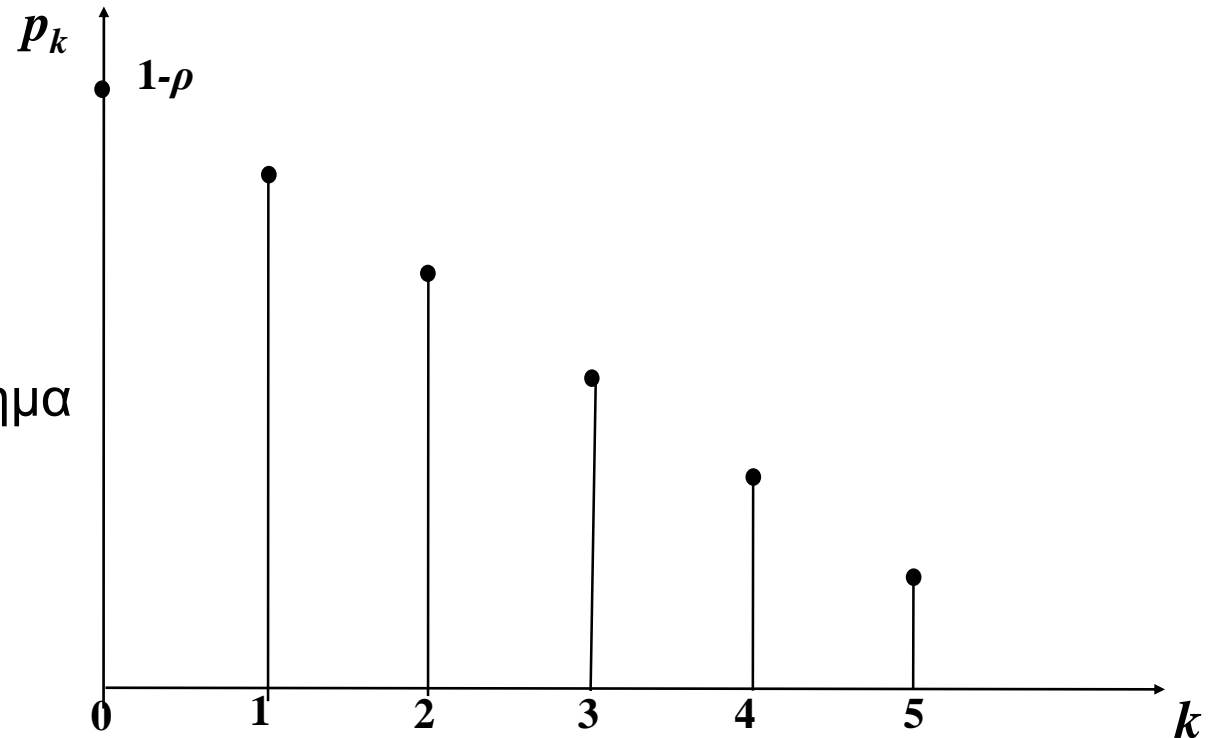
- Από τη γενική λύση των διαδικασιών Γ-Θ (ή την προσέγγιση της προηγούμενης διαφάνειας):

$$p_k = (1 - \rho) \cdot \rho^k \quad \text{για} \quad k = 0, 1, 2, \dots$$

- Περιέχεται το: $p_0 = 1 - \rho$

Λύση συστήματος $M/M/1$ (συν)

- Τα p_k ακολουθούν τη Γεωμετρική Κατανομή
- Εξαρτώνται από τα λ και μ , μόνο μέσω του λόγου ΤΟΥΣ ρ

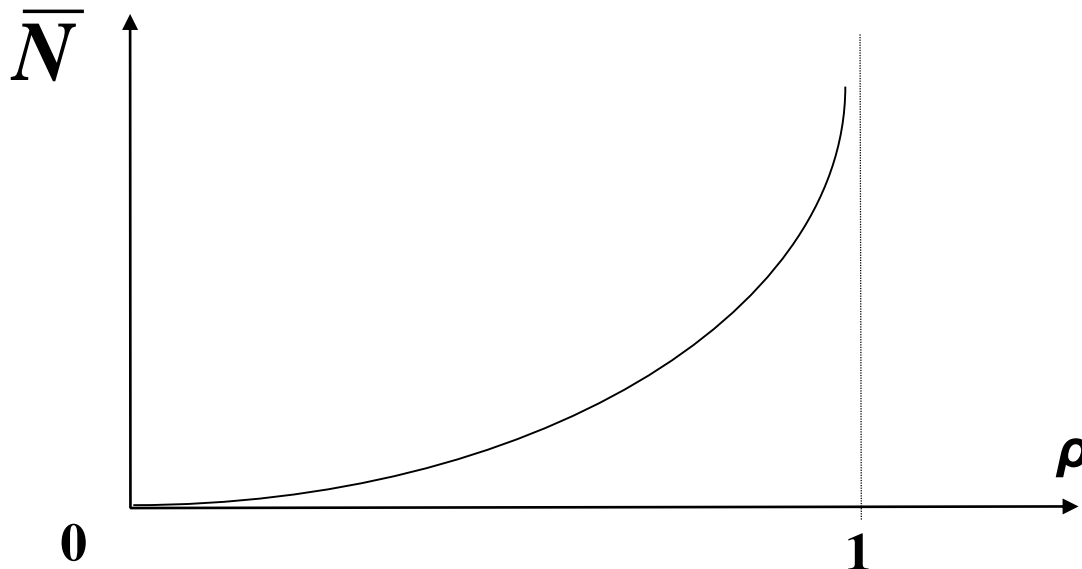


Τα p_k στο σύστημα $M/M/1$.

Μετρικές απόδοσης στο $M/M/1$

- Μέσος αριθμός πελατών στο σύστημα

$$\bar{N} = \sum_{k=0}^{\infty} k p_k = (1 - \rho) \sum_{k=0}^{\infty} k \rho^k = (1 - \rho) \frac{\rho}{(1 - \rho)^2} = \frac{\rho}{1 - \rho}$$

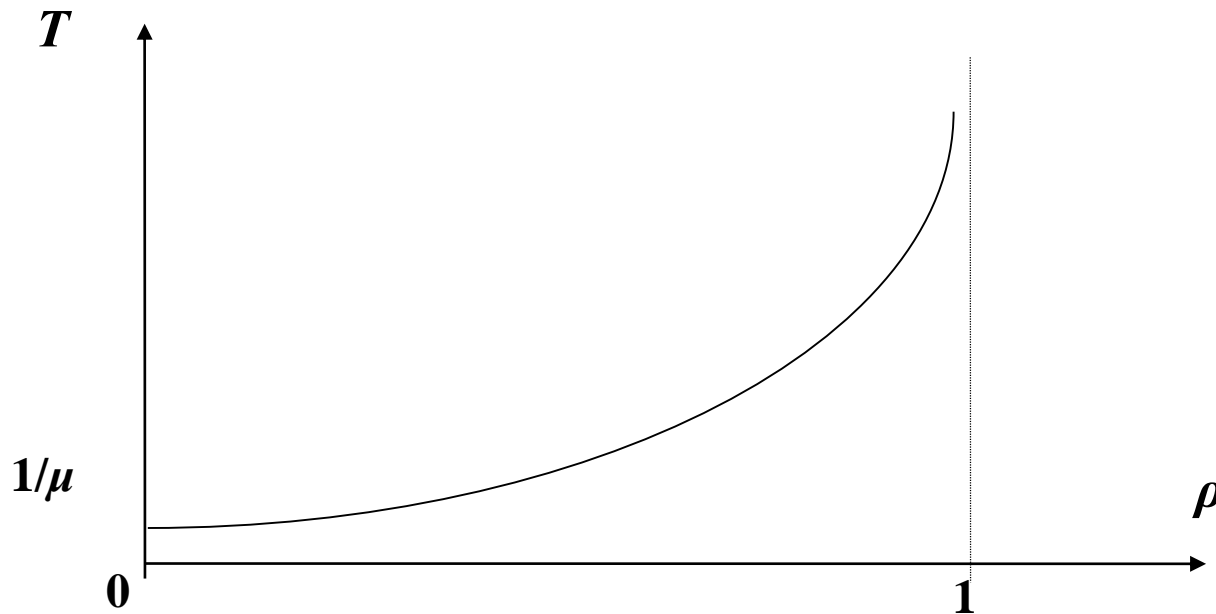


Μετρικές απόδοσης στο $M/M/1$ (συν.)

- Μέσος χρόνος ενός πελάτη στο σύστημα
(*Response Time*)

Με χρήση του *Νόμου του Little*:

$$T = \frac{\bar{N}}{\lambda} = \frac{1/\mu}{1-\rho}$$



Μετρικές απόδοσης στο $M/M/1$ (συν.)

- Μέσος χρόνος αναμονής ενός πελάτη στην ουρά

$$W = T - \bar{x} = T - 1/\mu = \frac{\rho}{\mu \cdot (1 - \rho)}$$

- Μέσος αριθμός πελατών στην ουρά

$$\bar{N}_q = \bar{N} - \rho = \frac{\rho^2}{1 - \rho}$$

- Πιθανότητα να υπάρχουν τουλάχιστον n πελάτες στο σύστημα

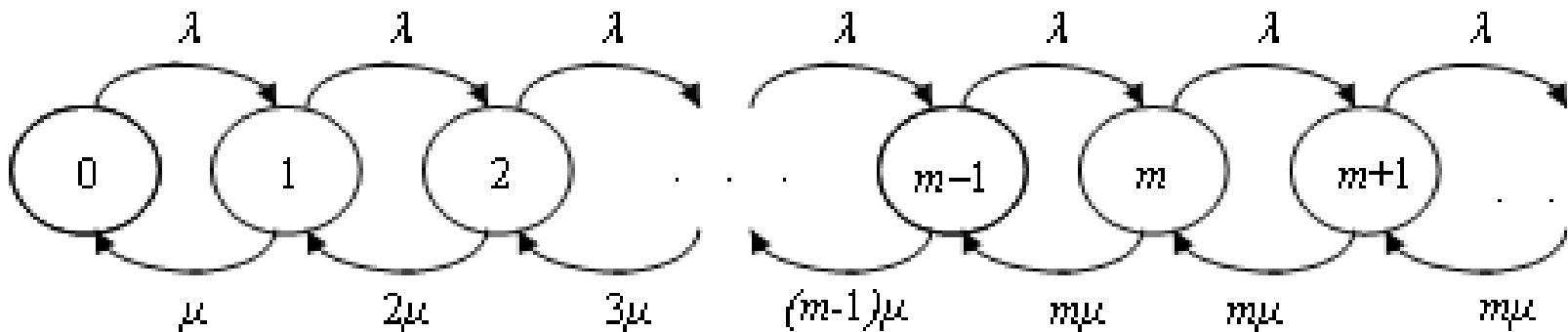
$$P_{(n)} = \text{Prob}[n \text{ ή περισσότεροι πελάτες στο σύστημα}]$$

$$P_{(n)} = \sum_{k=n}^{\infty} p_k = (1 - \rho) \sum_{k=n}^{\infty} \rho^k = (1 - \rho) \rho^n \sum_{k=0}^{\infty} \rho^k = (1 - \rho) \rho^n \frac{1}{1 - \rho} = \rho^n$$

Το σύστημα αναμονής $M/M/m$

- m ίδιοι εξυπηρετητές
- Ο καθένας με ρυθμό εξυπηρέτησης μ
- Τα υπόλοιπα χαρακτηριστικά, ίδια με του $M/M/1$
- $\lambda_k = \lambda$ για $k = 0, 1, 2, \dots$

- $$\mu_k = \begin{cases} k\mu & \text{για } 1 \leq k \leq m \\ m\mu & \text{για } m \leq k \end{cases}$$



Το σύστημα αναμονής $M/M/m$ (συν)

■ Χρησιμοποίηση

$$\rho = \frac{\lambda \bar{x}}{m} = \frac{\lambda}{m\mu}$$

Συνθήκη Σταθερότητας

$$0 < \rho = \frac{\lambda}{m\mu} < 1$$

■ Λύση μόνιμης κατάστασης

$$P_k = \begin{cases} p_0 \frac{(m\rho)^k}{k!} & \text{για } 1 \leq k \leq m \\ p_0 \frac{\rho^k m^m}{m!} & \text{για } k \geq m \end{cases}$$

$$p_0 = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \left(\frac{(m\rho)^m}{m!} \right) \left(\frac{1}{1-\rho} \right) \right]^{-1}$$

Μετρικές απόδοσης στο $M/M/m$

- Πιθανότητα να χρειαστεί να περιμένει στην ουρά ένας πελάτης:

$\Pi = Prob[m \text{ ή περισσότεροι πελάτες στο σύστημα}]$

$$\Pi = \sum_{k=m}^{\infty} p_k = \sum_{k=m}^{\infty} p_0 \frac{\rho^k m^m}{m!} = p_0 \frac{m^m}{m!} \sum_{k=m}^{\infty} \rho^k = p_0 \frac{m^m}{m!} \rho^m \frac{1}{1-\rho} = p_0 \frac{(m\rho)^m}{m!(1-\rho)}$$

- Μέσος αριθμός πελατών στο σύστημα

$$\bar{N} = \sum_{k=0}^{\infty} k p_k = m\rho + \frac{\rho\Pi}{1-\rho}$$

Μετρικές απόδοσης στο $M/M/m$ (συν)

- Μέσος χρόνος ενός πελάτη στο σύστημα (response time)

$$T = \frac{\bar{N}}{\lambda} = \frac{1}{\mu} \left(1 + \frac{\Pi}{m(1-\rho)} \right) \quad (\text{N. Little})$$

- Μέσος χρόνος αναμονής ενός πελάτη στην ουρά

$$W = T - \bar{x} = T - 1/\mu = \frac{\Pi}{m\mu(1-\rho)}$$

- Μέσος αριθμός πελατών στην ουρά

$$\bar{N}_q = \bar{N} - m\rho = \frac{\rho\Pi}{1-\rho}$$

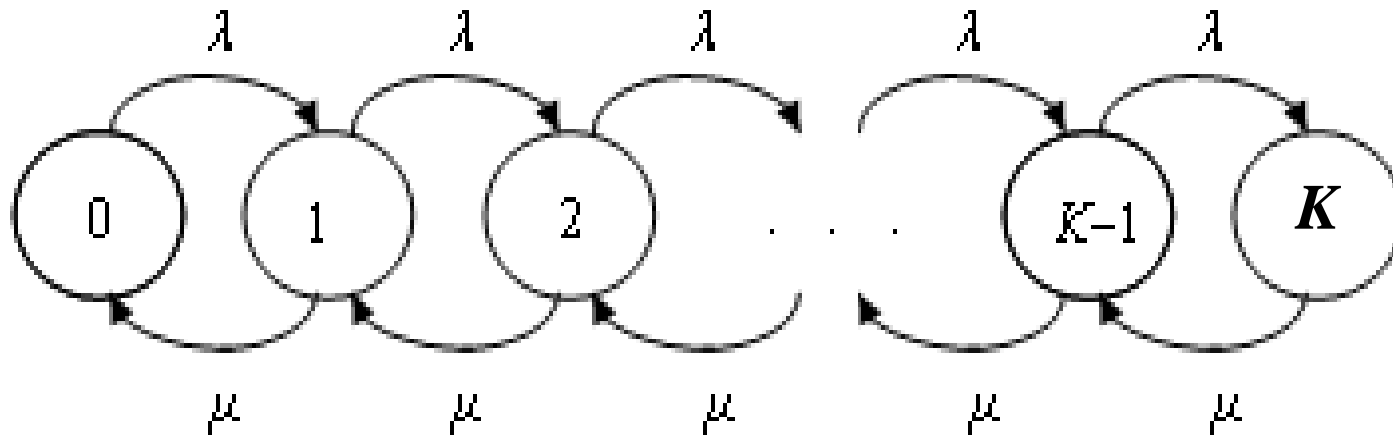
Το σύστημα αναμονής $M/M/1/K$

- Ίδια χαρακτηριστικά με το $M/M/1$, αλλά περιορισμένη χωρητικότητα σε πελάτες.
- Στο σύστημα μπορούν να βρίσκονται το πολύ K πελάτες (στην ουρά και στον εξυπηρετητή).
- Πελάτες που φθάνουν και βρίσκουν γεμάτο το σύστημα, χάνονται.
- Οι ρυθμοί αφίξεων και εξυπηρέτησης του $M/M/1/K$:

$$\lambda_k = \begin{cases} \lambda & \text{για } 0 \leq k < K \\ 0 & \text{για } k \geq K \end{cases}$$

$$\mu_k = \begin{cases} \mu & \text{για } 1 \leq k \leq K \\ 0 & \text{για } k > K \end{cases}$$

Το σύστημα αναμονής $M/M/1/K$ (συν)



■ Λύση συστήματος

$$P_k = \begin{cases} \frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{K+1}} \left(\frac{\lambda}{\mu}\right)^k & \text{για } 0 \leq k \leq K \\ 0 & \text{αλλιώς} \end{cases}$$

Το σύστημα αναμονής $M/M/1/K$ (συν)

ΜΕΤΡΙΚΕΣ

- Μέσος αριθμός πελατών στο σύστημα

$$\bar{N} = \sum_{k=0}^K k p_k = \frac{\lambda/\mu}{1 - \lambda/\mu} - \frac{(K+1)(\lambda/\mu)^{K+1}}{1 - (\lambda/\mu)^{K+1}}$$

- Μέσος αριθμός πελατών στην ουρά

$$\bar{N}_q = \sum_{k=2}^K (k-1) p_k = \frac{\lambda/\mu}{1 - \lambda/\mu} - (\lambda/\mu) \cdot \frac{1 + K(\lambda/\mu)^K}{1 - (\lambda/\mu)^{K+1}}$$

Το σύστημα αναμονής $M/M/1/K$ (συν)

■ Παράδειγμα: Το μοντέλο μιας τηλεφωνικής συσκευής χωρίς κράτηση κλήσεων (παλιό αναλογικό σύστημα):

$M/M/1/1$

$$p_k = \begin{cases} \frac{1}{1 + \lambda/\mu} & \text{για } k = 0 \\ \frac{\lambda/\mu}{1 + \lambda/\mu} & \text{για } k = 1 \\ 0 & \text{αλλιώς} \end{cases}$$

p_0 : Πιθανότητα να μιλήσει, κάποιος που καλεί

p_1 : Πιθανότητα να βρει κατειλημμένη τη συσκευή, κάποιος που καλεί

λ : Μέσος ρυθμός με τον οποίο γίνονται κλήσεις στη συσκευή

$\bar{x} = 1/\mu$: Μέση χρονική διάρκεια μιας συνομιλίας