Απλές Μαρκοβιανές Ουρές & Εφαρμογές

Ι. Δημητρίου

Πανεπιστήμιο Πατρών
Τμήμα Μαθηματικών

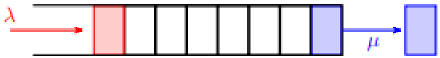
idimit@math.upatras.gr

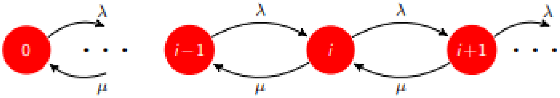
Περιεχόμενα

Το M/M/1 σύστημα

Οι διαδικασίες αφίξεων και εξυπηρετήσεων είναι Poisson
$\Rightarrow$ Birth & Death process.

1. **Αφίξεις:** Poisson, $\lambda > 0$ (φθάνουν λ πελάτες στην μονάδα του χρόνου)
2. **Εξυπηρέτηση:** $exp(\mu)$, $\mu > 0$, (εξυπηρετούνται μ πελάτες στην μονάδα του χρόνου)
3. Άπειρη χωρητικότητα.
4. Διαδικασία αφίξεων-εξυπηρετήσεων ανεξάρτητες.
5. FIFO (FCFS) (First In First Out or First Come First Served)
6. Υποθέτουμε ότι $\rho = \frac{\lambda}{\mu} < 1$ (**ΓΙΑΤΙ;**)





Το M/M/1 σύστημα

Έστω $\{N(t), 0 \leq t < \infty\}$, στοχαστική διαδικασία που παριστά τον αριθμό των πελατών στο σύστημα την χρονική στιγμή t, με χώρο καταστάσεων $E = \{0, 1, 2, \ldots\}$. Έστω

$$p_n(t) = P(N(t) = n), n \geq 0,$$

με $N(0) = 0$. Άρα $p_0(0) = 1$ $(p_n(0) = 0, n = 1, 2, \ldots)$.
Το σύστημα αλλάζει κατάσταση:

1. όταν έρθει πελάτης,
2. όταν ολοκληρωθεί η εξυπηρέτηση ενός πελάτη και αυτός αναχωρήσει.

- $\{N(t), 0 \leq t < \infty\}$ είναι Μαρκοβιανή και χρονικά ομογενής (Γιατί;).
- Το $M/M/1$ είναι ειδική περίπτωση BD process με

$$\lambda_n = \lambda,\, n \geq 0,\, \mu_n = \mu,\, n \geq 1\,.$$

και απειροστό γεννήτορα

$$\mathbb{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \dots \\ 0 & \mu & -(\lambda+\mu) & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

- Ενδιαφερόμαστε για αλλαγές των καταστάσεων στο χρονικό διάστημα $(t, t+dt)$ (Chapman-Kolmogorov).
- Λόγω των ιδιοτήτων της διαδικασίας Poisson και της εκθετικής κατανομής θα έχουμε τις ακόλουθες μεταβάσεις στο $(t, t+dt)$:
    1. $P(1 \text{ άφιξη στο } (t, t+dt)) = \lambda dt + o(dt)$,
    2. $P(\text{καμία άφιξη στο } (t, t+dt)) = 1 - \lambda dt + o(dt)$,
    3. $P(\text{περισσότερες από μια αφίξεις στο } (t, t+dt)) = o(dt)$.

    1. $P(1 \text{ αναχώρηση στο } (t, t+dt)) = \mu dt + o(dt)$,
    2. $P(\text{καμία αναχώρηση στο } (t, t+dt)) = 1 - \mu dt + o(dt)$,
    3. $P(\text{περισσότερες από μια αναχωρήσεις στο } (t, t+dt)) = o(dt)$.
- $\lim_{dt \to 0} \frac{o(dt)}{dt} = 0$.

Εξισώσεις Chapman-Kolmogorov

1. $n > 0$

$$
\begin{aligned}
p_n(t+dt) = {} & p_n(t)p_{n,n}(dt) + p_{n-1}(t)p_{n-1,n}(dt) \\
& + p_{n+1}(t)p_{n+1,n}(dt) \\
= {} & p_n(t)[1-(\lambda+\mu)dt + o(dt)] \\
& + p_{n-1}(t)[\lambda dt + o(dt)] + p_{n+1}(t)[\mu dt + o(dt)] \\
& \Longleftrightarrow \\
\frac{p_n(t+dt)-p_n(t)}{dt} = {} & -p_n(t)(\lambda+\mu) + p_{n-1}(t)\lambda + p_{n+1}(t)\mu + \frac{o(dt)}{dt}
\end{aligned}
$$

2. $n = 0$

$$
\begin{aligned}
p_0(t+dt) = {} & p_0(t)p_{0,0}(dt) + p_1(t)p_{1,0}(dt) \\
= {} & p_0(t)[1-\lambda dt + o(dt)] + p_1(t)[\mu dt + o(dt)] \\
& \Longleftrightarrow \\
\frac{p_0(t+dt)-p_0(t)}{dt} = {} & -p_0(t)\lambda + p_1(t)\mu + \frac{o(dt)}{dt}.
\end{aligned}
$$

Όταν $dt \to 0$, προκύπτει το ακόλουθο σύστημα *διαφορικών εξισώσεων-διαφορών:*

$$\begin{aligned} \frac{dp_0(t)}{dt} &= -\lambda p_0(t) + \mu p_1(t), \\ \frac{dp_n(t)}{dt} &= -(\lambda+\mu)p_n(t) + \mu p_{n+1}(t) + \lambda p_{n-1}(t), \ n > 0. \end{aligned} \tag{1}$$

Η επίλυση του (1) γίνεται με δύο τρόπους:

1. **Difference-equation method:**
   Χρήση Laplace μετασχηματισμού και μετατροπή του (1) σε μια ομογενή εξίσωση διαφορών 2ης τάξης.
2. **z-transform method:**
   Χρήση z-μετασχηματισμού και μετατροπή του (1) σε μια ομογενή διαφορική εξίσωση 1ης τάξης.

Difference-equation method

Έστω $\pi_n(s) = \mathcal{L}(p_n(t)) = \int_0^\infty e^{-st} p_n(t)dt, \; Re(s) > 0$.
Χρησιμοποιώντας την ιδιότητα

$$\mathcal{L}(\frac{dp_n(t)}{dt}) = s\mathcal{L}(p_n(t)) - p_n(0),$$

έχουμε

$$\begin{aligned} (s+\lambda)\pi_0(s) = & \quad 1 + \mu\pi_1(s), \\ (s+\lambda+\mu)\pi_n(s) = & \quad \mu\pi_{n+1}(s) + \lambda\pi_{n-1}(s), \, n > 0. \end{aligned} \tag{2}$$

Η δευτερη είναι εξίσωση διαφορών 2ης τάξης με γενική λύση

$$\pi_n(s) = Ax_+(s)^n + Bx_-(s)^n, \, n \geq 1,$$

όπου A, B σταθερές και $x_\pm(s)$ οι λύσεις της:

$$\mu x^2 - (s+\lambda+\mu)x + \lambda = 0,$$

όπου

$$x_\pm(s) = \frac{(s+\lambda+\mu) \pm \sqrt{(s+\lambda+\mu)^2 - 4\lambda\mu}}{2\mu}.$$

Difference-equation method (συνεχ.)

Θεώρημα 1

(Θεωρημα Rouché) Αν $f(x)$, $g(x)$ είναι αναλυτικές συναρτήσεις μέσα και πάνω σε μια κλειστή καμπύλη C, και αν $|g(x)| < |f(x)|$, $x \in C$, τότε η $f(x)$ και η $f(x) + g(x)$ έχουν τον ίδιο αριθμό ριζών εντός της C.

Στην περίπτωση μας θεωρούμε ως C, τον μοναδιαίο κύκλο $|x| = 1$, $f(x) = -(s + \lambda + \mu)x$, $g(x) = \lambda + \mu x^2$, $Re(s) > 0$. Τότε,

$$\begin{aligned} |f(x)| = & \; |(s + \lambda + \mu)x| = |s + \lambda + \mu| \\ \geq & \; |\lambda + \mu| = \lambda + \mu \\ \geq & \; |\lambda + \mu x^2| = |g(x)|. \end{aligned}$$

- Άρα η χαρακτηριστική εξίσωση

$$f(x) + g(x) = \mu x^2 - (s + \lambda + \mu)x + \lambda = 0,$$

έχει ακριβώς μια ρίζα εντός του μοναδιαίου κύκλου C.
- Δηλαδή $|x_-(s)| < 1 < |x_+(s)|$.

Difference-equation method (συνεχ.)

Όμως

$$\sum_{n=0}^{\infty} p_n(t) = 1 \iff \sum_{n=0}^{\infty} \pi_n(s) = 1/s,\ Re(s) > 0.$$

Αρα $\sum \pi_n(s) = \sum (Ax_+(s)^n + Bx_-(s)^n)$ πρέπει να συγκλίνει και μιας και $|x_+(s)| > 1$, $A = 0$ και

$$\pi_n(s) = Bx_-(s)^n.$$

Χρησιμοποιώντας την

$$\frac{1}{s} = \sum_{n=0}^{\infty} \pi_n(s) = B \sum_{n=0}^{\infty} x_-(s)^n = \frac{B}{1 - x_-(s)} \iff B = \frac{1 - x_-(s)}{s}.$$

Επομένως

$$\pi_n(s) = \frac{(1 - x_-(s))x_-(s)^n}{s}, n \geq 0,\ \ Re(s) > 0. \tag{3}$$

Οριακή συμπεριφορά

- Υποθέτουμε ότι $\rho = \frac{\lambda}{\mu} < 1$ (συνθήκη ευστάθειας).
- Έστω ότι $t \to \infty$. Ενδιαφερόμαστε για

$$\pi_n = \lim_{t\to\infty} p_n(t).$$

A) Χρήση του αποτελέσματος της χρ. εξαρτημένης συμπεριφοράς

Γνωρίζουμε ότι,

$$\pi_n = \lim_{t\to\infty} p_n(t) = \lim_{s\to 0} s\pi_n(s) = \lim_{s\to 0}(1 - x_-(s))x_-(s)^n.$$

Όμως

$$\begin{aligned} \lim_{s\to 0} x_-(s) = & \lim_{s\to 0} \tfrac{s+\lambda+\mu-\sqrt{(s+\lambda+\mu)^2-4\lambda\mu}}{2\mu} \\ = & \begin{cases} \rho, & \lambda < \mu, \\ 1, & \lambda \geq \mu. \end{cases} \end{aligned}$$

Άρα

$$\pi_n = (1-\rho)\rho^n, \; n \geq 0.$$

Αναδρομική επίλυση–Ολικές εξισώσεις ισορροπίας (Μέθοδος 1)

- Μιας και $\lim_{t\to\infty} \frac{dp_n(t)}{dt} = 0$ θα έχουμε από την (1) τις ολικές εξισώσεις ισορροπίας (GBE): $\pi\mathbb{Q} = \mathbf{0} \Rightarrow$

$$\begin{aligned}\lambda\pi_0 &= \mu\pi_1,\\ (\lambda+\mu)\pi_n &= \mu\pi_{n+1} + \lambda\pi_{n-1}, n > 0.\end{aligned}$$

- Ρυθμός εξόδου από την j=Ρυθμός εισόδου στην j.

Διαδικασία επίλυσης

- Λύνοντας την 1η

$$\pi_1 = \frac{\lambda}{\mu}\pi_0 = \rho\pi_0. \tag{4}$$

- Θέτοντας $n = 1$ στην 2η και χρησιμοποιώντας την (4)

$$\mu\pi_2 = \frac{\lambda}{\mu}\pi_0(\lambda+\mu-\mu) \Rightarrow \pi_2 = \rho^2\pi_0. \tag{5}$$

- Θέτοντας $n = 2$ στην 2η και χρησιμοποιώντας τις (4), (5)

$$\mu\pi_3 = \rho^2\pi_0(\lambda+\mu-\mu) \Rightarrow \pi_3 = \rho^3\pi_0. \tag{6}$$

- Συνεχίζοντας με την ίδια πρακτική, προκύπτει ότι

$$\pi_n = \rho^n \pi_0, \; n \geq 0 \tag{7}$$

- Όμως πρέπει

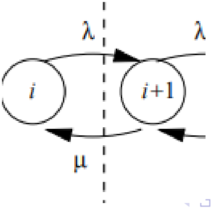
$$1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \sum_{n=0}^{\infty} \rho^n$$

- Για να συγκλίνει η παραπάνω σειρά θα πρέπει $\rho < 1$.
- Άρα όταν $\rho < 1$ (συνθήκη ευστάθειας)

$$1 = \pi_0 \frac{1}{1-\rho} \iff \pi_0 = 1 - \rho \iff \pi_n = (1-\rho)\rho^n.$$

Αναδρομική επίλυση–**Τοπικές** εξισώσεις ισορροπίας (Μέθοδος 2)

- Μια BD process είναι χρονικά αντιστρέψιμη.
- Σε αυτή την περίπτωση μπορούμε να εφαρμόσουμε την μέθοδο των τοπικών εξισώσεων ισορροπίας.
- Έχουν την ίδια ερμηνεία με τις ολικές ε.ι., αλλά εφαρμόζονται σε "συγκεκριμένα σετ" καταστάσεων.
- Προκύπτουν από "κατάλληλη αποδόμηση" των ολικών ε.ι.$\Rightarrow$ Αθροίζοντας τις τοπικές ε.ι. θα πάρουμε τις ολικές ε.ι..
- **Μειονέκτημα:** Δεν υπάρχει συγκεκριμένη μέθοδος επιλογής των "κατάλληλων καταστάσεων" για να τις εφαρμόσεις.



Τοπικές εξισώσεις ισορροπίας

Για την περίπτωση του M/M/1

$$
\begin{aligned}
\lambda\pi_0 &= \mu\pi_1 \Rightarrow \pi_1 = \rho\pi_0 \\
\lambda\pi_1 &= \mu\pi_2 \Rightarrow \pi_2 = \rho\pi_1 = \rho^2\pi_0 \\
\lambda\pi_2 &= \mu\pi_3 \Rightarrow \pi_3 = \rho\pi_2 = \rho^3\pi_0 \\
\vdots & \quad \vdots \\
\lambda\pi_{n-1} &= \mu\pi_n \Rightarrow \pi_n = \rho\pi_{n-1} = \rho^n\pi_0 \\
\vdots & \quad \vdots
\end{aligned}
$$

Όμως

$$1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \sum_{n=0}^{\infty} \rho^n$$

Άρα όταν $\rho < 1$ (συνθήκη ευστάθειας)

$$1 = \pi_0 \frac{1}{1-\rho} \Longleftrightarrow \pi_0 = 1-\rho \Longleftrightarrow \pi_n = (1-\rho)\rho^n.$$

Άμεση επίλυση (Μέθοδος 3)

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1, \\ (\lambda+\mu)\pi_n &= \mu\pi_{n+1} + \lambda\pi_{n-1}, n>0. \end{aligned} \tag{8}$$

- Η γενική λύση της 2ης στην (8) είναι

  $$\pi_n = c_1 x_1^n + c_2 x_2^n, n = 0, 1, 2, \ldots$$

  όπου x_1, x_2 είναι ρίζες της χαρακτηριστικής εξίσωσης

  $$\mu x^2 - (\lambda+\mu)x + \lambda = 0,$$

- Αυτή έχει λύσεις τις $x_1 = 1$, $x_2 = \lambda/\mu = \rho$. Άρα

  $$\pi_n = c_1 + c_2\rho^n, n = 0, 1, 2, \ldots \tag{9}$$

- c_1, c_2: Εφαρμόζοντας την (9) στην 1η της (8):

  $$\lambda(c_1 + c_2) = \mu(c_1 + c_2\rho) \iff c_1 = c_1\rho \iff c_1 = 0.$$

- Άρα $\pi_n = c_2\rho^n$ και αφού $\sum_n \pi_n = 1$

  $$c_2 \sum_n \rho^n = 1 \iff c_2 \frac{1}{1-\rho} = 1 \iff c_2 = 1-\rho \iff \pi_n = (1-\rho)\rho^n.$$

z-μετασχηματισμός (Μέθοδος 4)

Έστω $\Pi(z) = \sum_{n=0}^{\infty} \pi_n z^n$, $|z| \leq 1$. Εφαρμόζοντας στις (8)

$$\begin{aligned} \lambda \pi_0 z^0 &= \mu \pi_1 z^0, \\ (\lambda + \mu) \pi_n z^n &= \mu \pi_{n+1} z^n + \lambda \pi_{n-1} z^n, \; n > 0. \end{aligned}$$

Αθροίζοντας κατά μέλη

$$\lambda \Pi(z) + \mu[\Pi(z) - \pi_0] = \frac{\mu}{z}[\Pi(z) - \pi_0] + \lambda z \Pi(z)$$

Μετά από πράξεις προκύπτει

$$\Pi(z)[(z-1)(1-\rho z)] = \pi_0(z-1) \iff \Pi(z) = \frac{\pi_0}{1 - z\rho}.$$

Όμως $\Pi(1) = 1$ και επομένως $\pi_0 = 1 - \rho$. Άρα

$$\Pi(z) = \frac{1-\rho}{1-z\rho} = (1-\rho) \sum_{n=0}^{\infty} \rho^n z^n = \sum_{n=0}^{\infty} (1-\rho) \rho^n z^n.$$

Επομένως $\pi_n = (1-\rho)\rho^n$,

Μέτρα απόδοσης

- Αναμενόμενος αριθμός πελατών στο σύστημα.

$$E(N) = \sum_{n=0}^{\infty} n\pi_n = \frac{d}{dz}\Pi(z)|_{z=1} = (1-\rho)\frac{\rho}{(1-\rho)^2} = \frac{\rho}{1-\rho}.$$

Ισοδύναμα

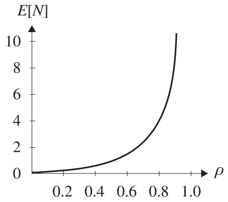
$$\begin{aligned} E(N) = & \textstyle\sum_{n=0}^{\infty} n\pi_n = (1-\rho)\sum_{n=0}^{\infty} n\rho^n \\ = & \textstyle(1-\rho)\rho\sum_{n=0}^{\infty} \frac{d}{d\rho}\rho^n = (1-\rho)\rho\frac{d}{d\rho}\sum_{n=0}^{\infty}\rho^n \\ = & \textstyle(1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) = (1-\rho)\rho\frac{1}{(1-\rho)^2} = \frac{\rho}{1-\rho}. \end{aligned}$$

- Αναμενόμενος αριθμός πελατών στην ουρά: Τότε $\bar{\pi}_0 = \pi_0 + \pi_1$, $\bar{\pi}_n = \pi_n$

$$E(N_q) = \sum_n n\bar{\pi}_n = \ldots = \frac{\rho^2}{1-\rho}. \qquad (10)$$

- server utilization: $U = \sum_{n=1}^{\infty} \pi_n = 1 - \pi_0 = \rho$
  $\Rightarrow$ Το ποσοστό του χρόνου που ο υπάλληλος είναι απασχολημένος.

Μέτρα απόδοσης (συνεχ.)



Plot of the expected number of customers in the M/M/1 system vs. ρ.

- Throughput: $\mu \sum_{n=1}^{\infty} \pi_n = \mu(1-\pi_0) = \mu\rho = \lambda.$
- Αναμενόμενος χρόνος **παραμονής** στο σύστημα: Από νόμο του Little

$$E(S) = \frac{E(N)}{\lambda} = \frac{1}{\mu(1-\rho)}.$$

- Αναμενόμενος χρόνος **αναμονής** ενός πελάτη στην ουρά:

$$E(W) = E(S) - E(B) = \frac{1}{\mu(1-\rho)} - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}.$$

Ανάλυση Μέσης Τιμής

- $E(N)$, $E(S)$ μπορούν να υπολογιστούν με χρήση Ν. Little και Αρχής PASTA (Poisson Arrivals See Time Averages).
- Έστω $t_k \equiv$ χρ. στιγμή άφιξης k–οστού πελάτη
- $a_n = \lim_{k\to\infty} P(N(t_k^-) = n) \equiv$ η οριακή πιθανότητα να βρει κάποιος πελάτης φθάνοντας n πελάτες
  $\Rightarrow$ οριακή κατανομή σε στιγμές αφίξεων.
- **Αρχή PASTA**: Αν οι πελάτες καταφθάνουν σύμφωνα με την διαδικασία Poisson και οι μελλοντικές αφίξεις δεν επηρεάζονται από την κατάσταση του συστήματος, τότε

$$a_n = \pi_n.$$

- Υποθέτουμε πειθαρχία FCFS.
- Το S εξαρτάται από τον αριθμό των πελατών που βρίσκει φθάνοντας ένας πελάτης

Ανάλυση Μέσης Τιμής (συνεχ.)

- Έστω $N^a \equiv$ αρ. πελατών στο σύστημα σε στιγμές αφίξεων, δηλ. $N^a = \lim_{k\to\infty} N(t_k^-)$.
- Τότε,

$$\begin{aligned} E(S) = & \sum_{n=0}^{\infty} E(S|N^a = n)P(N^a = n) \\ = & \sum_{n=0}^{\infty} \frac{n+1}{\mu} P(N = n) \text{ (λόγω αρχής PASTA)} \\ = & \frac{1}{\mu}[\sum_{n=0}^{\infty} n\pi_n + 1] = \frac{1}{\mu}[E(N) + 1]. \end{aligned}$$

- Από N. Little $E(N) = \lambda E(S)$. Συνδυάζοντας

$$E(S) = \frac{1}{\mu}[E(N) + 1] = \frac{1}{\mu}[\lambda E(S) + 1] = \rho E(S) + \frac{1}{\mu}$$

$$\Rightarrow E(S) = \frac{1}{\mu(1-\rho)} \text{ και επομένως } E(N) = \frac{\rho}{1-\rho}.$$

Κατανομή του χρόνου παραμονής (Sojourn time)

- Το S εξαρτάται από τον αριθμό των πελατών που βρίσκει φθάνοντας ο target πελάτης.
- Έστω B_k ο χρόνος εξυπηρέτησης του k-οστού πελάτη, $B_k \sim \exp(\mu)$.
- Ο target πελάτης, αν βρει $j(>0)$ πελάτες κατά την άφιξη του, πρέπει να περιμένει: $j-1$ "ολόκληρους" χρόνους εξυπηρέτησης + τον "υπολοιπόμενο" χρόνο εξυπηρέτησης του πελάτη που εξυπηρετείται + τον δικό του χρ. εξυπηρέτησης.
- Αφού οι χρόνοι εξυπηρέτησης είναι $\exp(\mu)$:

$$S = \sum_{k=1}^{N^a+1} B_k.$$

Κατανομή του χρόνου παραμονής (Sojourn time) (συνεχ.)

- Δεσμεύοντας στον αριθμό N^a των πελατών που βρίσκει φθάνοντας.
- Λαμβάνοντας υπόψιν την ανεξαρτησία των N^a, B_k,

$$\begin{aligned} P(S > t) &= P(\textstyle\sum_{k=1}^{N^a+1} B_k > t) \\ &= \textstyle\sum_{n=0}^{\infty} P(\sum_{k=1}^{n+1} B_k > t | N^a = n) P(N^a = n) \end{aligned}$$

- Λόγω PASTA $P(N^a = n) = P(N = n) = (1-\rho)\rho^n$.
- $\sum_{k=1}^{n+1} B_k \sim Gamma(n+1, \mu)$ με $P(\sum_{k=1}^{n+1} B_k > t) = \sum_{k=0}^{n} e^{-\mu t} \frac{(\mu t)^k}{k!}$.
- Τότε

$$\begin{aligned} P(S > t) &= \textstyle\sum_{n=0}^{\infty} \sum_{k=0}^{n} e^{-\mu t} \frac{(\mu t)^k}{k!} (1-\rho)\rho^n \\ &= \textstyle\sum_{k=0}^{\infty} \sum_{n=k}^{\infty} e^{-\mu t} \frac{(\mu t)^k}{k!} (1-\rho)\rho^n \\ &= \textstyle\sum_{k=0}^{\infty} e^{-\mu t} \frac{(\mu \rho t)^k}{k!} \sum_{n=k}^{\infty} (1-\rho)\rho^{n-k} \\ &= \textstyle\sum_{k=0}^{\infty} e^{-\mu t} \frac{(\mu \rho t)^k}{k!} = e^{-\mu(1-\rho)t}. \end{aligned}$$

- Άρα $P(S \le t) = 1 - e^{-\mu(1-\rho)t}$ και $S \sim \exp(\mu(1-\rho))$!!!

Sojourn time distribution (εναλλακτικά)

- Χρήση Laplace μετασχηματισμών.
  $S^*(s) = E(e^{-sS}) = \int_0^\infty e^{-st} dS(t)$, με $S(t) = P(S \le t)$. Τότε,

$$\begin{aligned} S^*(s) = &\ E(e^{-sS}) = E(e^{-s\sum_{k=1}^{N^a+1} B_k}) \\ = &\ \sum_{n=0}^\infty E(e^{-s\sum_{k=1}^{n+1} B_k}) P(N^a = n) \\ = &\ \sum_{n=0}^\infty E(e^{-s\sum_{k=1}^{n+1} B_k}) P(N = n) \ (\text{λόγω PASTA}) \\ = &\ \sum_{n=0}^\infty E(e^{-sB_1}) \times \ldots \times E(e^{-sB_{n+1}})(1-\rho)\rho^n \\ &\ (\text{ανεξαρτησία των } B_k) \end{aligned}$$

- Αφού $B_k \sim \exp(\mu)$, τότε $B^*(s) = E(e^{-sB_k}) = \frac{\mu}{\mu+s}$. Άρα

$$\begin{aligned} S^*(s) = &\ \sum_{n=0}^\infty \left(\frac{\mu}{\mu+s}\right)^{n+1} (1-\rho)\rho^n = \frac{\mu(1-\rho)}{\mu+s} \sum_{n=0}^\infty \left(\frac{\mu\rho}{\mu+s}\right)^n \\ = &\ \frac{\mu(1-\rho)}{\mu+s} \frac{1}{1-\frac{\mu\rho}{\mu+s}} = \frac{\mu(1-\rho)}{s+\mu(1-\rho)}. \end{aligned}$$

- Άρα $S \sim \exp(\mu(1-\rho))$!!!

Κατανομή του χρόνου αναμονής (Waiting time)

- Παρατήρησε ότι $S = W + B$, όπου $B \equiv$ ο χρόνος εξυπηρέτησης.
- Αφού οι W, B είναι ανεξάρτητες τ.μ., τότε

$$S^*(s) = W^*(s)B^*(s) = W^*(s)\frac{\mu}{s+\mu}.$$

- Άρα

$$W^*(s) = \frac{\mu+s}{\mu} \times \frac{\mu(1-\rho)}{s+\mu(1-\rho)} = (1-\rho) \times 1 + \rho \times \frac{\mu(1-\rho)}{s+\mu(1-\rho)}.$$

- Άρα ο W είναι με πιθανότητα $1-\rho$ ίσος με 0 και με πιθανότητα ρ ίσος με μια εκθετικά κατανεμημένη τ.μ. X με παράμετρο $\mu(1-\rho)$.
- Οπότε

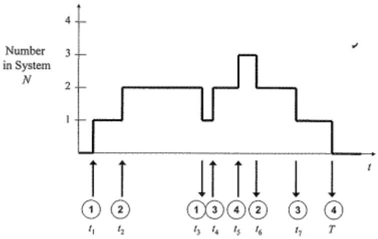
$$P(W > t) = 0 \times (1-\rho) + \rho \times P(X > t) = \rho e^{-\mu(1-\rho)t}.$$

- Παρατήρησε ότι

$$E(W) = -\frac{d}{ds}W^*(s)|_{s=0} = \frac{\rho\mu}{\mu(1-\rho)}.$$

Busy period

- Η "ζωή" ενός συστήματος εναλλάσσεται μεταξύ μεταξύ περιόδων συνεχούς απασχόλησης (busy period (BP)) και περιόδων αεργίας (idle period (IP)).
- Busy cycle = BP+IP
- $E(IP) = 1/\lambda$ (ΓΙΑΤΙ;)
- Busy cycle: περίοδος από την στιγμή που ένας πελάτης φθάνοντας βρίσκει 'άδειο" το σύστημα και αρχίζει να εξυπηρετείται μέχρι την αμέσως επόμενη τέτοια στιγμή.



Mean busy period

- Σε χρονικές στιγμές έναρξης κύκλου απασχόλησης, το σύστημα 'άναγεννάται" πιθανοθεωρητικά.
- Δηλ. η μελλοντική του συμπεριφορά είναι πιθανοθεωρητικά ένα 'άντίγραφο" της προηγούμενης.
- Άρα

$$\rho = \frac{E(BP)}{E(BP) + E(IP)} = \frac{E(BP)}{E(BP) + 1/\lambda} \quad \textbf{(ΓΙΑΤΙ;)}$$

- Οπότε,

$$E(BP) = \frac{1}{\mu(1-\rho)}.$$

Κατανομή της busy period

- C_n : Ο χρόνος μέχρι να αδειάσει το σύστημα όταν ένας πελάτης φθάνοντας βρει $n-1$ πελάτες.
- Άρα $BP = C_1$. Ψαχνω την κατανομή της C_1, $c_1(t)$.

Τότε για $n > 0$

$$C_n = X + \begin{cases} C_{n+1}, & \textit{with prob. } \frac{\lambda}{\lambda+\mu}, \\ C_{n-1}, & \textit{with prob. } \frac{\mu}{\lambda+\mu}, \end{cases}$$

με $X \equiv$ Χρόνος παραμονής στην κατάσταση n και $X \sim \exp(\lambda+\mu)$. Αν

$$c_n^*(s) = \int_0^\infty e^{-st} c_n(t)dt, f_X^*(s) = \int_0^\infty e^{-st} f_X(t)dt = \frac{\lambda+\mu}{s+\lambda+\mu}.$$

Απο ιδιότητες *Laplace* μετασχηματισμών

$$\begin{aligned} c_n^*(s) = &\ \tfrac{\lambda+\mu}{s+\lambda+\mu}[\tfrac{\lambda}{\lambda+\mu}c_{n+1}^*(s) + \tfrac{\mu}{\lambda+\mu}c_{n-1}^*(s)] \Longleftrightarrow \\ (s+\lambda+\mu)c_n^*(s) = &\ \lambda c_{n+1}^*(s) + \mu c_{n-1}^*(s). \end{aligned}$$

με γενική λύση

$$c_n^*(s) = d_1 x_1^n(s) + d_2 x_2^n(s),$$

όπου τα $x_{1,2}(s)$ με $x_1(s) \leq 1 < x_2(s)$, λύσεις της

$$\lambda x^2 - (s+\lambda+\mu)x + \mu = 0.$$

Αφού $0 \leq c_n^*(s) \leq 1$ πρέπει $d_2 = 0$. Επιπλέον μιας και $C_0 = 0$, τότε $c_0^*(s) = 1$. Επομένως $c_n(s) = x_1^n(s)$. Άρα,

$$\begin{aligned} c_1^*(s) = &\ x_1(s) = \tfrac{s+\lambda+\mu-\sqrt{(s+\lambda+\mu)^2-4\lambda\mu}}{2\lambda}, \\ E(BP) = &\ -\tfrac{d}{ds}c_1^*(s)|_{s=0} = \tfrac{1}{\mu(1-\rho)}. \end{aligned}$$

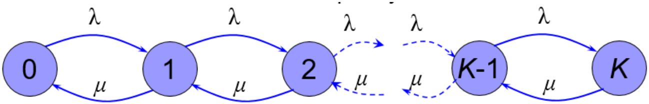
M/M/1/K (finite capacity)

Οι διαδικασίες αφίξεων και εξυπηρετήσεων είναι Poisson
$\Rightarrow$ Birth & Death process.

1. **Αφίξεις:** Poisson, $\lambda > 0$.
2. **Εξυπηρέτηση:** $exp(\mu)$, $\mu > 0$.
3. Πεπερασμένη χωρητικότητα K.
4. Διαδικασία αφίξεων-εξυπηρετήσεων ανεξάρτητες.
5. $N(t)$: αριθ. πελατών στο σύστημα την χρ. στιγμη t.
6. Η $\{N(t) : t \geq 0\}$ είναι BD process με $\lambda_n = \lambda$, $\mu_n = \mu$, $n \in \{1, \ldots, K\}$. Με χρήση τοπικών εξισώσεων ισορροπίας

$$\begin{aligned} \lambda\pi_j = &\quad \mu\pi_{j+1}, j = 0, 1, \ldots, K-1 \\ \Rightarrow \pi_j = &\quad \left(\frac{\lambda}{\mu}\right)^j \pi_0 = \rho^j \pi_0, j = 0, 1, \ldots, K, \end{aligned}$$

7. Οπότε $1 = \sum_{j=0}^{K} \pi_j = \pi_0 \sum_{j=0}^{K} \rho^j \Rightarrow \pi_0 = \frac{1-\rho}{1-\rho^{K+1}}$.



Μέτρα απόδοσης

- Server Utilization

$$E[U] = 1 - \pi_0 = 1 - \frac{(1-\rho)}{1-\rho^{K+1}} = \frac{\rho\left(1-\rho^K\right)}{1-\rho^{K+1}}$$

- Throughput

$$E[R] = \mu(1-\pi_0) = \lambda\frac{1-\rho^K}{1-\rho^{K+1}} < \lambda$$

- Blocking Probability

$$P_B = \pi_K = \frac{(1-\rho)\rho^K}{1-\rho^{K+1}}$$

Probability that an arriving customer finds the queue full (at state K)

Μέτρα απόδοσης

- Expected Queue Length

$$E[X]=\sum_{j=0}^{K} j\pi_j=\frac{(1-\rho)}{1-\rho^{K+1}}\sum_{j=0}^{K} j\rho^j=\frac{(1-\rho)\rho}{1-\rho^{K+1}}\sum_{j=0}^{K}\frac{d\{\rho^j\}}{d\rho}=$$

$$=\frac{(1-\rho)\rho}{1-\rho^{K+1}}\frac{d}{d\rho}\left\{\sum_{j=0}^{K}\rho^j\right\}=\frac{(1-\rho)\rho}{1-\rho^{K+1}}\frac{d}{d\rho}\left\{\frac{1-\rho^{K+1}}{(1-\rho)}\right\}=$$

$$=\frac{(1-\rho)\rho}{1-\rho^{K+1}}\left(\frac{(1-\rho^{K+1})-(1-\rho)(K+1)\rho^K}{(1-\rho)^2}\right)=$$

$$=\frac{\rho}{1-\rho^{K+1}}\left(\frac{1-\rho^K}{1-\rho}-K\rho^K\right)$$

- System time

Net arrival rate (no losses)

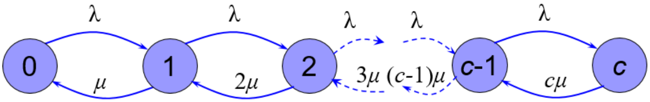
$$E[X]=\lambda(1-\pi_K)E[S]$$

M/M/c/c

- Αφίξεις Poisson, c υπάλληλοι, χρόνοι εξυπηρέτησης $\exp(\mu)$.
- Χωρητικότητα συστήματος $c \Rightarrow$ Δεν υπάρχει ουρά.
- $N(t)$: αριθ. πελατών στο σύστημα την χρ. στιγμη t.
- Η $\{N(t) : t \geq 0\}$ είναι BD process με $\lambda_n = \lambda$, $\mu_n = n\mu$, $n \in \{0, 1, \ldots, c\}$. Με χρήση τοπικών εξισώσεων ισορροπίας

$$\begin{aligned} \lambda\pi_j &= (j+1)\mu\pi_{j+1}, j = 0, 1, \ldots, c \\ \Rightarrow \pi_j &= \left(\frac{\lambda}{\mu}\right)^j \frac{\pi_0}{j!}, j = 0, 1, \ldots, c, \end{aligned}$$

- Οπότε
$1 = \sum_{j=0}^{c} \pi_j = \pi_0 \sum_{j=0}^{c} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} \Rightarrow \pi_0 = [\sum_{j=0}^{c} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!}]^{-1}$.



Erlang B formula

- Το σημαντικότερο μέτρο απόδοσης είναι η πιθανότητα blocking ενός πελάτη.
- Αποτελεί το απλούστερο μοντέλο για call center (Erlang (1917)).
- Τότε από αρχή PASTA, αν $R = \frac{\lambda}{\mu}$ (offered load)

$$P(blocking) = \pi_c = B(c,R) = \frac{\frac{R^c}{c!}}{\sum_{j=0}^{c} \frac{R^j}{j!}}.$$

- Δύσκολος ο υπολογισμός για μεγάλο c.
- Χρήση του ακόλουθου αλγόριθμου:

$$B(c,R) = \frac{\frac{R^c}{c!}}{\sum_{j=0}^{c} \frac{R^j}{j!}} = \frac{\frac{R^{c-1}}{(c-1)!}\frac{R}{c}}{\sum_{j=0}^{c-1} \frac{R^j}{j!} + \frac{R^c}{c!}} = \frac{\frac{R^c}{c!}}{\sum_{j=0}^{c} \frac{R^j}{j!}} = \frac{B(c-1,R)\frac{R}{c}}{1+\frac{R}{c}B(c-1,R)} = \frac{B(c-1,R)R}{1+RB(c-1,R)},$$

ξεκινώντας με $B(0,c) = 1$.

* https://www.erlang.com/calculator/erlb/

Μέτρα απόδοσης

- Ποσοστό “χαμένων” πελατών: $\lambda\pi_c$ (PASTA)
- Πραγματικός ρυθμός εισόδου (Throughput):

$$\lambda(1-\pi_c) = \lambda(1-B(c,R)) < \lambda.$$

- Mean number of customers in the systems, mean number of busy servers:

$$\begin{aligned} E(N) = \ & \textstyle\sum_{j=1}^{c} j\pi_j = \sum_{j=1}^{c} j\frac{R^j}{j!}\pi_0 = R\sum_{j=1}^{c}\frac{R^{j-1}}{(j-1)!}\pi_0 \\ & = R\pi_0 \textstyle\sum_{m=0}^{c-1}\frac{R^m}{m!} = R(1-\pi_c) = R(1-B(c,R)). \end{aligned}$$

- Little's law: Μέσος χρόνος παραμονής στο σύστημα

$$E(S) = \frac{E(N)}{\lambda(1-B(c,R))}.$$

Παράδειγμα

Customers arrive according to a Poisson process at a parking lot near a small shopping center with a rate of 60 cars per hour. The parking time is exponentially distributed with mean 2.5 hours and the parking lot offers place to 150 cars. When the parking lot is full, an arriving customer has to park his car somewhere else. Find the fraction of customers finding all places occupied on arrival.[1]

c	$B(c, 150)$
150	0.062
155	0.044
160	0.028
165	0.017
170	0.009

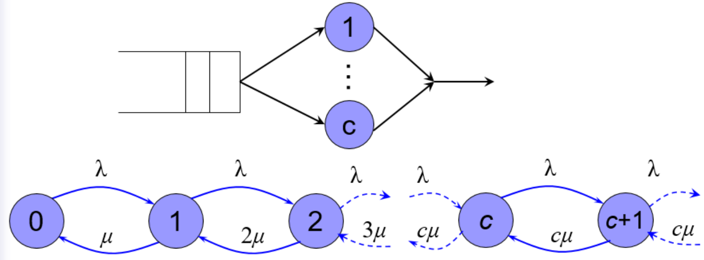
Σχήμα 1: Table 1.

[1] *Hint: Table 1 contains the blocking probability $B(c, R)$ for $R = 150$ and several values of c.*

M/M/c

Οι διαδικασίες αφίξεων και εξυπηρετήσεων είναι Poisson
$\Rightarrow$ Birth & Death process.

1. $c > 1$ υπάλληλοι.
2. **Αφίξεις:** Poisson, $\lambda > 0$
3. **Εξυπηρέτηση *σε κάθε* υπάλληλο:** $exp(\mu)$, $\mu > 0$
4. Άπειρη χωρητικότητα.
5. Διαδικασία αφίξεων-εξυπηρετήσεων ανεξάρτητες.
6. FIFO (FCFS) (First In First Out or First Come First Served)
7. Υποθέτουμε ότι $\rho = \frac{\lambda}{c\mu} = \frac{R}{c} < 1$ (**ΓΙΑΤΙ;**)



Κατανομή του αριθμού των πελατών στο σύστημα

- $N(t) \equiv$ αριθμός πελατών στο σύστημα την χρ. στιγμή t.
- Το $M/M/c$ είναι ειδική περίπτωση BD process με

$$\begin{aligned} \lambda_n = &\ \lambda, n \geq 0, \\ \mu_n = &\ min(n,c)\mu = \begin{cases} n\mu, & n = 1, \ldots, c-1, \\ c\mu, & n \geq c. \end{cases} \end{aligned}$$

- Ολικές εξισώσεις ισορροπίας

$$\begin{aligned} \lambda\pi_0 = &\ \mu\pi_1, \\ (\lambda + n\mu)\pi_n = &\ \lambda\pi_{n-1} + (n+1)\mu\pi_{n+1}, n = 1, \ldots, c-1, \\ (\lambda + c\mu)\pi_n = &\ \lambda\pi_{n-1} + c\mu\pi_{n+1}, n \geq c. \end{aligned}$$

- Αναδρομική επίλυση (οριακή συμπεριφορά BD processes)

$$\begin{aligned} \pi_n = &\ \begin{cases} \left(\frac{\lambda}{\mu}\right)^n \frac{\pi_0}{n!}, & n = 0, \ldots, c-1, \\ \left(\frac{\lambda}{\mu}\right)^n \frac{\pi_0}{c!c^{n-c}}, & n \geq c, \end{cases} \\ \pi_0 = &\ [\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)}]^{-1}. \end{aligned}$$

Τοπικές εξισώσεις ισορροπίας (LBE)

- Εφαρμόζοντας τις LBE

$$
\begin{aligned}
\lambda\pi_0 &= \mu\pi_1 \Rightarrow \pi_1 = \tfrac{\lambda}{\mu}\pi_0 = R\pi_0 \\
\lambda\pi_1 &= 2\mu\pi_2 \Rightarrow \pi_2 = \tfrac{\lambda}{2\mu}\pi_1 = R^2\tfrac{\pi_0}{2} \\
\lambda\pi_2 &= 3\mu\pi_3 \Rightarrow \pi_3 = \tfrac{\lambda}{3\mu}\pi_2 = R^3\tfrac{\pi_0}{2\times 3} \\
&\vdots \quad \vdots \\
\lambda\pi_{c-1} &= c\mu\pi_c \Rightarrow \pi_c = \tfrac{\lambda}{c\mu}\pi_{c-1} = \ldots = R^c\tfrac{\pi_0}{2\times 3\times\ldots\times c} \\
\lambda\pi_c &= c\mu\pi_{c+1} \Rightarrow \pi_{c+1} = \tfrac{\lambda}{c\mu}\pi_c = \ldots = R^{c+1}\tfrac{\pi_0}{cc!} \\
\lambda\pi_{c+1} &= c\mu\pi_{c+2} \Rightarrow \pi_{c+2} = \tfrac{\lambda}{c\mu}\pi_{c+1} = \left(\tfrac{\lambda}{c\mu}\right)^2\pi_c \ldots = R^{c+2}\tfrac{\pi_0}{c^2c!} \\
&\vdots \quad \vdots \\
&\vdots \quad \vdots
\end{aligned}
$$

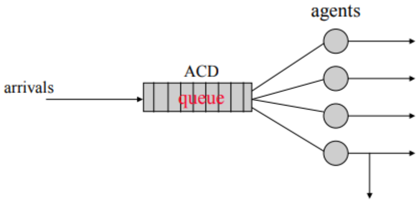
LBE (συνεχ.)

- Οπότε
$$\pi_n = \begin{cases} R^n \frac{\pi_0}{n!}, & n = 0, \ldots, c-1, \\ R^n \frac{\pi_0}{c! c^{n-c}}, & n \geq c, \end{cases}$$
- $R = \lambda/\mu$ είναι η προσφερόμενη ροή πελατών $\Rightarrow$ αναμενόμενος χρόνος εργασίας που φθάνει στη μονάδα του χρόνου.
- Όμως
$$\begin{aligned} 1 &= \sum_{n=0}^{\infty} \pi_n = \pi_0 \left[\sum_{n=0}^{c-1} \frac{R^n}{n!} + \sum_{n=c}^{\infty} \frac{R^n}{c^{n-c} c!} \right] \\ &= \pi_0 \left[\sum_{n=0}^{c-1} \frac{R^n}{n!} + \frac{R^c}{c!} \sum_{n=c}^{\infty} \left(\frac{R}{c} \right)^{n-c} \right] \\ &= \pi_0 \left[\sum_{n=0}^{c-1} \frac{R^n}{n!} + \frac{R^c}{c!} \sum_{n=c}^{\infty} \rho^{n-c} \right] \quad (\text{επειδή } \rho < 1) \\ &= \pi_0 \left[\sum_{n=0}^{c-1} \frac{R^n}{n!} + \frac{R^c}{c!} \frac{1}{1-\rho} \right]. \end{aligned}$$
- Άρα
$$\pi_0 = \left[\sum_{n=0}^{c-1} \frac{R^n}{n!} + \frac{R^c c}{c!(c-R)}. \right]^{-1}.$$

Erlang C formula

- $C(c,R) = P(W > 0) \equiv$ η πιθανότητα ένας πελάτης φθάνοντας να βρει όλους τους υπάλληλους απασχολημένους $\equiv$ Blocking probability.
- Βασικό μέτρο απόδοσης όταν το M/M/c χρησιμοποιείται για την μοντελοποίηση **call centers**.
- Τότε, από αρχή PASTA

$$
\begin{aligned}
C(c,R) &= \sum_{n=c}^{\infty} \pi_n \\
&= \pi_c[1+\rho+\rho^2+\ldots] = \frac{\pi_c}{1-\rho} \quad (\text{αφού } \rho < 1) \\
&= \frac{\frac{R^c}{c!}\frac{c}{c-R}}{\sum_{n=0}^{c-1}\frac{R^n}{n!}+\frac{R^c c}{c!(c-R)}}.
\end{aligned}
$$



Erlang C formula: Stable recursion scheme

- Κοστοβόρος (υπολογιστικά) ο υπολογισμός του $C(c, R)$ για μεγάλο c, εξαιτίας των όρων $R^n/n!$.
- Μπορούμε να υπολογίσουμε το $C(c, R)$ με τον ακόλουθο αναδρομικό αλγόριθμο:

1. Έχοντας την

$$C(c,R) = \frac{\frac{R^c}{c!}\frac{c}{c-R}}{\sum_{n=0}^{c-1}\frac{R^n}{n!}+\frac{R^c c}{c!(c-R)}} = \frac{\frac{R^{c-1}}{c-1!}\frac{c-1}{c-1-R}\left[\frac{R(c-1-R)}{(c-1)(c-R)}\right]}{\sum_{n=0}^{c-2}\frac{R^n}{n!}+\frac{R^{c-1}(c-1)}{(c-1)!(c-1-R)}-\frac{R^{c-1}}{c-1!}\frac{c-1}{c-1-R}\frac{R}{(c-1)(c-R)}},$$

2. Διαίρεσε τον αριθμητή και παρονομαστή με την ποσότητα $\sum_{n=0}^{c-2}\frac{R^n}{n!}+\frac{R^{c-1}(c-1)}{(c-1)!(c-1-R)}$.
3. Τότε

$$C(c,R) = \frac{R(c-1-R)C(c-1,R)}{(c-1)(c-R)-R\times C(c-1,R)}, \; c \geq 2, \qquad (11)$$

με $C(1,R) = \rho$.

Erlang C through Erlang B

- Υπολογισμός της Erlang C formula μεσω της Erlang B.
- Πράγματι

$$C(c,R) = \frac{\frac{R^c}{c!}\frac{c}{c-R}}{\sum_{n=0}^{c-1}\frac{R^n}{n!}+\frac{R^c c}{c!(c-R)}} = \frac{\frac{R^c}{c!}\frac{c}{c-R}}{\sum_{n=0}^{c}\frac{R^n}{n!}+\frac{R^c}{c!}[1-\frac{c}{c-R}]} = \frac{B(c,R)\frac{c}{c-R}}{1+\frac{R}{c-R}B(c,R)}$$

$$= \frac{B(c,R)c}{c-R+R\times B(c,R)}$$

$$= \frac{R\times B(c-1,R)}{c-R+R\times B(c-1,R)},$$

με $B(0,R)=1$.

- $C(c,R) > B(c,R)$

** Erlang C calculator: http://www.gerkoole.com/CCO/erlang-c.php

** Erlang C calculator: δες το σχετικο excel add-in στην παραπάνω σελίδα καθώς και στο e-class, folder Excel sheets.

** Erlang C calculator: https://www.erlang.com/calculator/erlc/

Waiting time distribution

- Παρατήρησε ότι

$$W = \sum_{k=1}^{N^a+1} D_k, \ \ (\text{ΓΙΑΤΙ;})$$

όπου $D_k \sim \exp(c\mu)$, και N^a ο αριθμός των πελατών στο σύστημα σε στιγμές αφίξεων.

- Τότε, $\sum_{k=1}^{n+1} D_k \sim Gamma(n+1, c\mu)$ και

$$\begin{aligned}
P(W > t) = & \sum_{n=0}^{\infty} P(\sum_{k=1}^{n+1} D_k > t | N^a = n+c) P(N^a = n+c) \\
= & \sum_{n=0}^{\infty} \sum_{k=0}^{n} e^{-c\mu t} \frac{(c\mu t)^k}{k!} \pi_{n+c} \\
= & \sum_{n=0}^{\infty} \sum_{k=0}^{n} e^{-c\mu t} \frac{(c\mu t)^k}{k!} \pi_c \rho^n \\
= & \pi_c \sum_{k=0}^{\infty} e^{-c\mu t} \frac{(c\mu t)^k}{k!} \sum_{n=k}^{\infty} \rho^n \\
= & \pi_c \sum_{k=0}^{\infty} e^{-c\mu t} \frac{(c\mu \rho t)^k}{k!} \sum_{n=k}^{\infty} \rho^{n-k} \\
= & \frac{\pi_c}{1-\rho} \sum_{k=0}^{\infty} e^{-c\mu t} \frac{(c\mu \rho t)^k}{k!} = \frac{\pi_c}{1-\rho} e^{-c\mu(1-\rho)t}
\end{aligned}$$

Waiting time distribution (συνέχ.)

- Οπότε,

$$P(W > t) = C(c, c\rho)e^{-c\mu(1-\rho)t} = C(c, R)e^{-\mu(c-R)t}.$$

- $P(W > t) \equiv$ η πιθανότητα να περιμένουμε πάνω από t μονάδες του χρόνου για δεδομένες παραμέτρους c, R.
- Επιπλέον,

$$P(W > t | W > 0) = \frac{P(W > t)}{P(W > 0)} = e^{-\mu(c-R)t}.$$

οπότε $W | W > 0 \sim \exp(\mu(c - R))$.

Call center Performance Metrics

- Telephone Service Factor (TSF) or Service Level (SL) $\equiv$ ποσοστό κλήσεων με χρόνο αναμονής κάτω από ένα προκαθορισμένο επίπεδο (target service-level) T.

$$P(W < T) = 1 - C(c, R)e^{-\mu(c-R)T}$$

- Average Speed of Answer (ASA) $\equiv$ μέσος χρόνος αναμονής μέχρι να ξεκινήσει η εξυπηρέτηση.

$$E(Wait) = P(W > 0)E(W|W > 0) = C(c, R)E(W|W > 0) = \frac{C(c, R)}{\mu(c - R)}.$$

- Γνησίως μονότονα ως προς c: $C(c, R)$ φθίνουσα, ASA φθίνουσα και TSF αύξουσα.

Call Center: Workforce management (WFM)

- Workforce management: Καθορισμός αριθμού υπαλλήλων ώστε να ικανοποιηθεί συγκεκριμένο επίπεδο παροχής υπηρεσιών.
- SLA: Service Level Agreement

Step 1: Πρόβλεψη αφίξεων στο target χρονικό διάστημα (Forecasting the Arrivals).

Step 2: Καθορισμός του staffing level: Χρήση της Erlang C formula. $\Rightarrow$ π.χ., Βρες το $c_{opt} = \{c : TSF \leq TSF^*\}$, όπου TSF^* ένα προκαθορισμένο επίπεδο αποδεκτού ποσοστου πελατών που θα πρέπει να περιμένουν για κάποιο δοθέν χρόνικό όριο.

Example

Consider a call center with a forecast (FC) of 100 for the 10:00-10:15 interval. The average call handling time (AHT) is 3.5 min. Find the minimum number of agents required to answer 80% of calls within 20 seconds.

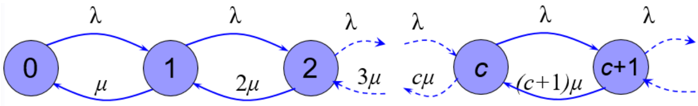
Λύση:

- Εκφράζουμε όλες τις ποσότητες στην ίδια μον. χρόνου (minute).
- Άρα: $\lambda = \frac{100}{15} = 6.66$, $1/\mu = 3.5$ και $R = 6.66 \times 3.5 = 23.33$.
- Χρειαζόμαστε τουλάχιστον 24 υπαλλήλους.
- Όταν $c = 24$, $P(W < 20/60) \approx 1 - C(24, 23.33) \approx 21\%$.
- Δηλ. περίπου το 21% των κλήσεων αναμένει κάτω από 20 sec.
- Αυξάνοντας έναν έναν τον αρ. των υπαλλήλων καταλλήγουμε ότι με 28 υπαλλήλους εξασφαλίζουμε το ζητούμενο service level.

M/M/∞

Οι διαδικασίες αφίξεων και εξυπηρετήσεων είναι Poisson
$\Rightarrow$ Birth & Death process.

1. $N(t) \equiv$ αρ. πελατών στο σύστημα την χρ. στ. t.
2. $\{N(t) : t \geq 0\}$ BD process.
3. $c = \infty$ υπάλληλοι.
4. **Αφίξεις:** Poisson, $\lambda > 0$
5. **Εξυπηρέτηση *σε κάθε* υπάλληλο:** $exp(\mu)$, $\mu > 0$
6. Διαδικασία αφίξεων-εξυπηρετήσεων ανεξάρτητες.



- Τι μπορούμε να πούμε για την ευστάθεια του συστήματος;
- Έστω $\pi_j = \lim_{t\to\infty} P(N(t) = j)$.

- Το M/M/∞ ειναι BD process με

$$\lambda_n = \lambda, \quad \mu_n = n\mu, n \geq 0$$

- Οι τοπικές εξισώσεις ισορροπίας είναι

$$\lambda \pi_n = (n+1)\mu \pi_{n+1}, n \geq 0.$$

- Η αναδρομική επίλυση τους θα δώσει

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \frac{\pi_0}{n!}, n \geq 0.$$

- Τότε

$$1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} = \pi_0 e^{\lambda/\mu} \Rightarrow \pi_0 = e^{-\frac{\lambda}{\mu}}.$$

- Άρα

$$\pi_n = e^{-\frac{\lambda}{\mu}} \frac{(\lambda/\mu)^n}{n!}, \; n \geq 0.$$

- $N \sim Poisson(\frac{\lambda}{\mu})$.

Μέτρα απόδοσης

- System utilization: $E(U) = 1 - \pi_0 = e^{-\frac{\lambda}{\mu}}$,
- Throughput: $\lambda \sum_{n \geq 0} \pi_n = \lambda$,
- Μέσος αριθμός πελατών στο σύστημα:

$$E(N) = \sum_{n=0}^{\infty} n\pi_n = e^{-\frac{\lambda}{\mu}} \sum_{n=0}^{\infty} n \frac{(\lambda/\mu)^n}{n!} = \frac{\lambda}{\mu} e^{-\frac{\lambda}{\mu}} \sum_{n=1}^{\infty} \frac{(\lambda/\mu)^{n-1}}{(n-1)!} = \frac{\lambda}{\mu}$$

  $E(N)$ είναι επίσης ο μέσος αριθμός απασχολημένων υπαλλήλων.
- Από N. Little

$$E(S) = \frac{E(N)}{\lambda} = \frac{1}{\mu} = \text{ Μέσος χρόνος εξυπηρέτησης}$$

$\Rightarrow$ No queueing in M/M/∞.

Staffing: Square root staffing rule

- For an M/M/c with **average arrival rate** λ **and service rate** μ, the resource requirement is denoted by $R = \lambda/\mu$ can be viewed as the minimum number of servers needed to keep the system stable, or as the expected number of servers that are busy, or as the expected number of jobs in service.
- We now argue that, if R is a large number, then by using relatively few servers more, say $c = R + \sqrt{R}$, we can get $P(W > 0)$ down below the 20% range. Our argument is based on the M/M/∞ queue
- **Question:** What is the probability of having more than $R + \sqrt{R}$ jobs in the M/M/∞?
  Answer: The number of jobs in the M/M/∞ is **Poisson distributed with mean** R. Given that R is large, the $Poisson(R) \sim Normal(R, R)$. We look for the probability $P(N > R + \sqrt{R})$ with $N \sim Normal(R, R)$, or equivalently $P(Z = \frac{N-R}{\sqrt{R}} > 1) = 1 - \Phi(1) \approx 16\%$. Thus, for an M/M/$\infty$ the probability that we use more than $R + \sqrt{R}$ servers is only 16%.

Staffing: Square root staffing rule

Θεώρημα 2

Given an M/M/c with arrival rate λ and server speed μ and $R = \lambda/\mu$, where R is large, let c^* denote the ***least number of servers needed to ensure that*** *$C(c^*, R) < a$. Then,*

$$c^* = R + \beta\sqrt{R},$$

where β is the solution to the equation

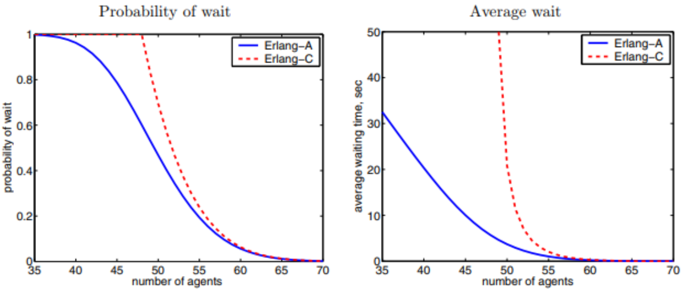
$$\frac{\beta\Phi(\beta)}{\varphi(\beta)} = \frac{1-a}{a},$$

where $\Phi(.)$ (resp $\varphi(.)$) denotes the c.d.f. (resp. p.d.f.) of the standard Normal.

- In practice β is quite small. When $a = 0.2$ then $\beta \approx 1$
- Thus, to ensure that only 20% of jobs queue up, it suffices to use just $R + \sqrt{R}$ servers.

M/M/c + M: modeling impatience (Palm/Erlang-A)

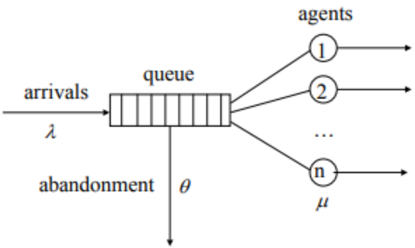
- Erlang-C (M/M/c) does not acknowledge abandonment
- More than 40% of the call centers set a target for the fraction of abandonment, but in most cases this target is not achieved[2].
- Moreover, the lack of understanding of the abandonment phenomenon and the scarcity of models that acknowledge it, has lead practitioners to ignore it altogether.

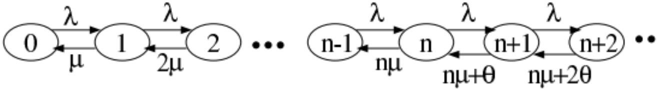


[2]Help Desk and Customer Support Practices Report, 1997

Birth-Death representation

- λ : (Poisson) ρυθμός αφίξεων (calls per unit of time).
- μ : (Exponential) ρυθμός εξυπηρέτησης ($1/\mu$ the average service duration).
- n ; αριθμός υπαλλήλων.
- θ : (Exponential) ρυθμός ανυπομονησίας.
- $d_j = min(j, n)\mu + max(0, j-n)\theta \equiv$ ρυθμός αναχώρησης.





Παρατηρήσεις

1. Παρατήρησε ότι

$$a_j := j * min(\mu, \theta) \leq d_j \leq b_j := j * max(\mu, \theta)$$

   - Η ποσότητα a_j (αντιστ. b_j) αποτελεί τον ρυθμό αναχωρήσεων σε ένα M/M/∞ σύστημα με ρυθμό εξυπηρέτησης a_j (αντιστ. b_j).
   - Τα δυο παραπάνω M/M/∞ συστήματα αποτελούν το κατώτερο και το *ανώτερο (στοχαστικό) όριο* του Erlang-A.
   - Με τον όρο *στοχαστικό όριο*, εννοούμε ότι οι οριακές πιθανότητες των καταστάσεων του Erlang-A, είναι μεταξύ των αντίστοιχων κατανομών των εν λόγω M/M/∞ (σε όρους στοχαστικής διάταξης).
   - Η ύπαρξη του κάτω ορίου εξασφαλίζει την ευστάθεια του Erlang-A.
2. Αν $\mu = \theta$, τότε το Erlang-A γίνεται ένα τροποποιημένο M/M/∞.

Stationary distribution

- Οι τοπικές εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda\pi_j &= (j+1)\mu\pi_{j+1},\ j = 0, 1, \ldots, n-1, \\ \lambda\pi_j &= (n\mu + (j+1-n))\pi_{j+1},\ j \geq n. \end{aligned}$$

- Η αναδρομική επίλυσή τους (ασκηση) είναι

$$\pi_j = \begin{cases} \pi_0 \frac{(\lambda/\mu)^j}{j!}, & j = 1, \ldots, n, \\ \pi_0 \frac{(\lambda/\mu)^n}{n!} \prod_{k=n+1}^{j} \left(\frac{\lambda}{n\mu + (k-n)\theta}\right), & j \geq n, \end{cases}$$

όπου

$$\pi_0 = \left[\sum_{j=0}^{n} \frac{(\lambda/\mu)^j}{j!} + \sum_{j=n+1}^{\infty} \prod_{k=n+1}^{j} \left(\frac{\lambda}{n\mu + (k-n)\theta}\right)\right]^{-1}.$$

- **Υπολογιστικό μειονέκτημα:** Εύρεση αθροίσματος απείρων όρων στον υπολογισμό του π_0.

Erlang-A is always stable!

- Όλα εξαρτώνται από την *σύγκλιση* της σειράς απείρων όρων στον υπολογισμό του π_0.
- Όμως

$$\begin{aligned} \pi_0^{-1} = & \sum_{j=0}^{n} \frac{(\lambda/\mu)^j}{j!} + \sum_{j=n+1}^{\infty} \prod_{k=n+1}^{j} \left(\frac{\lambda}{n\mu + (k-n)\theta} \right) \\ \leq & \sum_{j=0}^{\infty} \frac{(\lambda/min(\mu,\theta))^j}{j!} = e^{-(\lambda/min(\mu,\theta))}, \end{aligned}$$

αφού $k * min(\mu, \theta) \leq n\mu + (k-n)\theta$, για όλα τα $k \geq n$.

Erlang-A steady-state distribution & *special functions*

- Συνάρτηση Γάμμα

  $\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt,\ x > 0,\ \gamma(x,y) := \int_0^y t^{x-1} e^{-t} dt,\ x > 0, y \geq 0.$

- Από Palm (1957)[3]

$$A(x,y) := \frac{x e^y}{y} \gamma(x,y) = 1 + \sum_{j=1}^{\infty} \frac{y^j}{\prod_{k=1}^{j} (x+k)},\ x > 0, y \geq 0.$$

- Τότε, αν $B(n, \lambda/\mu)$ η Erlang B formula του M/M/n/n

$$\pi_j = \begin{cases} \pi_n \frac{n!}{j!(\lambda/\mu)^{n-j}}, & j = 1, \dots, n, \\ \pi_n \frac{(\lambda/\theta)^{j-n}}{\prod_{k=1}^{j-n} (\frac{n\mu}{\theta} + k)}, & j \geq n, \end{cases}$$

με

$$\pi_n = \frac{B(n, \lambda/\mu)}{1 + [A(\frac{n\mu}{\theta}, \frac{\lambda}{\theta}) - 1] B(n, \lambda/\mu)}^4$$

[3] Palm C. (1957) Research on telephone traffic carried by full availability groups. Tele, vol.1, 107

[4] Δες Appendix του "The Palm/Erlang-A Queue, with Applications to Call Centers". (Reading material, eclass)

Performance metrics

- **Delay probability:** Ποσοστό πελατών που πρέπει να περιμένουν. Από αρχή PASTA

$$\begin{aligned} P(W>0) &= \sum_{j=n}^{\infty} \pi_j = \pi_n \left[1 + \sum_{j=n+1}^{\infty} \frac{(\lambda/\theta)^{j-n}}{\prod_{k=1}^{j-n}(\frac{n\mu}{\theta}+k)} \right] \\ &= \frac{B(n,\lambda/\mu)}{1+[A(\frac{n\mu}{\theta},\frac{\lambda}{\theta})-1]B(n,\lambda/\mu)} A(\tfrac{n\mu}{\theta},\tfrac{\lambda}{\theta}). \end{aligned}$$

- **Fraction of abandonment:** Ποσοστό πελατών που θα αναχωρήσουν χωρίς εξυπηρέτηση.
  - Έστω P_j^{serv} η πιθανότητα να εξυπηρετηθεί ένας πελάτης που φθάνοντας θα βρει j πελάτες και όλους τους υπάλληλους απασχολημένους (δηλ. $j+n$ στο σύστημα).

$$P_0^{serv} = \frac{n\mu}{n\mu+\theta}.$$

  - ¨Δεσμεύοντας¨ στο 1ο γεγονός (είτε μια εξυπηρέτηση είτε μια αναχώρηση) μετά από μια άφιξη που θα βρει $n+1$ πελάτες στο σύστημα (first-step analysis)

$$P_1^{serv} = \frac{n\mu}{n\mu+2\theta}P_0^{serv} + \frac{\theta}{n\mu+2\theta}P_0^{serv} + \frac{\theta}{n\mu+2\theta}0 = \frac{n\mu}{n\mu+2\theta}.$$

(συνέχ)

- Επαγωγικά

$$P_j^{serv} = \frac{n\mu + j\theta}{n\mu + (j+1)\theta} P_{j-1}^{serv} = \frac{n\mu}{n\mu + (j+1)\theta}, j \geq 1.$$

- Τότε η πιθανότητα αναχώρησης για τον πελάτη που φθάνοντας βρει $n + j$ πελάτες στο σύστημα είναι

$$P_j^{Aband} = 1 - P_j^{serv} = \frac{(j+1)\theta}{n\mu + (j+1)\theta}.$$

- Τότε $P(Aband) = P(Aband|W > 0)P(W > 0)$ και

$$\begin{aligned}
P(Aband|W > 0) &= \sum_{j=0}^{\infty} P(Aband \ \& \ j \ in \ system|W > 0) \\
&= \sum_{j=n}^{\infty} P(Aband \ \& \ j \ in \ system|W > 0) \\
&= \frac{\sum_{j=n}^{\infty} P(Aband \ \& \ j \ in \ system)}{P(W > 0)} \\
&= \frac{\sum_{j=n}^{\infty} P(Aband| \ j \ in \ system)P(j \ in \ system)}{P(W > 0)} \\
&= \frac{\sum_{j=n}^{\infty} P(Aband|\text{βρίσκει } j-n \text{ στην ουρά})}{P(W > 0)} = \frac{\sum_{j=n}^{\infty} P_{j-n}^{Aband} \pi_j}{P(W > 0)}.
\end{aligned}$$

Reading material

1. W.J. Stewart (2009). PROBABILITY, MARKOV CHAINS, QUEUES, AND SIMULATION. Princeton University Press. **(Chapter 11)**
2. L. Lakatos, L. Szeidl, M. Telek. Introduction to Queueing Systems with Telecommunication Applications, Springer. **(Chapter 6,7)**
3. V. Kulkarni, (2010). Modeling and Analysis of Stochastic Systems, CRC Press. (**Chapter 7**, Sec. 7.3).
4. S. Ross (2010), Introduction to Probability Models, 10th ed., Academic Press. (**Chapter 8. Sec. 8.3, 8.9**).
5. M. Harchol-Balter (2013), Performance Modeling and Design of Computer Systems: Queueing Theory in Action. Cambridge University Press. **(Chapters 1,2,13,14,15)**

6. See also reading material in the corresponding **eclass folder**.