ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ (Queuing Systems)

Εισαγωγή

Καθηγητής Συμεών Παπαβασιλείου

Εθνικό Μετσόβιο Πολυτεχνείο
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Επικοινωνιών, Ηλεκτρονικής & Συστημάτων Πληροφορικής

(E-mail: papavass@mail.ntua.gr Τηλ: 210 772-2550
Γραφείο: Β.3.15 Νέο Κτίριο Ηλεκτρολόγων)

27 Φεβρουαρίου, 2020

Γενικά στοιχεία του μαθήματος

Προγραμματισμός Διαλέξεων & Εργαστηρίων

Διαλέξεις: κάθε Πέμπτη, 08:45-10:30, Αμφιθέατρο 5 (Νέο Κτ. Ηλεκτρολόγων)

Έναρξη: Πέμπτη, 27 Φεβρουαρίου 2020

Εργαστήριο: Μετά τις 2 πρώτες εβδομάδες μαθήματος, κάθε Δευτέρα, 10:45-12:30 (PC Lab Σχολής)

Μέθοδοι αξιολόγησης:

Ο βαθμός μαθήματος θα προκύψει από το βαθμό του εργαστηρίου (30%) και το βαθμό της εξέτασης στο «θεωρητικό» μέρος του μαθήματος (70%).

Βιβλιογραφία

[1] Α.-Γ. Σταφυλοπάτης και Γ. Σιόλας, "Ανάλυση Επίδοσης Υπολογιστικών Συστημάτων: Αναλυτικά Μοντέλα, Προσομοίωση, Μετρήσεις", Kallipos Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα & Βοηθήματα, 2015

https://repository.kallipos.gr/bitstream/11419/6055/4/master-%CE%9A%CE%9F%CE%A5.pdf

[2] Thomas G. Robertazzi, "Computer Networks and Systems: Queuing Theory and Performance Evaluation", Springer-Verlag, 2012.

http://mycourses.ntua.gr/course_description/index.php?cidReq=ECE1045

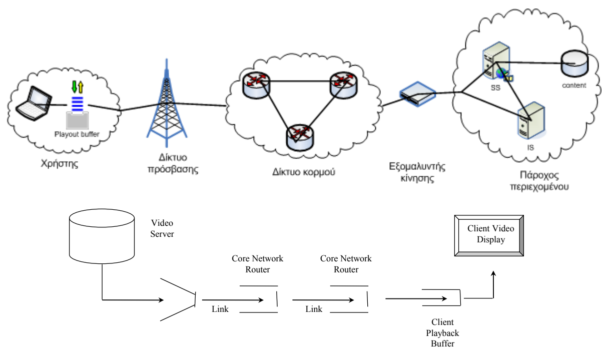
ΘΕΜΑΤΟΛΟΓΙΑ-ΠΕΡΙΕΧΟΜΕΝΑ (1/2)

1. **Εισαγωγή**
   - Περιεχόμενα
   - Γενική Περιγραφή Συστημάτων Αναμονής
   - Τεχνικές Μελέτης & Αξιολόγησης Επίδοσης Συστημάτων Αναμονής
   - Μοντέλα Τηλεπικοινωνιακών & Υπολογιστικών Συστημάτων
2. **Εισαγωγή στη Θεωρία Ουρών**.
   - Χαρακτηριστικά & Παράμετροι Συστημάτων Αναμονής
   - Μήκος Ουράς, Χρόνος Καθυστέρησης
   - Νόμος Little
3. **Γνώσεις από Θεωρία Πιθανοτήτων**
   - Εκθετική Κατανομή
   - Κατανομή Poisson
   - Ιδιότητα Απώλειας Μνήμης (Markov)
   - Στοχαστικές Ανελίξεις
4. **Μοντέλο Γεννήσεων - Θανάτων (Birth - Death Processes)**
5. **Συστήματα Markov και Εξισώσεις Ισορροπίας**
   - Ανάλυση απλών ουρών M/M/1

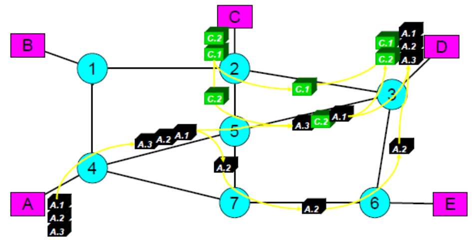
ΘΕΜΑΤΟΛΟΓΙΑ-ΠΕΡΙΕΧΟΜΕΝΑ (2/2)

6. **Άλλες ουρές Markov**
   - Μεταβάσεις Εξαρτώμενες από την Κατάσταση
   - Ουρές με Απώλειες (M/M/1/N)
   - Ουρές με Πολλαπλούς Εξυπηρετητές: M/M/m, M/M/m/K, M/M/m/m (Erlang – B)
7. **Προσομοίωση Απλών Συστημάτων Αναμονής**
8. **Ανοικτά και Κλειστά Δίκτυα Ουρών**
9. **Ουρές με μη Εκθετική Εξυπηρέτηση – M/G/1**
10. **Παραδείγματα & Εφαρμογές**
    - Ανάλυση Υπολογιστικών Συστημάτων
    - Ανάλυση & Σχεδίαση Τηλεφωνικών Κέντρων
    - Ανάλυση Δικτύων Internet & Ανάλυση Συστημάτων Πολυμέσων

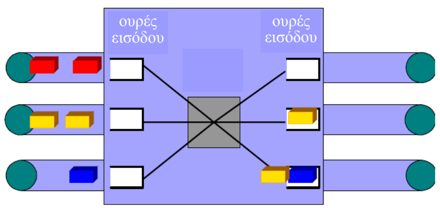
Δομή συστήματος παροχής υπηρεσιών πολυμέσων



Μεταγωγή αυτοδύναμων πακέτων



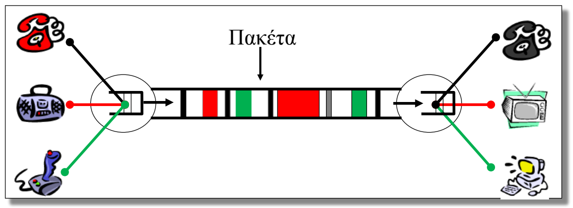
Μεταγωγέας πακέτου



Μεταγωγέας πακέτου

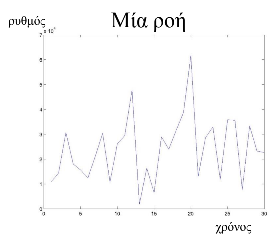
- Μεταφέρει πακέτα από τις ζεύξεις εισόδου στις ζεύξεις εξόδου.
- Υπάρχουν καθυστερήσεις αναμονής.

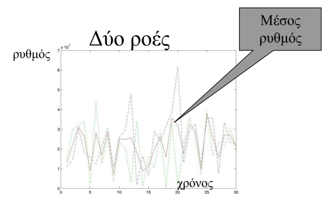
Μεταγωγή πακέτου – στατιστική πολυπλεξία

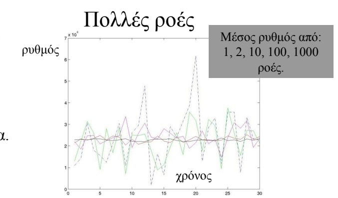


- Όταν εισέρχονται πολλές ροές στην ίδια ζεύξη, ο κόμβος μεταγωγής κρίνει ποια είσοδος θα εξυπηρετηθεί.
- Κάθε πακέτο οδεύει ανεξάρτητα στη ζεύξη.
- Δεν δεσμεύονται πόροι στη ζεύξη εκ των προτέρων. Η μεταγωγή πακέτου εκμεταλλεύεται τη στατιστική πολυπλεξία.

Στατιστική πολυπλεξία (1/3)

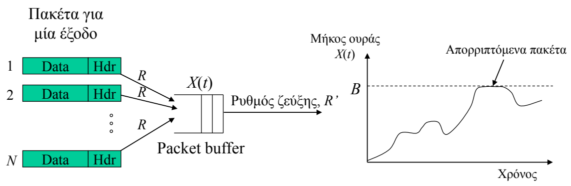






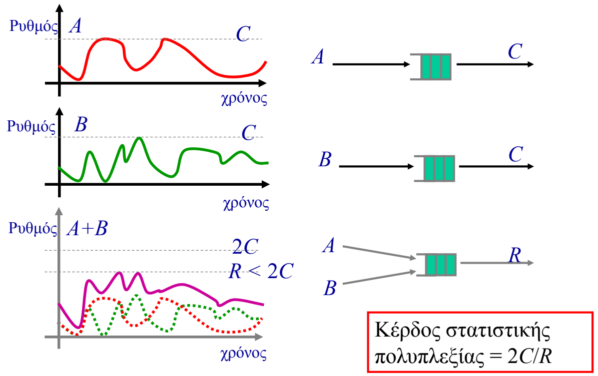
- Όταν η κίνηση είναι εκρηκτική, ο ρυθμός αλλάζει συχνά.
- Τα μέγιστα από διαφορετικές ροές εμφανίζονται γενικά ετεροχρονισμένα.
- Αποτέλεσμα: οι πολλές ροές έχουν ομαλότερη κίνηση.

Στατιστική πολυπλεξία (2/3)



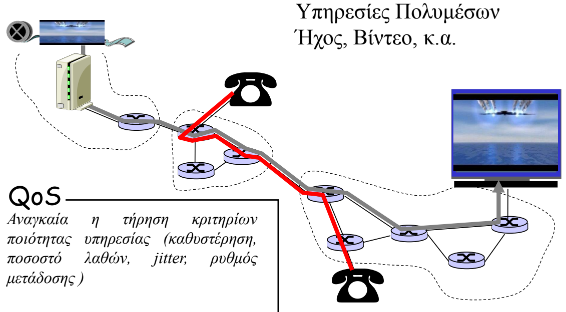
- Επειδή το buffer απορροφά τις εκρήξεις, η ζεύξη εξόδου δεν χρειάζεται να λειτουργεί με ρυθμό $N \times R$.
- Αλλά το buffer έχει πεπερασμένο μήκος B, οπότε θα υπάρχουν απώλειες.

Στατιστική πολυπλεξία (3/3)



Ποιότητα Υπηρεσίας
(Quality of Service - QoS)

Υπηρεσίες Πολυμέσων
Ήχος, Βίντεο, κ.α.



QoS

Αναγκαία η τήρηση κριτηρίων ποιότητας υπηρεσίας (καθυστέρηση, ποσοστό λαθών, jitter, ρυθμός μετάδοσης)

ΠΟΙΟΤΗΤΑ ΥΠΗΡΕΣΙΑΣ
(End-to-End User QoS Requirements for Real-time Services)

Medium	Application	Degree of symmetry	Key performance parameters and values		
			One-way Delay	**Delay variation**	**Information loss**
Audio	Conversational voice	Two-way	<150 msec preferred <400 msec limit	< 1 msec	< 3% PLR †
Audio	Voice messaging	Primarily one-way	< 1 sec for playback < 2 sec for record	< 1 msec‡	< 3% PLR
Audio	High quality streaming audio	Primarily one-way	< 10 sec	< 1 msec‡	< 1% PLR
Video	Videophone	Two-way	< 150 msec preferred <400 msec		< 1% PLR
Video	One-way	One-way	< 10 sec		< 1% PLR
Data	Telemetry - two-way control	Two-way	< 250 msec	N.A	Zero

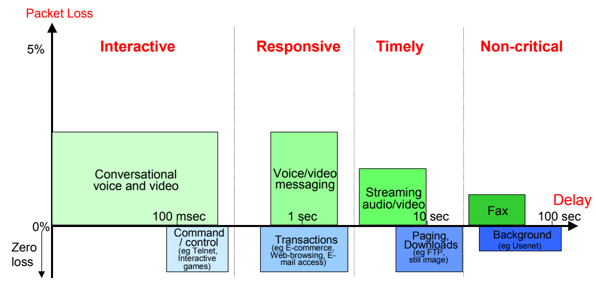
Typical QoS parameters
- ✓ Bandwidth: the data rate
- ✓ Delay: the latency of transmission
- ✓ Delay jitter: the variation in delay
- ✓ Loss ratio

ΠΟΙΟΤΗΤΑ ΥΠΗΡΕΣΙΑΣ
(End-to-End User QoS Requirements for Non Real-time Services)

Medium	Application	Degree of symmetry	Customer Demand *	Key Performance parameters		
				One-way delay	**Delay variation**	**Information loss**
Data	Fax (real-time)	Primarily one-way	Med	< 30 sec /page	N.A	<10-6 BER
Data	Fax (store & forward)	Primarily one-way	Med	Can be several minutes	N.A.	<10-6 BER
Data	Email (server to server transfer)	One-way	High	Can be several minutes	N.A	Zero
Data	Transaction services – lower priority	Primarily one-way? Two-way?	Med	< 30sec	N.A	Zero

Typical QoS parameters
- ✓ Bandwidth: the data rate
- ✓ Delay: the latency of transmission
- ✓ Delay jitter: the variation in delay
- ✓ Loss ratio

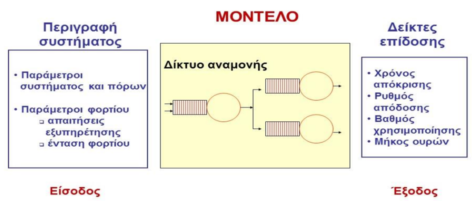
ΠΟΙΟΤΗΤΑ ΥΠΗΡΕΣΙΑΣ



ΣΧΕΔΙΑΣΗ & ΑΞΙΟΛΟΓΗΣΗ ΕΠΙΔΟΣΗΣ

Φάσεις:
- 1η: Φύση εφαρμογών που θα εξυπηρετηθούν από το σύστημα και φόρτος εργασίας (κυκλοφοριακή κίνηση)
- 2η: Αρχική αρχιτεκτονική του συστήματος (στοιχεία συστήματος – υλικό & λογισμικό)
- 3η: Ποσοτικός προσδιορισμός των τμημάτων/στοιχείων του συστήματος
- 4η: Μελέτη και μοντελοποίηση αλληλεπίδρασης τμημάτων του συστήματος
- Αξιολόγηση επίδοσης και επανεκτίμηση σχεδίασης – Ανάλυση ποιοτικών και ποσοτικών επιλογών

ΜΟΝΤΕΛΟ ΣΥΣΤΗΜΑΤΟΣ – ΔΙΚΤΥΟ ΑΝΑΜΟΝΗΣ



ΤΕΧΝΙΚΕΣ ΑΝΑΛΥΣΗΣ ΑΞΙΟΛΟΓΗΣΗΣ ΕΠΙΔΟΣΗΣ & ΜΟΝΤΕΛΟΠΟΙΗΣΗ

Τεχνικές:
- 1η: Μετρήσεις με πραγματικές τιμές και ανάλυση αποτελεσμάτων
  - Προέκταση σε μεγαλύτερης κλίμακας συστήματα: συνήθως η συμπεριφορά δεν είναι αναμενόμενη (π.χ. Γραμμική) σε αλλαγή φόρτου εργασίας
- 2η: Χρήση μοντέλων – μοντελοποίηση

  Γενικευμένη αναπαράσταση του συστήματος (αφαιρετική): περιλαμβάνει τα κύρια χαρακτηριστικά και αφαιρεί λεπτομέρειες που εκτιμάται ότι δεν επηρεάζουν σημαντικά την απόδοση του συστήματος (υποθέσεις). Κυκλοφορία (απαιτήσεις) χρηστών είναι στοχαστική (τυχαιότητα).
  - Αναλυτικά μοντέλα: χρήση μαθηματικής περιγραφής του συστήματος (βασισμένα κυρίως σε θεωρία αναμονής – queueing theory) και αλγορίθμων .
  - Προσομοίωση: Ανάπτυξη προγράμματος που ακολουθεί και αναπαριστά τη δυναμική εξέλιξη του συστήματος στο χρόνο.

ΚΡΙΤΗΡΙΑ ΕΠΙΛΟΓΗΣ ΤΕΧΝΙΚΗΣ ΑΝΑΛΥΣΗΣ

Κριτήριο	Αναλυτικό Μοντέλο	Προσομοίωση	Μετρήσεις
1. Στάδιο κύκλου ζωής	Οποιοδήποτε	Οποιοδήποτε	Υπάρχον σύστημα
2. Απαιτούμενος χρόνος	Μικρός	Μέτριος	Ποικίλλει
3. Απαιτούμενα εργαλεία	Θεωρία αναμονής	Γλώσσες προγραμματισμού	Οργανα μέτρησης
4. Ακρίβεια	Χαμηλή	Μέτρια	Ποικίλλει
5. Αποτίμηση εναλλακτικών λύσεων	Εύκολη	Μέτρια	Δύσκολη
6. Κόστος	Χαμηλό	Μέτριο	Υψηλό
7. Απήχηση	Χαμηλή	Μέτρια	Υψηλή

ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ (Queuing Systems)

Παράμετροι Ουρών και Συστημάτων Αναμονής

5 Μαρτίου, 2020

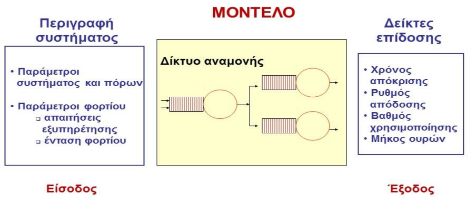
ΤΕΧΝΙΚΕΣ ΑΝΑΛΥΣΗΣ ΑΞΙΟΛΟΓΗΣΗΣ ΕΠΙΔΟΣΗΣ & ΜΟΝΤΕΛΟΠΟΙΗΣΗ (Επανάληψη)

Τεχνικές:

- 1η: Μετρήσεις με πραγματικές τιμές και ανάλυση αποτελεσμάτων
  - Προέκταση σε μεγαλύτερης κλίμακας συστήματα: συνήθως η συμπεριφορά δεν είναι αναμενόμενη (π.χ. Γραμμική) σε αλλαγή φόρτου εργασίας
- 2η: Χρήση μοντέλων – μοντελοποίηση

  Γενικευμένη αναπαράσταση του συστήματος (αφαιρετική): περιλαμβάνει τα κύρια χαρακτηριστικά και αφαιρεί λεπτομέρειες που εκτιμάται ότι δεν επηρεάζουν σημαντικά την απόδοση του συστήματος (υποθέσεις). Κυκλοφορία (απαιτήσεις) χρηστών είναι στοχαστική (τυχαιότητα).
  - Αναλυτικά μοντέλα: χρήση μαθηματικής περιγραφής του συστήματος (βασισμένα κυρίως σε θεωρία αναμονής – queueing theory) και αλγορίθμων .
  - Προσομοίωση: Ανάπτυξη προγράμματος που ακολουθεί και αναπαριστά τη δυναμική εξέλιξη του συστήματος στο χρόνο.

ΜΟΝΤΕΛΟ ΣΥΣΤΗΜΑΤΟΣ – ΔΙΚΤΥΟ ΑΝΑΜΟΝΗΣ - ΜΟΝΤΕΛΑ ΣΥΜΦΟΡΗΣΗΣ



- Κυκλοφοριακή κίνηση
- Ουρές σε δρομολογητές, καταστήματα, ταχυδρομεία, τράπεζες
  - Πολλαπλοί εξυπηρετητές (servers)
  - Κοινή ουρά ή παράλληλες ουρές, προτεραιότητες
- Τηλεφωνικά κέντρα (πολλαπλοί εξυπηρετητές)
- Κόμβοι δικτύων τύπου Internet
- Πόροι υπολογιστικών συστημάτων (CPU, Μνήμη, Δίσκοι)

ΚΟΙΝΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ (1/2)

- **Πελάτης:** Πελάτης τράπεζας, τηλεφωνική κλήση, πακέτο δεδομένων Internet…
- Εξυπηρετητής (**server**): Ταμίας, τηλεπικοινωνιακός πόρος (γραμμή) αφιερωμένος σε τηλεφωνική κλήση ή προώθηση πακέτου…
- Τυχαία είσοδος πελατών – «γεννήσεις», μέσος ρυθμός αφίξεων: λ πελάτες/sec
- Χρόνος μεταξύ δύο διαδοχικών αφίξεων - τυχαία μεταβλητή a, μέσος όρος: $\mathrm{E}(a) = 1/\lambda$ sec
- Μέσος ρυθμός εξυπηρέτησης πελατών: μ πελάτες/sec
- Χρόνος εξυπηρέτησης πελάτη – τυχαία μεταβλητή s, μέσος όρος: $\mathrm{E}(s) = 1/\mu$ sec/πελάτη

ΚΟΙΝΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ (2/2)

- Ουρά αναμονής (**queue**) για εξομάλυνση στατιστικών μεταβολών και απομόνωση (**buffering**) διακυμάνσεων εισόδου – εξυπηρέτησης
- Χωρητικότητα συστήματος αποθήκευσης (**queue size**) συμπεριλαμβανομένων των πελατών υπό εξυπηρέτηση
- Αριθμός εξυπηρετητών
- Πρωτόκολλο εξυπηρέτησης: First Come First Served - **FCFS** ή First In First Out - **FIFO**, Last In First Out - **LIFO**, Processor Sharing, προτεραιότητες
- **Κατάσταση συστήματος** $n(t)$: Αριθμός πελατών στο σύστημα αναμονής (ουρά + εξυπηρέτηση) σε μια χρονική στιγμή. Χρονοσειρά - time series - ή στοχαστική ανέλιξη - stochastic process - διακριτής κατάστασης & συνεχούς χρόνου
- Δρομολόγηση από ουρά σε ουρά σε περιπτώσεις δικτύων ουρών αναμονής

ΠΑΡΑΔΕΙΓΜΑΤΑ ΠΑΡΑΜΕΤΡΩΝ ΣΥΣΤΗΜΑΤΩΝ ΑΝΑΜΟΝΗΣ

- **Στοιχεία καθυστέρησης σε ένα σύστημα:** χρόνος επεξεργασίας, χρόνος αναμονής, χρόνος διάδοσης, χρόνος μετάδοσης
- **Δίκτυο μεταγωγής κυκλωμάτων (circuit switching):** ρυθμός αφίξεων κλήσεων, διάρκεια κλήσεων, ποσοστό απόρριψης κλήσεων
- **Δίκτυο μεταγωγής πακέτων (packet switching):** ρυθμός αφίξεων πακέτων, μέγεθος πακέτων, ποσοστό απόρριψης πακέτων, καθυστέρηση σε κόμβους του Internet
- **Υπολογιστικό σύστημα πολυεπεξεργασίας (windows):** αριθμός παράλληλων εντολών/προγραμμάτων υπό επεξεργασία, χρόνος ύπνωσης (sleeping time) ανά ενεργό παράθυρο, χρόνος αναζήτησης/ανταλλαγής δεδομένων στη μνήμη (I/O time), μέσος ρυθμός διεκπεραίωσης εντολών (ρυθμαπόδοση - throughput), χρόνος απόκρισης

Στοιχεία καθυστέρησης σε ένα σύστημα

- Processing Delay (χρόνος επεξεργασίας) is the time associated with the system analyzing a packet header and determining where the packet must be sent. This depends heavily on the entries in the routing table, the execution of data structures in the system, and the hardware implementation.
- Queueing Delay (χρόνος αναμονής) is the time between a packet being queued and it being sent. This varies depending on the amount of traffic, the type of traffic, and what router queue algorithms are implemented.
- Transmission Delay (χρόνος μετάδοσης) is the time needed to push a packet's data bits into the wire. This changes based on the size of the packet and the bandwidth. This does not depend on the distance of the wire, as it is solely the time to push a packet's bits into the wire, not to travel down the wire to the receiving endpoint.
- Propagation Delay (χρόνος διάδοσης) is the time associated with the first bit of the packet traveling from the sending endpoint to the receiving endpoint. This is often referred to as a delay by distance, and as such is influenced by the distance the bit must travel and the propagation speed.

ΠΑΡΑΜΕΤΡΟΙ (1/4)

- **Ένταση φορτίου (traffic intensity)**

  Σε περίπτωση 1 ουράς, 1 εξυπηρετητή:

  {Μέσος ***Χρόνος εξυπηρέτησης***}/ {Μέσος ***Χρόνος μεταξύ διαδοχικών αφίξεων***}

  $$\rho \triangleq \frac{\left(\frac{1}{\mu}\right)}{\left(\frac{1}{\lambda}\right)} = \lambda E(s) = \lambda/\mu \quad \text{(Erlangs)}$$

  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

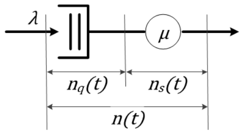
  Σε περίπτωση 1 ουράς, 1 εξυπηρετητή:

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

  όπου $\mathrm{P}\{\text{blocking}\}$ είναι η πιθανότητα να χαθεί ένας πελάτης επειδή βρήκε το σύστημα πλήρες
  - σε τηλεφωνικά δίκτυα: βαθμός ποιότητας, **Grade of Service** - **GoS**
  - σε δίκτυα δεδομένων: μία παράμετρος ποιότητας υπηρεσίας, **Quality of Service** - **QoS**

ΠΑΡΑΜΕΤΡΟΙ (2/4)

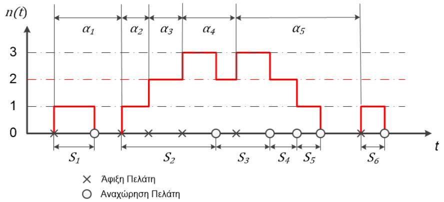
$\gamma = \lambda(1 - \mathrm{P}\{\text{blocking}\}) \leq \lambda,\ \gamma < \mu$



- **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
    $u \triangleq \gamma/\mu$

ΠΑΡΑΜΕΤΡΟΙ (3/4)

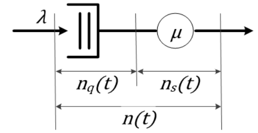
Εξέλιξη Αριθμού Πελατών στο Σύστημα



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

ΠΑΡΑΜΕΤΡΟΙ (4/4)

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



- *n(t)*: **Κατάσταση συστήματος αναμονής**
- *n_q(t)* : **Αριθμός πελατών στην αναμονή**
- *n_s(t)* : **Αριθμός πελατών στην εξυπηρέτηση**
- *n(t) = n_q(t) + n_s(t)*
- *E{n(t)} = E{n_q(t)} + E{n_s(t)}*
- **Χρόνος καθυστέρησης**: ***T = W + s***
- *E(T) = E(W) + E(s)*

ΚΑΤΑΣΤΑΣΗ ΣΥΣΤΗΜΑΤΟΣ

- $n(t) = 0,1,2,\ldots,K$: 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
  $n_q(t) = 0,1,2,\ldots,K-1$ ο αριθμός πελατών σε αναμονή
  $n_s(t) = 0,1$ ο αριθμός πελατών στην εξυπηρέτηση (αν έχω έναν εξυπηρετητή)

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

ΙΣΟΡΡΟΠΙΑ, ΕΡΓΟΔΙΚΕΣ ΚΑΤΑΣΤΑΣΕΙΣ ΣΥΣΤΗΜΑΤΟΣ

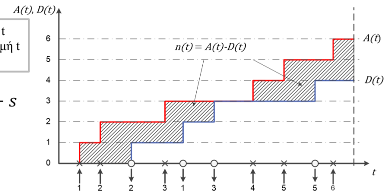
- **ΟΡΙΣΜΟΣ ΙΣΟΡΡΟΠΙΑΣ**: Αν μια στοχαστική ανέλιξη $n(t)$ **ισορροπήσει** μετά από παρέλευση μεγάλου χρονικού διαστήματος t, το μεταβατικό φαινόμενο παρέρχεται και το σύστημα παλινδρομεί τυχαία ανάμεσα σε απείρως επισκέψιμες (**γνησίως επαναληπτικές, positive recurrent**) καταστάσεις $n(t) = k$. Οι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

A(t): συνολικός # αφίξεων μέχρι τη στιγμή t
D(t): συνολικός # αναχωρήσεων μέχρι τη στιγμή t
n(t): συνολικός # πελατών τη στιγμή t



Χρόνος καθυστέρησης: $T = W + s$

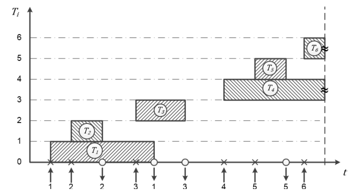
Τύπος Little:

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

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 $u = \frac{\gamma}{\mu} = \mathrm{P}\{n(t) > 0\}$)



ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ (Queuing Systems)

Εκθετική Κατανομή &Κατανομή Poisson
Διαδικασία Markov Γεννήσεων – Θανάτων
(Birth – Death Markov Processes)

Εβδομάδα 16 Μαρτίου, 2020

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*

Η εκθετική κατανομή (exponential distribution)

- Μια τυχαία μεταβλητή (τ.μ.) - random variable - ***X*** ακολουθεί ***Εκθετική Κατανομή*** *(Exponential Distribution)* με παράμετρο **λ** όταν:
- **CDF:** $F_X(t) = \mathrm{P}[X \leq t] = \begin{cases} 1 - e^{-\lambda t}, t \geq 0 \\ 0, \quad t < 0 \end{cases}$ και **PDF:** $f_X(t) = \frac{dF_X(t)}{dt} = \begin{cases} \lambda e^{-\lambda t}, t \geq 0 \\ 0, \quad t < 0 \end{cases}$
- $\mathrm{E}[X] = \int_{t=0}^{\infty} \lambda t e^{-\lambda t} dt = 1/\lambda$
- $\mathrm{E}[X^2] = \int_{t=0}^{\infty} \lambda t^2 e^{-\lambda t} dt = 2/\lambda^2,\ \sigma_X^2 = \mathrm{E}[X] - (\mathrm{E}[X])^2 = 1/\lambda^2$
- Ιδιότητα έλλειψης μνήμης:
  - $\mathrm{P}[X > t + s\,|X > s] = \frac{\mathrm{P}[X>t+s, X>s]}{\mathrm{P}[X>s]} = \frac{\mathrm{P}[X>t+s]}{\mathrm{P}[X>s]} = e^{-\lambda t} = \mathrm{P}[X > t] = 1 - F_X(t)$

  Η εκθετική κατανομή είναι η **μόνη κατανομή συνεχούς μεταβλητής** με την ιδιότητα αυτή *(Memoryless, Markov Property)*.
- Κατανομή ελαχίστου μεταξύ ανεξάρτητων τ.μ. εκθετικά κατανεμημένων
  X1: με παράμετρο λ1
  X2: με παράμετρο λ2
  X = min{X1,X2} είναι εκθετικά κατανεμημένη με παράμετρο: λ = λ1+λ2

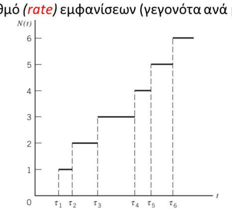
Στοχαστικές διαδικασίες
(Stochastic Processes – Time Series)

- Στάσιμες διαδικασίες (stationary stochastic processes) - οι από κοινού συναρτήσεις κατανομής πιθανότητας είναι αμετάβλητες σε μετατοπίσεις στο χρόνο
- Διαδικασίες Markov, ιδιότητα έλλειψης μνήμης
  $P[\boldsymbol{X}(t_{n+1})=x_{n+1}/\boldsymbol{X}(t_n)=x_n,\boldsymbol{X}(t_{n-1})=x_{n-1},\ldots,\boldsymbol{X}(t_1)=x_1] = P[\boldsymbol{X}(t_{n+1})=x_{n+1}/\boldsymbol{X}(t_n)=x_n]$
- Εργοδικότητα (ergodicity) ως προς τον μέσο όρο – μέση τιμή στο χρόνο συνάρτησης δείγματος είναι ίση με στατιστική μέση τιμή
- Διαδικασίες Γεννήσεων-Θανάτων (birth – death processes): αποτελούν μια κλάση των διαδικασιών Markov, με την επιπλέον συνθήκη ότι μεταβάσεις επιτρέπονται μόνο ανάμεσα σε γειτονικές καταστάσεις
- Διαδικασία απαρίθμησης γεγονότων (counter processes)
  $P[\boldsymbol{N}(t) = k]$: Πιθανότητα k γεγονότων στο διάστημα $(0, t)$
- Ανεξάρτητες αυξήσεις: αν οι αριθμοί των γεγονότων που λαμβάνουν χώρα σε μη επικαλυπτόμενα διαστήματα είναι ανεξάρτητοι μεταξύ τους
- Στάσιμες αυξήσεις (stationary increments): Ανεξάρτητα του χρόνου αναφοράς t *(εξάρτηση μόνο από το μήκος του διαστήματος)*
  $P[\boldsymbol{N}(t + \Delta t) - \boldsymbol{N}(t) = k] = P[\boldsymbol{N}(\tau + \Delta t) - \boldsymbol{N}(\tau) = k] = P[\boldsymbol{N}(\Delta t) = k]$

Η ΔΙΑΔΙΚΑΣΙΑ ΚΑΤΑΜΕΤΡΗΣΗΣ ΓΕΓΟΝΟΤΩΝ POISSON

Η τυχαία εμφάνιση παλμών περιγράφεται σαν μια Στοχαστική Ανέλιξη Καταμέτρησης *(Counting Process)* $N(t)$ που καταμετρά τυχαία γεγονότα (αφίξεις πελατών) στο διάστημα $(0, t)$.

Ο αριθμός εμφανίσεων στο διάστημα $(t, t+T)$ είναι διακριτή τυχαία μεταβλητή $\nu = N(t+T) - N(t)$. Κάτω από συνθήκες απρόβλεπτης εξέλιξης της ανέλιξης *(τα γεγονότα εμφανίζονται ανεξάρτητα από το παρελθόν και χωρίς να επηρεάζουν το μέλλον)*, η ν ακολουθεί την **κατανομή *Poisson* με μέσο αριθμό εμφανίσεων ανάλογο του διαστήματος** T: $\mathrm{E}_T[\nu] = \lambda T$. Η σταθερά λ ορίζει τον μέσο ρυθμό *(rate)* εμφανίσεων (γεγονότα ανά μονάδα χρόνου)



Η ΚΑΤΑΝΟΜΗ POISSON (1/3)

Διακριτή Τυχαία Μεταβλητή $\nu = N(t+T) - N(t)$ απαρίθμησης γεγονότων σε χρονικό διάστημα παρατήρησης T που εμφανίζονται **τυχαία** και **ανεξάρτητα** από παρελθούσες ή μελλοντικές εμφανίσεις γεγονότων στο δείγμα (υλοποίηση) της Στοχαστικής Ανέλιξης μετρητή $N(t)$ στο οποίο συνεισφέρουν (**ιδιότητα έλλειψης μνήμης *Markov***)

Ο μέσος όρος εμφανίσεων γεγονότων στο διάστημα T είναι $\mathrm{E}_T[\nu] = \lambda T$

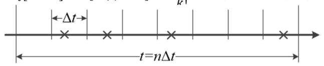
Εφαρμογές σε ανεξάρτητες εμφανίσεις τυχαίων γεγονότων:
- Τυχαίες εκρήξεις που προκαλούν τον **ΘΟΡΥΒΟ ΒΟΛΗΣ** σε ηλεκτρονικές συσκευές επικοινωνών
- Ανεξάρτητες τυχαίες **αφίξεις πελατών** σε **ΣΥΣΤΗΜΑΤΑ ΟΥΡΩΝ ΑΝΑΜΟΝΗΣ** με απαιτήσεις εξυπηρέτησης όπως:
  - Διεκπεραίωση Τηλεφωνικών Κλήσεων
  - Διακίνηση Πακέτων Δεδομένων στο Internet
  - Κυκλοφορία Αυτοκίνητων σε Οδικά Συστήματα
  - Αγορές και Πληρωμές σε Καταστήματα
  - Επεξεργασία Δεδομένων σε Κοινές Υπολογιστικές Υποδομές

Η ΚΑΤΑΝΟΜΗ POISSON (2/3)
Η Κατανομή Poisson σαν Όριο Διωνυμικής Κατανομής

Ανεξάρτητες εμφανίσεις $\{N(t) = k\}$ γεγονότων (*σημείων*) ***Poisson*** στο διάστημα $(0, t)$ με ρυθμό λ σημεία/sec ορίζουν Διακριτή Τυχαία Μεταβλητή *(Discrete Random Variable)* $\{\nu = k\}$ με Κατανομή Μάζας Πιθανότητας

$$\mathrm{P}_t[\nu = k] \triangleq \mathrm{P}[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \qquad k = 0,1,2,\cdots$$

Απόδειξη



- Διαιρώ το διάστημα t σε n υποδιαστήματα, $t = n\Delta t$
- Πραγματοποιώ n ανεξάρτητες δοκιμές Bernouilli, μια σε κάθε υποδιάστημα, με δύο εναλλακτικές: Εμφάνιση (***επιτυχία***) με πιθανότητα $p = \lambda\Delta t$, μη εμφάνιση (***αποτυχία***) με $1 - p$
- Η πιθανότητα k επιτυχιών σε n ανεξάρτητες δοκιμές δίνεται από την Διωνυμική Κατανομή:

$$\mathrm{P}[N(t) = k] = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0,1,\cdots,n$$

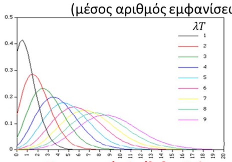
$$\mathrm{P}[N(t) = k] = \binom{n}{k} (\lambda\Delta t)^k (1-\lambda\Delta t)^{n-k} = \binom{n}{k}\left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}$$

- Στο όριο $\Delta t \to 0,\ n \to \infty, t = n\Delta t$ έχουμε $\frac{n!}{(n-k)!} \to n^k$, $\left(1 - \frac{\lambda t}{n}\right)^{n-k} \to e^{-\lambda t}$ και

$$\mathrm{P}[N(t) = k] = \frac{n!}{k!\,(n-k)!}\left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \to \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Η ΚΑΤΑΝΟΜΗ POISSON (3/3)
Κατανομή Poisson για Διαφορετικές Τιμές του $\lambda T = \mathrm{E}[N(T)]$

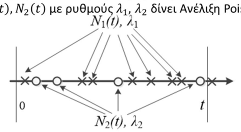
(μέσος αριθμός εμφανίσεων γεγονότων σε διάστημα T)



Οι συνεχείς καμπύλες στο σχήμα είναι οι περιβάλλουσες των Συναρτήσεων Μάζας Πιθανότητας (Ιστογράμματος) της Διακριτής Τυχαίας Μεταβλητής Poisson $\mathrm{P}_T[\nu = k] \triangleq \mathrm{P}[N(T) = k] = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$

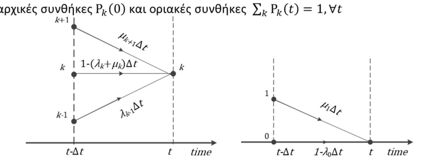
Ιδιότητες της Στοχαστικής Ανέλιξης Poisson

- **Μέση Τιμή & Διασπορά**: $\mathrm{E}[N(t)] = \sigma^2_{N(t)} = \lambda t$
  Απόδειξη: $\mathrm{E}[N(t)] = \lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}[N_i(t)] = \lim_{n\to\infty} \sum_{i=1}^n \lambda\Delta t = \lambda t$, $\sigma^2_{N(t)} = \lim_{n\to\infty} \sum_{i=1}^n \sigma^2_{N_i(t)} = \lambda t$
- Ο συνολικός αριθμός σημείων Στοχαστική Ανέλιξης Poisson ρυθμού λ σε **μη υπερ-καλυπτόμενα** χρονικά διαστήματα T_1, T_2 είναι διακριτή τυχαία μεταβλητή Poisson με μέση τιμή $\lambda(T_1 + T_2)$
- **Υπέρθεση** δυο **ανεξαρτήτων** Ανελίξεων Poisson $N_1(t), N_2(t)$ με ρυθμούς λ_1, λ_2 δίνει Ανέλιξη Poisson $N(t)$ με ρυθμό $\lambda = \lambda_1 + \lambda_2$
- **Διάσπαση** Ανέλιξης Poisson ρυθμού λ μέσω ανεξαρτήτων τυχαίων επαναλήψεων **Bernoulli** με πιθανότητες $p, q = 1 - p$
  Παράδειγμα: Τυχαία δρομολόγηση χωρίς μνήμη δημιουργεί ανεξάρτητες ανελίξεις (διαδικασίες) Poisson με μέσους ρυθμούς $\lambda_1 = p\lambda, \lambda_2 = q$



ΔΙΑΔΙΚΑΣΙΑ ΓΕΝΝΗΣΕΩΝ – ΘΑΝΑΤΩΝ (1/3)
Birth – Death Processes

- Παραδοχές:
  - Ανεξαρτησία γεννήσεων-θανάτων
  - Εξέλιξη της κατάστασης - πληθυσμού $n(t)$ βασισμένη μόνο στο παρόν (Ιδιότητα Markov)
- Σύστημα Διαφορικών εξισώσεων Διαφορών
  - Κατάσταση ισορροπίας (**steady state**)
  - Την χρονική στιγμή t το σύστημα καταλήγει σε πληθυσμό $n(t) = k$
  - Μπορεί να έχουν προηγηθεί οι ακόλουθες μεταβάσεις από την χρονική στιγμή $t - \Delta t, \Delta t \to 0$:
    - Μία άφιξη στο διάστημα Δt, με πιθανότητα $\lambda_{k-1}\Delta t$ αν $k > 0$
    - Μια αναχώρηση, με πιθανότητα $\mu_{k+1}\Delta t$ αν υπάρχει η κατάσταση $k+1$ (σε περίπτωση περιορισμού μέγιστου πληθυσμού K μπορούμε να θεωρήσουμε $\mu_{k+1} = 0$)
    - Τίποτα από τα δύο, με πιθανότητα $1 - (\lambda_k + \mu_k)\Delta t$ αν $k > 0$ ή $1 - \lambda_0\Delta t$ αν $k = 0$
- Οι εξισώσεις μετάβασης (***Chapman - Kolmogorov***) προκύπτουν από τον τύπο συνολικής πιθανότητας:
  - $\mathrm{P}_k(t) = \lambda_{k-1}\Delta t \mathrm{P}_{k-1}(t-\Delta t) + \mu_{k+1}\Delta t \mathrm{P}_{k+1}(t-\Delta t) + [1 - (\lambda_k + \mu_k)\Delta t]\mathrm{P}_k(t-\Delta t)$
  - $\mathrm{P}_{\mathbf{0}}(t) = \mu_1 \Delta t\, \mathrm{P}_1(t-\Delta t) + (1 - \lambda_0\Delta t)\, \mathrm{P}_{\mathbf{0}}(t-\Delta t)$
  - με αρχικές συνθήκες $\mathrm{P}_k(0)$ και οριακές συνθήκες $\sum_k \mathrm{P}_k(t) = 1, \forall t$



ΔΙΑΔΙΚΑΣΙΑ ΓΕΝΝΗΣΕΩΝ – ΘΑΝΑΤΩΝ (2/3)
Birth – Death Processes

Στο όριο, $\Delta t \approx dt \to 0$, $\frac{\mathrm{P}_k(t) - P_k(t-\Delta t)}{\Delta t} \to \frac{d\mathrm{P}_k(t)}{dt}$ και προκύπτει το γραμμικό σύστημα διαφορικών εξισώσεων διαφορών:

- $\frac{d\mathrm{P}_k(t)}{dt} = \lambda_{k-1}\mathrm{P}_{k-1}(t) + \mu_{k+1}\mathrm{P}_{k+1}(t) - (\lambda_k + \mu_k)\mathrm{P}_k(t),\ k > 0$
- $\frac{d\mathrm{P}_0(t)}{dt} = \mu_1\mathrm{P}_1(t) - \lambda_0\mathrm{P}_0(t)$
- με αρχικές συνθήκες $\mathrm{P}_k(0)$ και οριακές συνθήκες $\sum_k \mathrm{P}_k(t) = 1, \forall t$

Όταν $t \to \infty$ και υπό ορισμένες συνθήκες το σύστημα συγκλίνει σε σταθερή κατάσταση. Το μεταβατικό φαινόμενο παρέρχεται για ***επαναληπτικές*** καταστάσεις $n(t) = k$ (απείρως επισκέψιμες - ***positive recurrent***) ξεχνιέται η αρχική συνθήκη $\mathrm{P}_k(0)$ και οι $\mathrm{P}_k(t)$ συγκλίνουν στις οριακές πιθανότητες $P_k > 0$:

Για $t \to \infty$, $\frac{d\mathrm{P}_k(t)}{dt} = 0$, $\mathrm{P}_k(t) \to P_k > 0$: ***Εργοδικές Οριακές Πιθανότητες***

Σημείωση: Ισχύει η ***εργοδική*** ιδιότητα και οι οριακές πιθανότητες μπορούν να προσεγγισθούν σαν $P_k = \lim_{\mathrm{T}\to\infty}\{\frac{\mathrm{T}_k}{\mathrm{T}}\}$ όπου T_k είναι το σχετικό συνολικό χρονικό διάστημα όταν $n(t) = k$ σε μεγάλο χρονικό ορίζοντα T ***μιας*** καταγραφής της ανέλιξης $n(t)$ σε ***ισορροπία***.

Οι εργοδικές οριακές πιθανότητες προκύπτουν από τις γραμμικά ανεξάρτητες ***Εξισώσεις Ισορροπίας***:

- $(\lambda_k + \mu_k)P_k = \lambda_{k-1}P_{k-1} + \mu_{k+1}P_{k+1},\ k > 1$
- $\lambda_0 P_0 = \mu_1 P_1$
- $P_0 + P_1 + \cdots + P_k + \cdots = 1$

ΔΙΑΔΙΚΑΣΙΑ ΓΕΝΝΗΣΕΩΝ – ΘΑΝΑΤΩΝ (3/3)
Birth – Death Processes
Εφαρμογή σε Απλή Ουρά M/M/1

- Αφίξεις Poisson με μέσο ρυθμό λ αφίξεις/sec: $\lambda_k = \lambda,\ \ k = 0,1,2,3,\cdots$
- Χρόνοι εξυπηρέτησης εκθετικοί με μέση τιμή $\mathrm{E}(s) = \frac{1}{\mu}$sec: $\mu_k = \mu,\ k = 1,2,3,\cdots$
- $\rho = \frac{\lambda}{\mu} < 1$ Erlang (συνθήκη για οριακή ισορροπία – εργοδικότητα)
- Η εξέλιξη των πιθανοτήτων $\mathrm{P}[n(t) = k] = \mathrm{P}_k(t)$ προκύπτει από το σύστημα διαφορικών εξισώσεων:
  - $\frac{d\mathrm{P}_k(t)}{dt} = \lambda\mathrm{P}_{k-1}(t) + \mu\mathrm{P}_{k+1}(t) - (\lambda + \mu)\mathrm{P}_k(t),\ k > 0$
  - $\frac{d\mathrm{P}_0(t)}{dt} = \mu\mathrm{P}_1(t) - \lambda\mathrm{P}_0(t)$
  - με αρχικές συνθήκες $\mathrm{P}_k(0)$ και οριακές συνθήκες $\sum_{k=0}^{\infty} \mathrm{P}_k(t) = 1\ \ \forall t \geq 0$
- Στο όριο $t \to \infty$, $\frac{d\mathrm{P}_k(t)}{dt} = 0$, $\mathrm{P}_k(t) \to P_k > 0$, τις **εργοδικές πιθανότητες** που προκύπτουν από τις εξισώσεις ισορροπίας:
  - $\lambda P_0 = \mu P_1$ ή $P_1 = \left(\frac{\lambda}{\mu}\right)P_0 = \rho P_0$
  - $(\lambda + \mu)P_1 = \lambda P_0 + \mu P_2$ ή $P_2 = \rho^2 P_0$ και γενικά $P_k = \rho^k P_0,\ \ k > 0$
  - $P_0 + P_1 + \cdots + P_k + \cdots = 1 = P_0(1 + \rho + \rho^2 + \rho^3 + \cdots)$

  Εφόσον $0 < \rho < 1$ η άπειρη δυναμοσειρά$(1 + \rho + \rho^2 + \rho^3 + \cdots) \to \frac{1}{1-\rho}$ $\Rightarrow$ $P_0(\frac{1}{1-\rho}) = 1$ και

$$P_0 = (1-\rho),\ \ P_k = (1-\rho)\rho^k, \qquad k > 0$$

  Μέσο μήκος ουράς M/M/1 σε ισορροπία: $E[n(t)] \triangleq \sum_{k=1}^{\infty} kP_k = \frac{\rho}{1-\rho}$

ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ (Queuing Systems)

Συστήματα Γεννήσεων – Θανάτων:

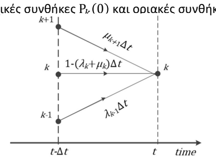
- Σφαιρικές & Λεπτομερείς Εξισώσεις Ισορροπίας
- Ουρές Markov M/M/1, M/M/1/N

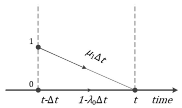
Εβδομάδα 23 Μαρτίου, 2020

ΔΙΑΔΙΚΑΣΙΑ ΓΕΝΝΗΣΕΩΝ – ΘΑΝΑΤΩΝ (1/3)

Birth – Death Processes *(επανάληψη)*

- Παραδοχές:
  - Ανεξαρτησία γεννήσεων-θανάτων
  - Εξέλιξη της κατάστασης - πληθυσμού $n(t)$ βασισμένη μόνο στο παρόν (Ιδιότητα Markov)
- Σύστημα Διαφορικών εξισώσεων Διαφορών
  - Κατάσταση ισορροπίας (**steady state**)
  - Την χρονική στιγμή t το σύστημα καταλήγει σε πληθυσμό $n(t) = k$
  - Μπορεί να έχουν προηγηθεί οι ακόλουθες μεταβάσεις από την χρονική στιγμή $t - \Delta t, \Delta t \to 0$:
    - Μία άφιξη στο διάστημα Δt, με πιθανότητα $\lambda_{k-1}\Delta t$ αν $k > 0$
    - Μια αναχώρηση, με πιθανότητα $\mu_{k+1}\Delta t$ αν υπάρχει η $k+1$ (σε περίπτωση περιορισμού μέγιστου πληθυσμού K μπορούμε να θεωρήσουμε $\mu_{k+1} = 0$)
    - Τίποτα από τα δύο, με πιθανότητα $1 - (\lambda_k + \mu_k)\Delta t$ αν $k > 0$ ή $1 - \lambda_0 \Delta t$ αν $k = 0$
- Οι εξισώσεις μετάβασης (***Chapman - Kolmogorov***) προκύπτουν από τον τύπο συνολικής πιθανότητας:
  - $\mathrm{P}_k(t) = \lambda_{k-1}\Delta t \mathrm{P}_{k-1}(t - \Delta t) + \mu_{k+1}\Delta t \mathrm{P}_{k+1}(t - \Delta t) + [1 - (\lambda_k + \mu_k)\Delta t]\mathrm{P}_k(t - \Delta t)$
  - $\mathrm{P}_{\mathbf{0}}(t) = \mu_1 \Delta t\, \mathrm{P}_1(t - \Delta t) + (1 - \lambda_0 \Delta t)\, \mathrm{P}_{\mathbf{0}}(t - \Delta t)$
  - με αρχικές συνθήκες $\mathrm{P}_k(0)$ και οριακές συνθήκες $\sum_k \mathrm{P}_k(t) = 1, \forall t$





ΔΙΑΔΙΚΑΣΙΑ ΓΕΝΝΗΣΕΩΝ – ΘΑΝΑΤΩΝ (2/3)

Birth – Death Processes *(επανάληψη)*

Στο όριο, $\Delta t \approx dt \to 0$, $\frac{\mathrm{P}_k(t) - P_k(t-\Delta t)}{\Delta t} \to \frac{d\mathrm{P}_k(t)}{dt}$ και προκύπτει το γραμμικό σύστημα διαφορικών εξισώσεων διαφορών:

- $\frac{d\mathrm{P}_k(t)}{dt} = \lambda_{k-1}\mathrm{P}_{k-1}(t) + \mu_{k+1}\mathrm{P}_{k+1}(t) - (\lambda_k + \mu_k)\mathrm{P}_k(t), \ k > 0$
- $\frac{d\mathrm{P}_0(t)}{dt} = \mu_1 \mathrm{P}_1(t) - \lambda_0 \mathrm{P}_0(t)$
- με αρχικές συνθήκες $\mathrm{P}_k(0)$ και οριακές συνθήκες $\sum_k \mathrm{P}_k(t) = 1, \forall t$

Όταν $t \to \infty$ και υπό ορισμένες συνθήκες το σύστημα συγκλίνει σε σταθερή κατάσταση. Το μεταβατικό φαινόμενο παρέρχεται για απείρως επισκέψιμες καταστάσεις $n(t) = k$ (***επαναληπτικές, positive recurrent***), ξεχνιέται η αρχική συνθήκη $\mathrm{P}_k(0)$ και οι $\mathrm{P}_k(t)$ συγκλίνουν στις οριακές πιθανότητες $P_k > 0$:

Για $t \to \infty$, $\frac{d\mathrm{P}_k(t)}{dt} = 0$, $\mathrm{P}_k(t) \to P_k > 0$: ***Εργοδικές Οριακές Πιθανότητες***

Σημείωση: Ισχύει η ***εργοδική*** ιδιότητα και οι οριακές πιθανότητες μπορούν να προσεγγισθούν σαν

$P_k = \lim_{T\to\infty}\{\frac{T_k}{T}\}$ όπου T_k είναι το σχετικό συνολικό χρονικό διάστημα T_k όταν $n(t) = k$ σε μεγάλο χρονικό

ορίζοντα T ***μιας*** καταγραφής της ανέλιξης $n(t)$ σε ***ισορροπία***.

Οι εργοδικές οριακές πιθανότητες προκύπτουν από τις γραμμικά ανεξάρτητες ***Εξισώσεις Ισορροπίας***:

- $(\lambda_k + \mu_k)P_k = \lambda_{k-1}P_{k-1} + \mu_{k+1}P_{k+1}, \ k > 1$
- $\lambda_0 P_0 = \mu_1 P_1$
- $P_0 + P_1 + \cdots + P_k + \cdots = 1$

ΔΙΑΔΙΚΑΣΙΑ ΓΕΝΝΗΣΕΩΝ – ΘΑΝΑΤΩΝ (3/3)

Birth – Death Processes *(επανάληψη)*

Εφαρμογή σε Απλή Ουρά M/M/1

- Αφίξεις Poisson με μέσο ρυθμό λ αφίξεις/sec: $\lambda_k = \lambda, \ k = 0,1,2,3,\cdots$
- Χρόνοι εξυπηρέτησης εκθετικοί με μέση τιμή $\mathrm{E}(s) = \frac{1}{\mu}$sec: $\mu_k = \mu, \ k = 1,2,3,\cdots$
- $\rho = \frac{\lambda}{\mu} < 1$ Erlang (συνθήκη για οριακή ισορροπία – εργοδικότητα)
- Η εξέλιξη των πιθανοτήτων $\mathrm{P}[n(t) = k] = \mathrm{P}_k(t)$ προκύπτει από το σύστημα διαφορικών εξισώσεων:
  - $\frac{d\mathrm{P}_k(t)}{dt} = \lambda \mathrm{P}_{k-1}(t) + \mu \mathrm{P}_{k+1}(t) - (\lambda + \mu)\mathrm{P}_k(t), \ k > 0$
  - $\frac{d\mathrm{P}_0(t)}{dt} = \mu \mathrm{P}_1(t) - \lambda \mathrm{P}_0(t)$
  - με αρχικές συνθήκες $\mathrm{P}_k(0)$ και οριακές συνθήκες $\sum_{k=0}^{\infty} \mathrm{P}_k(t) = 1 \ \forall t \geq 0$
- Στο όριο $t \to \infty$, $\frac{d\mathrm{P}_k(t)}{dt} = 0$, $\mathrm{P}_k(t) \to P_k > 0$, τις **εργοδικές πιθανότητες** που προκύπτουν από τις εξισώσεις ισορροπίας:
  - $\lambda P_0 = \mu P_1$ ή $P_1 = \left(\frac{\lambda}{\mu}\right) P_0 = \rho P_0$
  - $(\lambda + \mu)P_1 = \lambda P_0 + \mu P_2$ ή $P_2 = \rho^2 P_0$ και γενικά $P_k = \rho^k P_0, \ k > 0$
  - $P_0 + P_1 + \cdots + P_k + \cdots = 1 = P_0(1 + \rho + \rho^2 + \rho^3 + \cdots)$

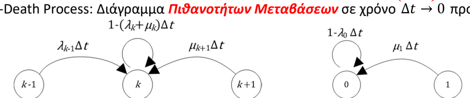
  Εφόσον $0 < \rho < 1$ η άπειρη δυναμοσειρά$(1 + \rho + \rho^2 + \rho^3 + \cdots) \to \frac{1}{1-\rho} \ \Rightarrow \ P_0(\frac{1}{1-\rho}) = 1$ και

$$P_0 = (1 - \rho), \quad P_k = (1 - \rho)\rho^k, \qquad k > 0$$

Μέσο μήκος ουράς M/M/1 σε ισορροπία: $E[n(t)] \triangleq \sum_{k=1}^{\infty} kP_k = \frac{\rho}{1-\rho}$

ΕΞΙΣΩΣΕΙΣ ΙΣΟΡΡΟΠΙΑΣ (1/2)

- Birth-Death Process: Διάγραμμα ***Πιθανοτήτων Μεταβάσεων*** σε χρόνο $\Delta t \to 0$ προς $n(t) = k$

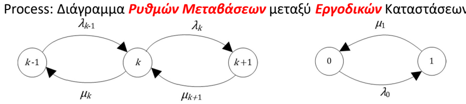


Εξισώσεις Μετάβασης (***Chapman - Kolmogorov***):

$$\mathrm{P}_k(t) = \lambda_{k-1}\Delta t \mathrm{P}_{k-1}(t - \Delta t) + \mu_{k+1}\Delta t \mathrm{P}_{k+1}(t - \Delta t) + [1 - (\lambda_k + \mu_k)\Delta t]\mathrm{P}_k(t - \Delta t), \qquad k \geq 1$$

$$\mathrm{P}_{\mathbf{0}}(t) = \mu_1 \Delta t\, \mathrm{P}_1(t - \Delta t) + (1 - \lambda_0 \Delta t)\, \mathrm{P}_{\mathbf{0}}(t - \Delta t)$$

- Birth-Death Process: Διάγραμμα ***Ρυθμών Μεταβάσεων*** μεταξύ ***Εργοδικών*** Καταστάσεων



Εξισώσεις Ισορροπίας:

$$(\lambda_k + \mu_k)P_k = \lambda_{k-1}P_{k-1} + \mu_{k+1}P_{k+1} \text{ για } k \geq 1 \text{ και } \lambda_0 P_0 = \mu_1 P_1$$

Σχετικές Πιθανότητες Μεταβάσεων $k \to (k+1), k \to (k-1)$:

$\mathrm{P}[k \to (k+1)/\text{μετάβαση}] = \lambda_k/(\lambda_k + \mu_k)$, $\mathrm{P}[k \to (k-1)/\text{μετάβαση}] = \mu_k/(\lambda_k + \mu_k)$

Dwell Time - Χρόνος Παραμονής στην $n(t) = k$ μέχρι την επόμενη μετάβαση

Εκθετική τυχαία μεταβλητή d_k **με μέσο** $1/(\lambda_k + \mu_k)$: Η μικρότερη δύο ανεξάρτητων εκθετικών τυχαίων μεταβλητών μέχρι *(1)* την επόμενη άφιξη με μέσο $1/\lambda_k$ ή *(2)* την ολοκλήρωση εξυπηρέτησης με μέσο $1/\mu_k$

$d_k = \min(x, y)$, $F_{d_k}(\tau) = \mathrm{P}\{d_k \leq \tau\} = 1 - \mathrm{P}\{d_k > \tau\} = 1 - e^{-(\lambda_k + \mu_k)\tau}$ διότι

$\mathrm{P}\{d_k > \tau\} = \mathrm{P}\{x > \tau, y > \tau\} = \mathrm{P}\{x > \tau\}\mathrm{P}\{y > \tau\} = e^{-\lambda_k \tau}e^{-\mu_k \tau} = e^{-(\lambda_k + \mu_k)\tau}$

ΕΞΙΣΩΣΕΙΣ ΙΣΟΡΡΟΠΙΑΣ (2/2)

- Απείρως επισκέψιμες επαναληπτικές καταστάσεις $\boldsymbol{s} = n(t)$ ***positive recurrent states***: Με μη μηδενικές εργοδικές πιθανότητες $\mathrm{P}\{n(t) = k\} = P_k(t) \to P_k > 0, \ k = 0,1,2, \ldots$
- Σε μεγάλο χρονικό διάστημα παρατήρησης Τ ισορροπούν οι αριθμοί μεταβάσεων από και προς την κατάσταση $\boldsymbol{s}$:

$$\#\{\text{ΜΕΤΑΒΑΣΕΩΝ ΠΡΟΣ ΤΗΝ ΚΑΤΑΣΤΑΣΗ } \boldsymbol{s}\} = \#\{\text{ΕΚΤΟΣ ΤΗΣ } \boldsymbol{s}\}$$

Εξισώσεις Σφαιρικής Ισορροπίας, ***Global Balance Equations***

- Υπό συγκεκριμένες συνθήκες (ισχύουν για διαδικασίες ***birth-death***) οι ***Εξισώσεις Σφαιρικής Ισορροπίας*** μπορεί να αντικαταστθούν από απλούστερες ***Εξισώσεις Λεπτομερούς (Ακριβούς) Ισορροπίας***, ***Detailed Balance Equations*** ως εξής:
  - Σε μεγάλο χρονικό διάστημα παρατήρησης Τ αν ισχύουν οι συνθήκες ***Λεπτομερούς Ισορροπίας*** ισορροπούν οι αριθμοί μεταβάσεων μεταξύ δύο αμφίδρομα γειτονικών καταστάσεων $\boldsymbol{s}_1$ και $\boldsymbol{s}_2$:

$$\#\{\text{ΜΕΤΑΒΑΣΕΩΝ } \boldsymbol{s}_1 \to \boldsymbol{s}_2\} = \#\{\text{ΜΕΤΑΒΑΣΕΩΝ } \boldsymbol{s}_2 \to \boldsymbol{s}_1\}$$

  - Λόγω ***εργοδικότητας*** σε μεγάλο χρονικό διάστημα παρατήρησης Τ, με T_1 και T_2 τους συνολικούς χρόνους παραμονής στις $\boldsymbol{s}_1, \boldsymbol{s}_2$:

(1) $\quad \#\{\text{ΜΕΤΑΒΑΣΕΩΝ } \boldsymbol{s}_1 \to \boldsymbol{s}_2\} = \mathrm{T}_1 \times r_{1,2}$

(2) $\quad \#\{\text{ΜΕΤΑΒΑΣΕΩΝ } \boldsymbol{s}_2 \to \boldsymbol{s}_1\} = \mathrm{T}_2 \times r_{2,1}$

όπου $r_{1,2}, r_{2,1}$ οι μέσοι ρυθμοί μεταβάσεων μεταξύ των $\boldsymbol{s}_1$ και $\boldsymbol{s}_2$

(1) = (2) και $r_{1,2} \lim_{\mathrm{T}\to\infty} \frac{\mathrm{T}_1}{\mathrm{T}} = r_{2,1} \lim_{\mathrm{T}\to\infty} \frac{\mathrm{T}_2}{\mathrm{T}}$ ή

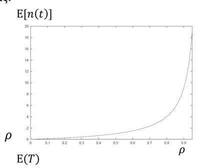
$r_{1,2}P_1 = r_{2,1}P_2$ ***Detailed Balance Equations***

ΟΥΡΑ MARKOV M/M/1 (απείρου μεγέθους)

- Αφίξεις Poisson με μέσο ρυθμό λ αφίξεις/sec: $\lambda_k = \lambda = \gamma, \ k = 0,1,2,3,\cdots$
- Χρόνοι εξυπηρέτησης εκθετικοί με μέση τιμή $\mathrm{E}(s) = \frac{1}{\mu}$sec: $\mu_k = \mu, \ k = 1,2,3,\cdots$
- $\rho = u = \frac{\lambda}{\mu} < 1$ Erlang (συνθήκη για οριακή ισορροπία – εργοδικότητα)
- Οι **εργοδικές πιθανότητες** προκύπτουν από τις εξισώσεις ισορροπίας:
  - $\lambda P_0 = \mu P_1$ ή $P_1 = \left(\frac{\lambda}{\mu}\right) P_0 = \rho P_0$
  - $(\lambda + \mu)P_1 = \lambda P_0 + \mu P_2$ ή $P_2 = \rho^2 P_0$ και $P_k = \rho^k P_0, \ k > 0$
  - $P_0 + P_1 + \cdots + P_k + \cdots = 1 = P_0(1 + \rho + \rho^2 + \rho^3 + \cdots)$

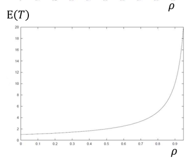
  Με $0 < \rho < 1$ η άπειρη δυναμοσειρά συγκλίνει, $P_0\left(\frac{1}{1-\rho}\right) = 1 \Rightarrow$

  $P_0 = (1 - \rho)$, $P_k = (1 - \rho)\rho^k, k > 0$ και $\mathrm{P}\{n(t) > 0\} = 1 - P_0 = \rho$



- **Μέση κατάσταση συστήματος** M/M/1 σε ισορροπία:

$$\mathrm{E}[n(t)] = \sum_{k=1}^{\infty} kP_k = \frac{\rho}{1 - \rho}$$



- **Μέσος χρόνος καθυστέρησης**: Τύπος Little

$$\mathrm{E}(T) = \frac{\mathrm{E}[n(t)]}{\gamma} = \frac{\mathrm{E}[n(t)]}{\lambda} = \frac{1/\mu}{1 - \rho}$$

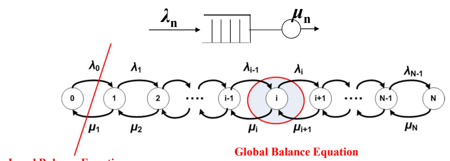
- **Μέσο μήκος ουράς & μέσος χρόνος αναμονής M/M/1**:

$$\mathrm{E}[n_q(t)] = \mathrm{E}[n(t)] - \rho, \qquad \mathrm{E}(W) = \mathrm{E}(T) - 1/\mu$$

Ουρές M/M/1 – Ρυθμοί εξαρτώμενοι από παρούσα κατάσταση (state dependent)

- Συστήματα M/M/1 με ρυθμούς άφιξης και ρυθμούς εξυπηρέτησης εξαρτώμενους από τον αριθμό των πελατών στο σύστημα (από την παρούσα κατάσταση του συστήματος)

  (State Dependent M/M/1 Queues)



Local Balance Equation

$\lambda_0 P_0 = \mu_1 P_1$

$\lambda_{i-1}P_{i-1} = \mu_i P_i$, i = 1, 2, …N

Global Balance Equation

$(\lambda_i + \mu_i)P_i = \lambda_{i-1}P_{i-1} + \mu_{i+1}P_{i+1}$, i = 0, 1,…, N

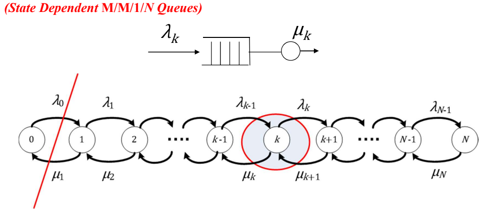
Κανονικοποίηση Εργοδικών Πιθανοτήτων

$P_0 + \ldots + P_N = 1$

ΟΥΡΑ M/M/1/*N* (1/2)

- Συστήματα M/M/1/*N* με ρυθμούς άφιξης και ρυθμούς εξυπηρέτησης εξαρτώμενους από τον αριθμό των πελατών στο σύστημα (από την παρούσα κατάσταση του συστήματος)

  (State Dependent M/M/1/N Queues)



Detailed Balance Equations

$\lambda_0 P_0 = \mu_1 P_1$

$\lambda_{k-1}P_{k-1} = \mu_k P_k \qquad k = 1, 2, \ldots N$

Global Balance Equations

$\lambda_0 P_0 = \mu_1 P_1$

$(\lambda_k + \mu_k)P_k = \lambda_{k-1}P_{k-1} + \mu_{k+1}P_{k+1} \qquad k = 1, \ldots, N$

Κανονικοποίηση Εργοδικών Πιθανοτήτων

$P_0 + \cdots + P_N = 1$

ΟΥΡΑ M/M/1/*N* (2/2)

- Σταθεροί μέσοι ρυθμοί αφίξεων (γεννήσεων)

  $\lambda_k = \lambda$, Poisson, $k = 1,2, \ldots, N$
- Σταθεροί μέσοι ρυθμοί εξυπηρέτησης (θανάτων)

  $\mu_k = \mu, \ k = 1,2, \ldots, N$

  Εκθετικοί ανεξάρτητοι χρόνοι εξυπηρέτησης $s, E(s) = 1/\mu$
- Εργοδικές πιθανότητες καταστάσεων

  $P_k = \rho^k P_0, \ k = 0,1,2, \ldots, N$

  $P_0 + P_1 + \cdots + P_{N-1} + P_N = 1$

  $\rho = \lambda/\mu$ Erlangs (η M/M/1/*N* είναι **πάντα ευσταθής** γιατί υπερβολικό φορτίο δεν προωθείται)
- Αντικαθιστώντας με τον τύπο πεπερασμένου αθροίσματος γεωμετρικής προόδου:

$$P_0 = \frac{1 - \rho}{1 - \rho^{N+1}}, \qquad \rho \neq 1$$

$$P_0 = \frac{1}{N + 1}, \qquad \rho = 1$$

- Χρησιμοποίηση Εξυπηρετητή (Server Utilization) $U = 1 - P_0$
- Ρυθμαπόδοση (throughput) $\gamma = \lambda(1 - P_N) = \mu(1 - P_0) = \mu U$
- Πιθανότητα απώλειας $P_{blocking} = P_N$
- Στάσιμος Εργοδικός μέσος όρος πληθυσμού – κατάστασης

$$\mathrm{E}[n(t)] \to \mathrm{E}(k) = \sum_{k=1}^{N} kP_k = \rho \frac{1 - (N + 1)\rho^N + N\rho^{N+1}}{(1 - \rho)(1 - \rho^{N+1})}$$

- Νόμος του Little: $\mathrm{E}(T) = \mathrm{E}(k)/\gamma = \mathrm{E}(k)/[\lambda(1 - P_N)]$

ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ (Queuing Systems)

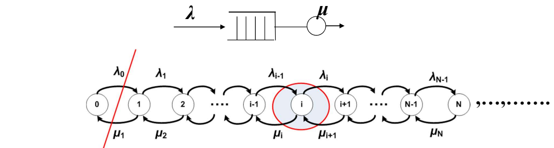
Διαδικασίες Birth-Death & Ουρές Markov:

- Προσομοιώσεις
- Συστήματα Αναμονής M/M/c (Erlang-C), M/M/2, M/M/N/K, M/M/c/c (Erlang-B)
- Παραδείγματα και Εφαρμογή: Ανάλυση & Σχεδιασμός
  - a) Τηλεφωνικών Κέντρων
  - b) Βελτιστοποίηση Μέσου Μήκους Πακέτου

Εβδομάδα 6ης Απριλίου & 27ης Απριλίου, 2020

Ουρές M/M/1 (Επανάληψη)
Local Balance Equation - Global Balance Equation

$\lambda_0 = \lambda_1 = \lambda_2 = \ldots = \lambda_i = \ldots\ldots$, i = 0, 1, 2, ….,
$\mu_1 = \mu_2 = \ldots = \mu_i = \ldots\ldots$, i = 1, 2, ….,



Local Balance Equation
$\lambda P_0 = \mu P_1$
$\lambda P_1 = \mu P_2$
……………
$\lambda P_{i-1} = \mu P_i$, i = 1, 2, …

Global Balance Equation
$(\lambda+\mu)P_i = \lambda_i P_{i-1} + \mu P_{i+1}$, i = 0, 1, 2, ….,

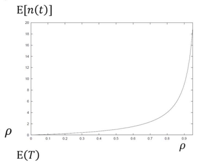
Κανονικοποίηση Εργοδικών Πιθανοτήτων
$P_0 + P_1 + \ldots + P_i + \ldots = 1$

ΟΥΡΑ MARKOV M/M/1 (Επανάληψη)

- Αφίξεις Poisson με μέσο ρυθμό λ αφίξεις/sec: $\lambda_k = \lambda = \gamma,\ \ k = 0,1,2,3,\cdots$
- Χρόνοι εξυπηρέτησης εκθετικοί με μέση τιμή $\mathrm{E}(s) = \frac{1}{\mu}$sec: $\mu_k = \mu,\ k = 1,2,3,\cdots$
- $\rho = u = \frac{\lambda}{\mu} < 1$ Erlang (συνθήκη για οριακή ισορροπία – εργοδικότητα)
- Οι **εργοδικές πιθανότητες** προκύπτουν από τις εξισώσεις ισορροπίας:
  - $\lambda P_0 = \mu P_1$ ή $P_1 = \left(\frac{\lambda}{\mu}\right) P_0 = \rho P_0$
  - $(\lambda+\mu)P_1 = \lambda P_0 + \mu P_2$ ή $P_2 = \rho^2 P_0$ και $P_k = \rho^k P_0,\ \ k > 0$
  - $P_0 + P_1 + \cdots + P_k + \cdots = 1 = P_0(1 + \rho + \rho^2 + \rho^3 + \cdots)$

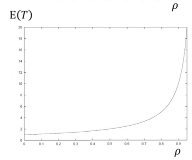
  Με $0 < \rho < 1$ η άπειρη δυναμοσειρά συγκλίνει, $P_0\left(\frac{1}{1-\rho}\right) = 1 \Rightarrow$

  $P_0 = (1-\rho),\ \ P_k = (1-\rho)\rho^k, k > 0$ και $\mathrm{P}\{n(t) > 0\} = 1 - P_0 = \rho$



- **Μέση κατάσταση συστήματος** M/M/1 σε ισορροπία:

$$\mathrm{E}[n(t)] = \sum_{k=1}^{\infty} kP_k = \frac{\rho}{1-\rho}$$



- **Μέσος χρόνος καθυστέρησης**: Τύπος Little

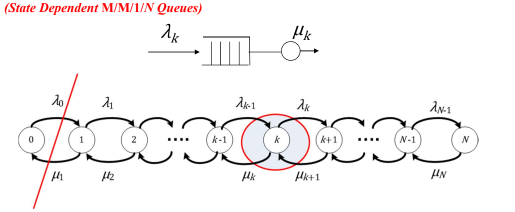
$$\mathrm{E}(T) = \frac{\mathrm{E}[n(t)]}{\gamma} = \frac{\mathrm{E}[n(t)]}{\lambda} = \frac{1/\mu}{1-\rho}$$

- **Μέσο μήκος ουράς & μέσος χρόνος αναμονής M/M/1**:

$$\mathrm{E}\left[n_q(t)\right] = \mathrm{E}[n(t)] - \rho, \qquad \mathrm{E}(W) = \mathrm{E}(T) - 1/\mu$$

ΟΥΡΑ M/M/1/N (Επανάληψη)

- Συστήματα M/M/1/N με ρυθμούς άφιξης και ρυθμούς εξυπηρέτησης εξαρτώμενους από τον αριθμό των πελατών στο σύστημα (από την παρούσα κατάσταση του συστήματος) *(State Dependent M/M/1/N Queues)*



Detailed Balance Equations
$\lambda_0 P_0 = \mu_1 P_1$
$\lambda_{k-1}P_{k-1} = \mu_k P_k \qquad k = 1, 2, \ldots N$

Global Balance Equations
$\lambda_0 P_0 = \mu_1 P_1$
$(\lambda_k + \mu_k)P_k = \lambda_{k-1}P_{k-1} + \mu_{k+1}P_{k+1} \qquad k = 1, \ldots, N$

Κανονικοποίηση Εργοδικών Πιθανοτήτων
$P_0 + \cdots + P_N = 1$

ΟΥΡΑ M/M/1/N (Επανάληψη)

- Σταθεροί μέσοι ρυθμοί αφίξεων (γεννήσεων)
  $\lambda_k = \lambda$, Poisson, $k = 1, 2, \ldots, N$
- Σταθεροί μέσοι ρυθμοί εξυπηρέτησης (θανάτων)
  $\mu_k = \mu,\ k = 1, 2, \ldots, N$
  Εκθετικοί ανεξάρτητοι χρόνοι εξυπηρέτησης $s, E(s) = 1/\mu$
- Εργοδικές πιθανότητες καταστάσεων
  $P_k = \rho^k P_0,\ \ k = 0, 1, 2, \ldots, N$
  $P_0 + P_1 + \cdots + P_{N-1} + P_N = 1$
  $\rho = \lambda/\mu$ Erlangs (η M/M/1/N είναι **πάντα ευσταθής** γιατί υπερβολικό φορτίο δεν προωθείται)
- Αντικαθιστώντας με τον τύπο πεπερασμένου αθροίσματος γεωμετρικής προόδου:

$$P_0 = \frac{1-\rho}{1-\rho^{N+1}}, \qquad \rho \neq 1$$

$$P_0 = \frac{1}{N+1}, \qquad \rho = 1$$

- Χρησιμοποίηση Εξυπηρετητή (Server Utilization) $U = 1 - P_0$
- Ρυθμαπόδοση (throughput) $\gamma = \lambda(1 - P_N) = \mu(1 - P_0) = \mu U$
- Πιθανότητα απώλειας $P_{blocking} = P_N$
- Στάσιμος Εργοδικός μέσος όρος πληθυσμού – κατάσταση

$$\mathrm{E}[n(t)] \rightarrow \mathrm{E}(k) = \sum_{k=1}^{N} kP_k = \rho\frac{1-(N+1)\rho^N + N\rho^{N+1}}{(1-\rho)(1-\rho^{N+1})}$$

- Νόμος του Little: $\mathrm{E}(T) = \mathrm{E}(k)/\gamma = \mathrm{E}(k)/[\lambda(1 - P_N)]$

ΠΡΟΣΟΜΟΙΩΣΗ ΣΤΟΧΑΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ BIRTH-DEATH ΟΜΟΙΟΓΕΝΩΝ ΑΦΙΞΕΩΝ (1/2)

- Σε στοχαστικό σύστημα ***Birth-Death*** με αφίξεις (γεννήσεις) σταθερού μέσου ρυθμού λ ανεξάρτητου του πληθυσμού $n(t) = k$ (***ομοιογενείς αφίξεις Poisson*** με ρυθμό $\lambda_k = \lambda$ αφίξεις/sec) οι εργοδικές πιθανότητες (αν υπάρχουν) μπορούν να υπολογισθούν σαν λόγος αφίξεων που βρίσκουν το σύστημα στη κατάσταση $n(t) = k$ στη χρονική διάρκεια T_k, προς τον συνολικό αριθμό αφίξεων σε μεγάλο χρονικό διάστημα παρατήρησης T μιας χρονικής εξέλιξης της διαδικασίας:

$$P_n = \lim_{\mathrm{T}\to\infty}\frac{\mathrm{T}_k}{\mathrm{T}} = \lim_{\mathrm{T}\to\infty}\frac{\lambda\mathrm{T}_k}{\lambda\mathrm{T}} \approx \frac{\#\{\text{ΑΦΙΞΕΩΝ στη } n(t) = k \text{ σε ΧΡΟΝΙΚΟ ΔΙΑΣΤΗΜΑ } \mathrm{T}_k\}}{\#\{\text{ΣΥΝΟΛΟΥ ΑΦΙΞΕΩΝ σε ΧΡΟΝΙΚΟ ΔΙΑΣΤΗΜΑ T}\}}$$

  Άρα μπορούμε να προσομοιώσουμε σύστημα Birth-Death με ομοιογενείς αφίξεις καταμετρώντας τις αφίξεις στις διάφορες καταστάσεις που μεταβαίνει

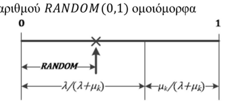
- Η εξέλιξη της κατάστασης (πληθυσμού) του συστήματος προκύπτει από τις πιθανότητες μετάβασης από την κατάσταση $n(t) = k$ στις $(k+1)$, $(k-1)$ με το δεδομένο ότι μια από τις δύο μεταβάσεις θα συμβεί με απόλυτη βεβαιότητα:

$$\mathrm{P}[k \rightarrow (k+1)/\text{μετάβαση}] = \lambda/(\lambda+\mu_k),\ \ \mathrm{P}[k \rightarrow (k-1)/\text{μετάβαση}] = \mu_k/(\lambda+\mu_k)$$

- Η προσομοίωση ενεργοποιεί τις μεταβάσεις με κλήση τυχαίου αριθμού $RANDOM(0,1)$ ομοιόμορφα κατανεμημένου μεταξύ $(0, 1)$:

$$0 \leq RANDOM(0,1) \leq \frac{\lambda}{\lambda+\mu_k} \Rightarrow \text{ΑΦΙΞΗ}, n(t) \rightarrow k+1$$

$$\frac{\lambda}{\lambda+\mu_k} < RANDOM(0,1) \leq 1 \Rightarrow \text{ΑΝΑΧΩΡΗΣΗ}, n(t) \rightarrow k-1$$



- Αν το σύστημα έχει μηδενικό πληθυσμό $n(t) = 0$, η επόμενη μετάβαση είναι πάντα ΑΦΙΞΗ και $n(t) \rightarrow 1$
- Αν το σύστημα δεν επιδέχεται αύξηση πληθυσμού, η $n(t) = K$ είναι blocking state και δεν ενεργοποιείται μετάβαση κατάστασης, αλλά η διαδικασία τυχαίας δημιουργίας επόμενου γεγονότος (ΑΦΙΞΗ ή ΑΝΑΧΩΡΗΣΗ) καθώς και η μέτρηση αφίξεων συνεχίζονται κανονικά

ΠΡΟΣΟΜΟΙΩΣΗ ΣΤΟΧΑΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ BIRTH-DEATH ΟΜΟΙΟΓΕΝΩΝ ΑΦΙΞΕΩΝ (2/2)

Στατιστική Σύγκλιση Προσομοίωσης

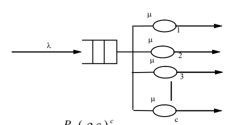
- Η σύγκλιση της προσομοίωσης ελέγχεται ως προς την ***στατιστική σύγκλιση*** μεγεθών ενδιαφέροντος, π.χ. εκτίμηση μέσου πληθυσμού $\mathrm{E}[n(t)] = \sum_{k=0}^{K} k * P_k$ που υπολογίζεται σε τακτά διαστήματα από την αρχή της προσομοίωσης π.χ. κάθε 1000 αφίξεις: (από 0 έως 1000 αφίξεις), (από 0 έως 2000 αφίξεις), (από 0 έως 3000 αφίξεις) κλπ.
- Λόγω ***εργοδικότητας*** δεν απαιτούνται επαναλήψεις: Η εκτέλεση του προγράμματος και οι μετρήσεις συνεχίζονται σε μία υλοποίηση που διακόπτεται προσωρινά π.χ. κάθε 1000 αφίξεις μέχρι την ικανοποίηση κριτηρίων σύγκλισης
- Η στατιστική σύγκλιση επιταχύνεται αν αγνοήσουμε στη καταμέτρηση αφίξεων στις διάφορες καταστάσεις τις πρώτες μεταβάσεις (π.χ. 1-1000 αφίξεις) που αντιστοιχούν στο ***μεταβατικό φαινόμενο*** προς την εργοδική κατάσταση
- **Γενική Παρατήρηση**: Η σύγκλιση μιας προσομοίωσης είναι σύνηθες να διερευνάται με μαθηματικά εργαλεία στατιστικής γιατί μπορεί να κρύβονται εξαρτήσεις (correlations) μεγεθών, περιοδικές συμπεριφορές που οδηγούν σε πρόωρους τερματισμούς κλπ. Μια πλήρης προσομοίωση περιλαμβάνει διαστήματα εμπιστοσύνης για τις εκτιμήσεις πιθανοτήτων και ροπών τυχαίων μεταβλητών (π.χ. μέσοι όροι, διασπορά)

ΠΡΟΣΟΜΟΙΩΣΗ ΟΥΡΑΣ M/M/1/10

```
RANDOM: Ομοιόμορφος τυχαίος αριθμός (0,1)
THRESHOLD: λ/(λ+μ)
ARRIVALS: Συνολικός αριθμός αφίξεων
ARRIVAL[STATE]: Αριθμός αφίξεων στην κατάσταση STATE = 0, 1, … ,10
COUNT: Αριθμός μεταβάσεων COUNT = 0, 1, … , MAXIMUM
STATE: Κατάσταση ουράς (πληθυσμός συστήματος M/M/1/10) , STATE = 0, 1, … , 10
P[STATE]: Εργοδική πιθανότητα STATE = 0, 1, … , 10
AVERAGE: Μέσος πληθυσμός συστήματος M/M/1/10

INITIALIZE: COUNT = 0, STATE = 0, ARRIVALS = 0, ARRIVAL[0…10] = 0, P[0…10] = 0
ARRIVAL:    ARRIVALS = ARRIVALS + 1
            ARRIVAL[STATE] = ARRIVAL[STATE] + 1
            COUNT = COUNT +1
            IF STATE = 10 : GO TO LOOP
            ELSE : STATE = STATE + 1
                          GO TO LOOP
LOOP :      IF STATE = 0 : GO TO ARRIVAL
            ELSE :
                 IF RANDOM < THRESHOLD : GO TO ARRIVAL
                 ELSE : GO TO DEPARTURE
DEPARTURE  :  COUNT  = COUNT +1 ; STATE = STATE - 1
                 IF COUNT < MAXIMUM : GO TO LOOP
                 ELSE : P[STATE=1…10] = ARRIVAL[STATE= 1…10] / ARRIVALS
                        AVERAGE = SUM { STATE * P[STATE] }, STATE = [1…10]
```

Παράδειγμα ανάλυσης ουράς Markov με m εξυπηρετητές M/M/c [Erlang –C]



Αφίξεις Poisson με ομοιόμορφο μέσο ρυθμό
$\lambda_n = \lambda$ (εξωτερικές κλήσεις - τηλεφωνήματα/sec)
c ανεξάρτητοι εκθετικοί εξυπηρετητές (εξωτερικές γραμμές τηλεφωνικού κέντρου)
Ρυθμοί εξυπηρέτησης στη κατάσταση k:
$\mu_k = k\mu,\ \ k = 1, 2, \ldots, c$
$\boldsymbol{\mu_k = c\mu, \quad k = c+1, c+2, \ldots}$

$$P_Q = \frac{P_0(\rho c)^c}{c!(1-\rho)}$$ Prob. All servers are busy

$$N_Q = P_Q\frac{\rho}{1-\rho}$$

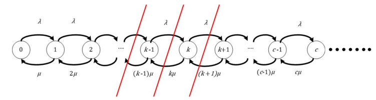
$$W = \frac{\rho \cdot P_Q}{\lambda(1-\rho)}$$

$$T = \frac{p_Q}{c\mu - \lambda} + \frac{1}{\mu}$$

$$N = \frac{\rho p_Q}{1-\rho} + c\rho$$

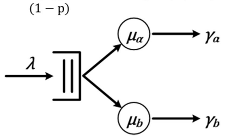
$$p_0 = \frac{1}{\left[\sum_{n=0}^{c-1}\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{c!}\left(\frac{\lambda}{\mu}\right)^c\frac{c\mu}{c\mu-\lambda}}$$

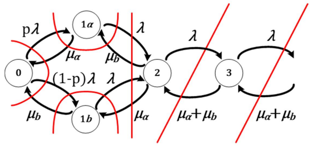
ρ’ = λ/μ Erlangs, ρ = ρ’/c< 1



ΟΥΡΑ M/M/2

- Αφίξεις Poisson με ομοιόμορφο μέσο ρυθμό $\lambda_k = \lambda$
- 2 ανεξάρτητοι εκθετικοί εξυπηρετητές $(\alpha), (b)$ με άνισους ρυθμούς μ_α και μ_b
- Άπειρη Χωρητικότητα
- Άφιξη σε άδειο σύστημα δρομολογείται στον (α) με πιθανότητα p και στον (b) με πιθανότητα $(1-\text{p})$





Εξισώσεις Ισορροπίας:

$\lambda P_0 = \mu_\alpha P_{1\alpha} + \mu_b P_{1b}$

$(\lambda + \mu_\alpha) P_{1\alpha} = \text{p}\lambda P_0 + \mu_b P_2$

$(\lambda + \mu_b) P_{1b} = (1-\text{p})\lambda P_0 + \mu_\alpha P_2$

$\lambda(P_{1\alpha} + P_{1b}) = (\mu_\alpha + \mu_b) P_2$, $\lambda P_k = (\mu_a + \mu_b) P_{k+1}$, $k = 2, 3, \dots$

$P_0 + P_{1\alpha} + P_{1b} + P_2 + P_3 + \cdots = 1$, $\lambda/(\mu_\alpha + \mu_b) < 1$ για σύγκληση (εργοδικότητα)

Βαθμοί Χρησιμοποίησης – Ρυθμαποδόσεις Εξυπηρετητών:

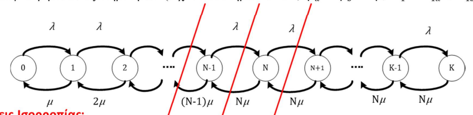
$U_\alpha = 1 - P_0 - P_{1b}$ $\quad \gamma_\alpha = \mu_\alpha U_\alpha$

$U_b = 1 - P_0 - P_{1\alpha}$ $\quad \gamma_b = \mu_b U_b$

$\gamma = \lambda = \gamma_a + \gamma_b$

ΟΥΡΑ M/M/N/K

- Αφίξεις Poisson με ομοιόμορφο μέσο ρυθμό $\lambda_k = \lambda$
- N ανεξάρτητοι εκθετικοί εξυπηρετητές με ίσους ρυθμούς μ
- Χωρητικότητα K, $\text{N} \leq \text{K}$ (π.χ. **call center** με N εξυπηρετητές & δυνατότητα αναμονής μέχρι K − N κλήσεις)
- Μέσοι ρυθμοί εξυπηρέτησης στη κατάσταση k:
  - $\mu_k = k\mu$, $k = 1, 2, \dots, \text{N}$
  - $\mu_k = \text{N}\mu$, $k = \text{N}, \text{N}+1, \dots, \text{K}-1, \text{K}$
- Εργοδική κατάσταση $n(t)$: Αριθμός πελατών στο σύστημα, αδιάφορα από χρήση συγκεκριμένων εξυπηρετητών (π.χ. σε σύστημα M/M/N, $\mu_\alpha = \mu_b = \mu$, $P_1 = P_{1\alpha} + P_{1b}$



Εξισώσεις Ισορροπίας:

$P_k = \left[\dfrac{\lambda}{k\mu}\right] P_{k-1}, \qquad k = 1, 2, \dots, \text{N}-1$

$P_k = \left[\dfrac{\lambda}{\text{N}\mu}\right] P_{k-1}, \qquad k = \text{N}, \text{N}+1, \dots, \text{K}-1, \text{K}$

$P_0 + P_1 + \cdots + P_{\text{K}-1} + \; + P_\text{K} = 1, \qquad P_\text{K} = P_{\text{blocking}}, \qquad \gamma = \lambda(1 - P_{\text{blocking}})$

$P_{\text{waiting}} = P_\text{N} + P_{\text{N}+1} + \cdots + P_{\text{K}-1}$

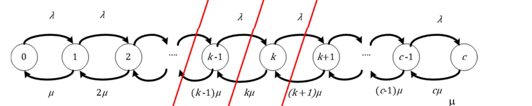
Για $\text{K} \to \infty$ σύστημα M/M/N και $P_{\text{waiting}} = P_\text{N} + P_{\text{N}+1} + \cdots = 1 - \sum_{k=0}^{\text{N}-1} P_k$

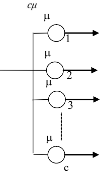
ΟΥΡΑ M/M/*c*/*c* (*Erlang-B*)

(τηλεφωνικό κέντρο με *c* εξωτερικές γραμμές, trunks)

- Αφίξεις Poisson με ομοιόμορφο μέσο ρυθμό $\lambda_n = \lambda$ (εξωτερικές κλήσεις - τηλεφωνήματα/sec)
- c ανεξάρτητοι εκθετικοί εξυπηρετητές (εξωτερικές γραμμές τηλεφωνικού κέντρου)
- Χωρητικότητα c πελάτες (τηλεφωνήματα, εξωτερικές κλήσεις)
- Ρυθμοί εξυπηρέτησης στη κατάσταση k:

  $\mu_k = k\mu$, $k = 1, 2, \dots, c$ $\quad \frac{1}{\mu} = \text{E}(s)$: Μέση Διάρκεια Τηλεφωνήματος (π.χ. 3 min ή 180 sec)





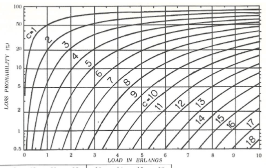
Εξισώσεις Ισορροπίας:

$P_k = \left[\frac{\lambda}{k\mu}\right] P_{k-1} = \left(\frac{\rho^k}{k!}\right) P_0$, $k = 1, 2, \dots, c \qquad \rho \triangleq \frac{\lambda}{\mu}$ Erlangs

$P_0 + P_1 + \cdots + P_{c-1} + P_c = 1 \Rightarrow \text{P}_0 = \frac{1}{\sum_{k=1}^{c} \frac{\rho^k}{k!}}$

$P_c = P_{\text{blocking}} = \dfrac{\rho^c/c!}{\sum_{k=0}^{c} \frac{\rho^k}{k!}} \triangleq \text{B}(\rho, c)$ (***Erlang-B Formula***)

ΠΙΝΑΚΕΣ Erlang $\text{B}(\rho, c)$



Αναδρομικός Υπολογισμός

$\text{B}(\rho, 0) = 1$

$\text{B}(\rho, n) = \dfrac{\rho \text{B}(\rho, n-1)}{\rho \text{B}(\rho, n-1) + n}, \qquad n = 1, 2, \dots, c$

ρ	$c=1$	2	3	4	5	6	7	8
0.00	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.25	.2000	.0244	.0020	.0001	**	*	*	*
0.50	.3333	.0769	.0127	.0016	.0002	**	*	*
0.75	.4286	.1385	.0335	.0062	.0010	.0001	**	*
1.00	.5000	.2000	.0625	.0154	.0031	.0005	.0001	**
1.25	.5556	.2577	.0970	.0294	.0073	.0015	.0003	**
1.50	.6000	.3103	.1343	.0480	.0142	.0035	.0008	.0001
1.75	.6364	.3577	.1726	.0702	.0240	.0070	.0017	.0004
2.00	.6667	.4000	.2105	.0952	.0367	.0121	.0034	.0009
2.25	.6923	.4378	.2472	.1221	.0521	.0192	.0061	.0017
2.50	.7143	.4717	.2822	.1499	.0697	.0282	.0100	.0031
2.75	.7333	.5021	.3152	.1781	.0892	.0393	.0152	.0052
3.00	.7500	.5294	.3462	.2062	.1101	.0522	.0219	.0081
3.25	.7647	.5541	.3751	.2336	.1318	.0666	.0300	.0121
3.50	.7778	.5765	.4021	.2603	.1541	.0825	.0396	.0170
3.75	.7895	.5968	.4273	.2860	.1766	.0994	.0506	.0232
4.00	.8000	.6154	.4507	.3107	.1991	.1172	.0628	.0304
4.25	.8095	.6324	.4725	.3343	.2213	.1355	.0760	.0388
4.50	.8182	.6480	.4929	.3567	.2430	.1542	.0902	.0483
4.75	.8261	.6624	.5119	.3781	.2643	.1730	.1051	.0587
5.00	.8333	.6757	.5287	.3983	.2849	.1919	.1205	.0701
5.25	.8400	.6880	.5463	.4176	.3048	.2106	.1364	.0822
5.50	.8462	.6994	.5618	.4358	.3241	.2290	.1525	.0949
5.75	.8519	.7101	.5764	.4531	.3426	.2472	.1688	.1082
6.00	.8571	.7200	.5902	.4696	.3604	.2649	.1851	.1219
6.25	.8621	.7293	.6031	.4852	.3775	.2822	.2013	.1359
6.50	.8667	.7380	.6152	.4999	.3939	.2991	.2174	.1501
6.75	.8710	.7462	.6267	.5140	.4096	.3155	.2333	.1644
7.00	.8750	.7539	.6376	.5273	.4247	.3313	.2489	.1788
7.25	.8788	.7611	.6478	.5401	.4392	.3467	.2642	.1932

ρ	$c=9$	10	11	12	13	14	15	16
0.00	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
0.25								
0.50								
0.75								
1.00	*							
1.25	**							
1.50	**	*						
1.75	.0001	**	*					
2.00	.0002	**	**					
2.25	.0004	.0001	**	*				
2.50	.0009	.0002	.0001	**	*			
2.75	.0016	.0004	.0001	**	**			
3.00	.0027	.0008	.0002	.0001	**	*		
3.25	.0043	.0014	.0004	.0001	**	**		
3.50	.0066	.0023	.0007	.0002	.0001	**	*	
3.75	.0096	.0036	.0012	.0004	.0001	**	**	
4.00	.0133	.0053	.0019	.0006	.0002	.0001	**	*
4.25	.0180	.0076	.0029	.0010	.0003	.0001	**	**
4.50	.0236	.0105	.0043	.0016	.0006	.0002	.0001	**
4.75	.0301	.0141	.0060	.0024	.0009	.0003	.0001	**
5.00	.0375	.0184	.0083	.0034	.0013	.0005	.0002	.0001
5.25	.0457	.0234	.0111	.0048	.0019	.0007	.0003	.0001
5.50	.0548	.0293	.0144	.0066	.0028	.0011	.0004	.0001
5.75	.0647	.0358	.0184	.0087	.0039	.0016	.0006	.0002
6.00	.0751	.0431	.0230	.0114	.0052	.0022	,0009	.0003
6.25	.0862	.0511	.0282	.0145	.0069	.0031	.0013	.0005
6.50	.0978	.0598	.0341	.0181	.0090	.0042	.0018	.0007
6.75	.1098	.0690	.0406	.0223	.0115	.0055	.0025	.0010
7.00	.1221	.0787	.0477	.0271	.0144	.0071	.0033	.0015
7.25	.1347	.0890	.0554	.0324	.0177	.0091	.0044	.0020

ΠΑΡΑΔΕΙΓΜΑ ΑΝΑΛΥΣΗΣ ΤΗΛΕΦΩΝΙΚΟΥ ΚΕΝΤΡΟΥ

- Τηλεφωνικό Κέντρο με 7 εξωτερικές γραμμές προωθεί κίνηση Θεωρώ ότι οι εξωτερικές κλήσεις ακολουθούν διαδικασία ***Poisson*** με μέσο ρυθμό $\lambda = 2$ κλήσεις/min και χρόνο εξυπηρέτησης ***Εκθετικό*** με μέση διάρκεια $1/\mu = 3$ min, άρα το συνολικό προσφερόμενο φορτίο (***offered traffic***) είναι

  $\rho = \lambda/\mu = 6$ Erlangs
- Υποθέτουμε πως οι κλήσεις που δεν βρίσκουν γραμμή χάνονται οριστικά. Άρα η πιθανότητα απώλειας δίνεται από τον τύπο

  $\text{B}(\rho, c) = \text{B}(6,7) = 18.51\%$
- Το εξυπηρετούμενο φορτίο (***carried traffic***) είναι

  $\rho[1 - \text{B}(\rho, c)] = \frac{\lambda}{\mu}[1 - \text{B}(\rho, c)] = \frac{\gamma}{\mu} = 4.8894$ Erlangs
- Το φορτίο υπερχείλισης (***overflow traffic***) είναι

  $\rho B(\rho, c) = 1.1106$ Erlangs

ΠΑΡΑΔΕΙΓΜΑ ΣΧΕΔΙΑΣΜΟΥ ΤΗΛΕΦΩΝΙΚΟΥ ΚΕΝΤΡΟΥ

- Τηλεφωνικό Κέντρο με c εξωτερικές γραμμές (trunks) προωθεί κίνηση με μέσο ρυθμό κλήσεων 2 κλήσεις το λεπτό με μέση διάρκεια κλήσης 3 min.
- Θεωρώ ότι οι εξωτερικές κλήσεις ακολουθούν διαδικασία ***Poisson*** με μέσο ρυθμό $\lambda = 2$ κλήσεις/min και χρόνο εξυπηρέτησης ***Εκθετικό*** με μέση διάρκεια $\frac{1}{\mu} = 3$ min, άρα το συνολικό προσφερόμενο φορτίο (***offered traffic***) είναι

  $\rho = \frac{\lambda}{\mu} = 6$ Erlangs
- Υποθέτουμε πως οι κλήσεις που δεν βρίσκουν γραμμή χάνονται οριστικά. Ζητείται ο απαιτούμενος αριθμός εξωτερικών γραμμών (trunks) c ώστε ο ρυθμός απωλειών (***Grade of Service, GOS***) να είναι μικρότερος από 0.3%
- Από τους πίνακες προκύπτει πως $\text{B}(6,13) = 0.52\%$ και $\text{B}(6,14) = 0.24\%$, άρα οι απαιτήσεις καλύπτονται με ελάχιστο αριθμό εξωτερικών γραμμών $c = 14$ trunks

ΠΑΡΑΔΕΙΓΜΑ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ΜΕΣΟΥ ΜΗΚΟΥΣ ΠΑΚΕΤΟΥ ΣΕ ΔΙΚΤΥΟ ΤΥΠΟΥ INTERNET

- *10* υπολογιστές (Η/Υ) διασυνδέονται σε Δίκτυο μέσω μεταγωγέα πακέτων (***Router*** ή ***Ethernet Switch***) που τα προωθεί προς τον προορισμό τους. Η ταχύτητα της πολυπλεγμένης εξόδου (***trunk port***) είναι $C = 100$ Mbps
- Τα δεδομένα παράγονται στους Η/Υ σε μορφή πακέτων (πλαισίων) μεταβλητού μήκους L bits (***data payload***). Θεωρείστε πως κάθε Η/Υ παράγει δεδομένα που αντιστοιχούν σε 1 Mbps κατά μέσο όρο
- Οι Η/Υ προσθέτουν σε κάθε πακέτο επικεφαλίδα (***header***) με υποχρεωτικές πληροφορίες πρωτοκόλλου (***protocol overhead*** με διευθύνσεις, σηματοδοσία ελέγχου, ανίχνευσης λαθών κλπ.) μήκους 200 bits
- Θεωρείστε πως ο μεταγωγέας έχει άπειρη χωρητικότητα αποθήκευσης πακέτων, το συνολικό μήκος πακέτου $(L + 200)$ bits είναι κατά προσέγγιση ***Εκθετικά*** κατανεμημένο και πως η συνολική ροή πακέτων γίνεται με διαδικασία ***Poisson***
- Βρείτε το μέσο ωφέλιμο μήκος πακέτου $\text{E}(L)$ που να βελτιστοποιεί την μέση καθυστέρηση προώθησης πακέτου στο μεταγωγέα
- Θεωρείστε πως η ανάστροφη ροή πακέτων ***Δίκτυο→ Η/Υ*** γίνεται ανεξάρτητα από την ροή ***Η/Υ→ Δίκτυο*** (FDX) και πως η στατιστική συμπεριφορά των δύο κατευθύνσεων είναι συμμετρική

Λύση

- Θεωρώ μοντέλο ουράς M/M/1 με

  $\lambda = \frac{10 \times 10^6}{\text{E}(L)} = 10^7/\text{E}(L)$ packets/sec

  $\mu = \frac{C}{200 + \text{E}(L)}$ sec^{-1} $= 10^8/[200 + \text{E}(L)]$ sec^{-1}
- Η μέση καθυστέρηση δίνεται από τον τύπο

$$\text{E}(T) = \frac{1/\mu}{1 - \lambda/\mu} = \frac{1}{\mu - \lambda}$$

$$\text{E}(T) = \frac{1}{\dfrac{10^8}{200 + \text{E}(L)} - \dfrac{10^7}{\text{E}(L)}}$$

- Με $\text{E}(L) = x$, ελαχιστοποιώ την συνάρτηση

$$f(x) = \frac{1}{\dfrac{10^8}{200 + x} - \dfrac{10^7}{x}} = \frac{x(200 + x)}{10^7(9x - 200)}$$

και βρίσκω το βέλτιστο μέσο payload ανά πακέτο $\text{E}(L) = x$ όταν $\frac{df(x)}{dx} = 0 \Rightarrow$

optimal$\{\text{E}(L)\} \cong 92.495$ bits

- **Προσοχή:** Για εργοδικότητα πρέπει $\rho = \lambda/\mu < 1$ και $\text{E}(L) > 200/9 = 22.222$ bits

ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ (Queuing Systems)

Ασκήσεις – Παραδείγματα

Εφαρμογής Συστημάτων Αναμονής M/M/1, M/M/1/K, M/M/m (Erlang-C), M/M/N/K, M/M/m/m (Erlang-B)

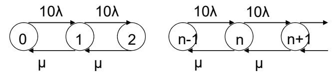
Εβδομάδα 4ης Μαΐου και 11ης Μαΐου, 2020

Παράδειγμα 1 : Πιθανότητες και εξισώσεις καταστάσεων ισορροπίας

- 10 τερματικά τροφοδοτούν κοινό στατιστικό πολυπλέκτη πακέτου (μεταγωγέα – switch ή δρομολογητή – router) που εξυπηρετεί δεδομένα σε πακέτα των 1000 bits κατά μέσο όρο. Η έξοδος του πολυπλέκτη είναι γραμμή των 10 Mbps (Megabits per sec). Τα τερματικά θεωρούνται ανεξάρτητα και ισότιμα.
- A) Προσεγγίστε τον πολυπλέκτη σαν ουρά M/M/1. Βρείτε το μέσο όρο ροής των δεδομένων ανά τερματικό ώστε η γραμμή να έχει χρησιμοποίηση 50%.
- B) Αν ο πολυπλέκτης δεν δύναται να αποθηκεύει πάνω από 3 πακέτα (μαζί με το πακέτο υπό εξυπηρέτηση) και ο μέσος ρυθμός ροής πακέτων ανά τερματικό είναι 500 packets/sec, βρείτε τα χαρακτηριστικά της ουράς. Υποθέστε Poisson διαδικασία άφιξης πακέτων και εκθετικά κατανεμημένους χρόνους εξυπηρέτησης πακέτων.

Λύση – Τμήμα Α

Χρησιμοποιείται μοντέλο M/M/1
Η ροή πακέτων ανά τερματικό είναι λ.
Ζητούμενο: λ=? pak/sec,
Ο ρυθμός εξυπηρέτησης είναι: μ=(10000 kbits/sec)/(1 kbits/pkt) = 10000 pkts/sec

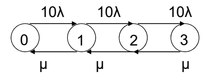


Η αθροιστική ροή πακέτων (από όλα τα τερματικά) στον πολυπλέκτη είναι: 10λ.

Ο βαθμός χρησιμοποίησης είναι: u=(10λ)/μ=0.5 → (10λ)/10000=0.5
→ λ=500 pkts/sec

Λύση – Τμήμα Β

Χρησιμοποιείται μοντέλο M/M/1/3
λ=500 pkts/sec, μ=10000 pkts/sec



$10\lambda P_0=\mu P_1$ → $P_1=(10\lambda/\mu)P_0$ → $P_1=0.5P0$
$10\lambda P_1=\mu P_2$ → $P_2=(10\lambda/\mu)P_0^2$ → $P_2=0.25P0$
$10\lambda P_2=\mu P_3$ → $P_3=(10\lambda/\mu)P_0^3$ → $P_3=0.125P0$

$P_0+P_1+P_2+P_3=1$ → $P_0=8/15$, $P_1=4/15$, $P_2=2/15$, $P_3=1/15$

Μέσο μήκος ουράς: E(n)=0*(8/15)+1*(4/15)+2*(2/15)+3*(1/15)=11/15 pkts

Πιθανότητα απωλειών: $P_{bl}=P_3=1/15$

Ρυθμαπόδοση: $\gamma=10\lambda(1-P_{bl})=500*(14/15)=1400/3$ pkts/sec
($\gamma=\mu(1-P_0)$)

Μέση Καθυστέρηση: E(τ)=E(n)/γ=(11/15)/(1400/3)=11/7000 sec

Παράδειγμα 2: Μοντελοποίηση - Σύγκριση

Δύο υπολογιστές επικοινωνούν με μια γραμμή 64 kbps (kbits/sec) και υποστηρίζει 8 συνόδους (sessions). Αν το μέσο μήκος πακέτου είναι 150 bytes, ο ρυθμός άφιξης ανά σύνοδο (arrival rate/session) είναι 4 packets/second και ακολουθεί Poisson κατανομή, και ο χρόνος εξυπηρέτησης πακέτου είναι εκθετικά κατανεμημένος:

Είναι καλύτερα το δίκτυο να παρέχει σε κάθε σύνοδο το δικό της αφιερωμένο (dedicated, αποκλειστική πρόσβαση) 8 kbps κανάλι, ή είναι προτιμότερο όλες οι σύνοδοι να μοιράζονται όλη τη χωρητικότητα της γραμμής;

Θεωρείστε ότι ο χρόνος καθυστέρησης του πακέτου είναι το πιο σημαντικό κριτήριο.

Λύση

Ας θεωρήσουμε πρώτα το δίκτυο να παρέχει σε κάθε σύνοδο το δικό της (αποκλειστική πρόσβαση) dedicated 8 kbits/sec κανάλι. Τότε κάθε υποσύστημα μπορεί να μοντελοποιηθεί σαν ένα ξεχωριστό M/M/1 σύστημα με λ=4 packets/sec και ρυθμό εξυπηρέτησης 8 kbits/sec ή ισοδύναμα

$$\mu=\frac{8\times10^3}{150\times8}\frac{packets}{\sec}\Rightarrow\mu=6.67\,packets/\sec \qquad T=\frac{1}{\mu-\lambda}=0.375\sec onds=375m\sec$$

Θεωρώντας την περίπτωση όπου οι σύνοδοι μοιράζονται όλη τη χωρητικότητα της γραμμής τότε συγχωνεύουμε όλες τις συνόδους και μοντελοποιούμε το σύστημα σαν M/M/1 σύστημα: με λ=8*4=32 packets/second και ρυθμό εξυπηρέτησης 64 kbits/sec ή ισοδύναμα

$$\mu=\frac{64\times10^3}{150\times8}=53.33\,packets/\sec \qquad T=\frac{1}{\mu-\lambda}=0.0468\sec onds=46.8m\sec$$

Προτιμότερη είναι η δεύτερη λύση αφού μειώνει την καθυστέρηση σημαντικά

Παράδειγμα 3: Μοντελοποίηση Τηλεφωνικής Ζεύξης

Μια τηλεφωνική εταιρεία εγκαθιστά σύνδεση μεταξύ δύο πόλεων όπου η αναμενόμενη κίνηση ακολουθεί κατανομή Poisson με ρυθμό 30 calls/min. Η διάρκεια των κλήσεων είναι ανεξάρτητες εκθετικά κατανεμημένες τυχαίες μεταβλητές με μέση τιμή 3 min. Πόσα κυκλώματα θα πρέπει να παρέχει η εταιρεία ώστε να εγγυηθεί ότι η πιθανότητα απόρριψης κλήσης (blocking) (επειδή όλα τα κυκλώματα είναι κατειλημμένα) είναι μικρότερη του 1%

Λύση

Θεωρούμε ένα σύστημα M/M/m/m όπου m είναι ο αριθμός κυκλωμάτων (γραμμών) που παρέχει η εταιρεία. Πρέπει να βρούμε τον μικρότερο αριθμό m για τον οποίο p_m<0.01 όπου p_m

$$p_m=\frac{(\lambda/\mu)^m/m!}{\sum_{n=0}^{m}(\lambda/\mu)^n/n!}$$

Έχουμε λ = 30 calls/min, $1/\mu$ = 3 min, άρα
$\rho=\lambda/\mu=30\cdot3$ = 90 Erlangs.
Με αντικατάσταση στην παραπάνω σχέση υπολογίζουμε την τιμή του m.

Παράδειγμα 4: (a) Αξιολόγηση Δρομολογητή

Ένας δρομολογητής πακέτων μπορεί να επεξεργάζεται πακέτα με μέσο ρυθμό 300 pkts/sec (packets per second). Τα πακέτα καταφθάνουν στον δρομολογητή κατά μέσο όρο με ρυθμό 200 pkts/sec. Αν μοντελοποιήσουμε το σύστημα σαν M/M/1

- Μέσος # πακέτων στο σύστημα:
  N = λ/(μ-λ) = 200/(300-200) = 2 πακέτα
- Μέσος # πακέτων σε αναμονή:
  $N_Q=\lambda^2/[\mu(\mu-\lambda)]$ = 200^2/([300(300-200)] = 1.33 πακέτα
- Μέσος χρόνος πακέτου στο σύστημα:
  T = 1/(μ-λ) = 1/(300-200) = 0.01 sec = N/λ (Νόμος Little)
- Μέσος χρόνος αναμονής (στην ουρά):
  W = λ/[μ(μ-λ)] = 2/300 = 0.00666 sec = N_Q/λ (ΝόμοςLittle)
- Πιθανότητα να είναι ο εξυπηρετητής απασχολημένος
  ρ = λ/μ = 2/3
- Πιθανότητα το σύστημα να είναι άδειο
  $P_0=1-\rho$ = 0.333

Παράδειγμα 4: (b) Βελτίωση Επίδοσης

Για να βελτιώσουμε την απόδοση του συστήματος έχουμε δύο επιλογές:
- Να εγκαταστήσουμε ένα γρηγορότερο επεξεργαστή (αντικαθιστώντας τον παλαιό)
- Να εγκαταστήσουμε και έναν δεύτερο εξυπηρετητή

1η Επιλογή
Αντικατάσταση του υπάρχοντος εξυπηρετητή με άλλον γρηγορότερο με ρυθμό μ = 400 pkts/second
Επαναπροσδιορισμός της απόδοσης του συστήματος M/M/1 με λ = 200 pkst/sec, μ = 400 pkts/sec και άπειρο μήκος ουράς
2η Επιλογή
Έχουμε ένα σύστημα με δύο εξυπηρετητές, ο κάθε ένας με ρυθμό μ = 200 pkts/sec. Η άφιξη των πακέτων γίνεται σε μία ουρά (buffer) με ρυθμό λ = 200 pkts/second και μετά δρομολογούνται στον πρώτο εξυπηρετητή που είναι ελεύθερος. Το σύστημα αυτό το μοντελοποιούμε σαν M/M/2 με άπειρο μήκος ουράς

ΑΣΚΗΣΗ: Ποίες είναι οι επιδόσεις (μέση καθυστέρηση) των δύο εναλλακτικών; Γιατί;

Παράδειγμα 5: Μοντελοποίηση-διάγραμμα καταστάσεων

Μηνύματα παραδίδονται σε ένα σύστημα αναμονής που αποτελείται από δύο εξυπηρετητές και κοινό χώρο αναμονής. Η διαδικασία άφιξης των μηνυμάτων είναι Poisson (λ=1 πελάτης/sec) και οι χρόνοι εξυπηρέτησης μηνυμάτων είναι εκθετικά κατανεμημένοι. Για τον πρώτο εξυπηρετητή ο ρυθμός εξυπηρέτησης είναι: μ_α και για το δεύτερο είναι: μ_β.

Κάθε καινούριο μήνυμα εξυπηρετείται πάντα από τον πρώτο εξυπηρετητή, αν αυτός είναι ελεύθερος. Αν ο πρώτος εξυπηρετητής είναι απασχολημένος τότε το μήνυμα εξυπηρετείται από τον δεύτερο εξυπηρετητή. Αν και οι δύο εξυπηρετητές είναι απασχολημένοι το μήνυμα αποθηκεύεται στην ουρά.

Βρείτε τις εργοδικές πιθανότητες καταστάσεως και τους βαθμούς χρησιμοποίησης των δυο εξυπηρετητών.

ΑΣΚΗΣΗ 1

- Θεωρήστε ένα σύστημα παρόμοιο με ένα M/M/1 με τη διαφορά ότι όταν το σύστημα αδειάζει η εξυπηρέτηση των πελατών αρχίζει όταν *k* πελάτες είναι παρόντες στο σύστημα (*k* γνωστό). Όταν η εξυπηρέτηση ξεκινήσει συνεχίζει κανονικά μέχρι το σύστημα να αδειάσει ξανά.
- Α) Σχεδιάστε το διάγραμμα καταστάσεων του συστήματος.
- Β) Βρείτε τις εργοδικές πιθανότητες καταστάσεων του αριθμού πελατών στο σύστημα
- Γ) Βρείτε το μέσο αριθμό πελατών στο σύστημα και τη μέση καθυστέρηση ανά πελάτη.

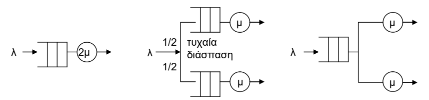
ΑΣΚΗΣΗ 2

Α) Δίνονται K ανεξάρτητες τυχαίες μεταβλητές $X_1, X_2, \ldots, X_k$, καθεμιά εκθετικά κατανεμημένη με παράμετρο λ. Να βρεθεί η συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) της τυχαίας μεταβλητής $\min(X_1, X_2, \ldots, X_k)$.

Β) Μια αθλητική εγκατάσταση έχει 5 γήπεδα τένις. Αθλητές φθάνουν στα γήπεδα με ρυθμό Poisson ενός ζευγαριού ανά 10min. και κάθε ζευγάρι χρησιμοποιεί το γήπεδο κατά ενα χρονικό διάστημα εκθετικά κατανεμημένο με μέση τιμή 40min. Υποθέστε ότι ένα ζευγάρι αθλητών φθάνει στην εγκατάσταση και βρίσκει όλα τα γήπεδα απασχολημένα και k ακόμα ζευγάρια περιμένουν στην αναμονή. Πόσο πρέπει να περιμένει κατά μέσο όρο το ζευγάρι που μόλις έφθασε για να παίξει σε ένα γήπεδο;

ΑΣΚΗΣΗ 3

- Για κάθε ένα από τα συστήματα του σχήματος, υπολογίστε το μέσο χρόνο πακέτου στο σύστημα. Συγκρίνετε και επιλέξτε το καλύτερο και το χειρότερο.
-



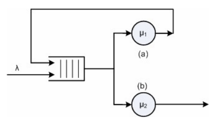
Α) M/M/1 Β) Γ) M/M/2

ΑΣΚΗΣΗ 4

Θεωρείστε το σύστημα δύο εξυπηρετητών του παρακάτω σχήματος. Όταν και οι δύο εξυπηρετητές είναι ανενεργοί, ένα εισερχόμενο πακέτο δρομολογείται πάντα στον δεύτερο εξυπηρετητή (b). Ένας δρομολογητής δεν μπορεί να είναι ανενεργός αν υπάρχει πακέτο στην ουρά αναμονής. Αναχωρήσεις από τον εξυπηρετητή (a), παραμένουν στο σύστημα. Πακέτα που ολοκληρώνουν την εξυπηρέτησή τους στον (b), φεύγουν από το σύστημα.

Α) Σχεδιάστε το διάγραμμα καταστάσεων του συστήματος.

Β) Βρείτε τις εργοδικές πιθανότητες καταστάσεων συναρτήσει μόνο της πιθανότητας άδειου συστήματος και των $\rho_1 = \lambda/\mu_1$, $\rho_2 = \lambda/\mu_2$, $w = \mu_2/\mu_1$.



ΑΣΚΗΣΗ 5

- Θεωρήστε ένα σύστημα αναμονής M/M/2/10 με 2 εξυπηρετητές και μέγιστο αριθμό πελατών 10 (συμπεριλαμβανομένων αυτών που εξυπηρετούνται).

  Εφόσον ο αριθμός των πελατών στο σύστημα είναι μικρότερος ή ίσος του k=4, οι αφίξεις δρομολογούνται πάντα στον εξυπηρετητή a ο δε b παραμένει ανενεργός (idle).

  Ο εξυπηρετητής b ενεργοποιείται μόνο όταν ο αριθμός των πελατών στο σύστημα ξεπεράσει το κατώφλι k=4.

  Σχεδιάστε το διάγραμμα καταστάσεων του συστήματος (Θεωρήστε ότι ο ρυθμός άφιξης πελατών στο σύστημα είναι λ, ο ρυθμός εξυπηρέτησης του εξυπηρετητή a είναι μ_a και ο ρυθμός εξυπηρέτησης του εξυπηρετητή b είναι μ_b).

ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ (Queuing Systems)

Δίκτυα Ουρών

- Ουρές M/M/1 εν σειρά
- Θεώρημα Burke
  Ανοικτά Δίκτυα Ουρών Markov
- Θεώρημα Jackson
- Εφαρμογή σε Δίκτυα Μεταγωγής Πακέτου
- Ασκήσεις/Παραδείγματα

Εβδομάδα 25ης Μαΐου και 1ης Ιουνίου, 2020

ΔΙΚΤΥΟ ΔΥΟ ΕΚΘΕΤΙΚΩΝ ΟΥΡΩΝ ΕΝ ΣΕΙΡΑ (1/2)

- ***Θεώρημα Burke***: Η έξοδος πελατών από ουρά *M/M/1* ακολουθεί κατανομή Poisson με ρυθμό τον ρυθμό εισόδου λ

ΑΠΟΔΕΙΞΗ

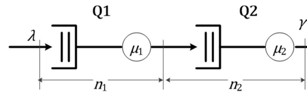
- Θεωρούμε δύο εκθετικές ουρές $\mathbf{Q1}$, $\mathbf{Q2}$ (π.χ. μεταγωγείς πακέτου) με χρόνους εξυπηρέτησης **ανεξάρτητες** εκθετικές μεταβλητές με μέσους όρους $1/\mu_1, 1/\mu_2$
- Προσέγγιση με **Παραδοχή Ανεξαρτησίας** *Leonard Kleinrock* **σε δίκτυα μεταγωγής πακέτου**: Οι χρόνοι εξυπηρέτησης (ανάλογοι του μήκους πακέτου) δεν διατηρούν τα μεγέθη τους όταν προωθούνται μεταξύ συστημάτων (ουρών) εξυπηρέτησης . Οι χρόνοι εξυπηρέτησης ανατίθενται σε κάθε σύστημα σαν **ανεξάρτητε**ς εκθετικές τυχαίες μεταβλητές
- Η είσοδος στην $\mathbf{Q1}$ είναι Poisson με ρυθμό λ (η $\mathbf{Q1}$ είναι ***M/M/1***), $\lambda < \{\mu_1, \mu_2\}$ για εργοδικότητα (ισορροπία)
- Η κατάσταση του συστήματος περιγράφεται από το διάνυσμα $\boldsymbol{n} = (n_1, n_2)$ όπου
  n_1# πελατών στην $\mathbf{Q1}$, n_2 # πελατών στην $\mathbf{Q2}$
- Καταστρώνουμε το διάγραμμα μεταβάσεων καταστάσεων Markov σε δύο διαστάσεις και γράφουμε τις εξισώσεις ισορροπίας
- Εξετάζουμε αν οι εργοδικές πιθανότητες έχουν **μορφή γινομένου** (***product form solution***)
  $P(\boldsymbol{n}) = P(n_1, n_2) \,?= P(n_1)P(n_2)? = (1-\rho_1){\rho_1}^{n_1}(1-\rho_2){\rho_2}^{n_2} = K{\rho_1}^{n_1}{\rho_2}^{n_2}$
  όπου $\rho_1 = \lambda/\mu_1$, $\rho_2 = \lambda/\mu_2$ και $K = (1-\rho_1)(1-\rho_2)$ η ***Σταθερά Κανονικοποίησης***:
  $$\sum_{\boldsymbol{n}} P(\boldsymbol{n}) = 1$$
- Οι εξισώσεις επαληθεύονται.
- Άρα οι δύο ουρές συμπεριφέρονται σαν **δύο ανεξάρτητες ουρές *M/M/1*** σε ισορροπία με ρυθμούς εισόδου λ και ρυθμούς εξυπηρέτησης μ_1, μ_2

$\boxed{\textit{Έπεται πως ο ρυθμός εξόδου της } \mathbf{Q1} \textit{ (και εισόδου στην } \mathbf{Q2}\textit{) είναι Poisson με ρυθμό } \lambda}$

ΔΙΚΤΥΟ ΔΥΟ ΕΚΘΕΤΙΚΩΝ ΟΥΡΩΝ ΕΝ ΣΕΙΡΑ (2/2)

Επαλήθευση Υπόθεσης Γινομένου

$$P(\boldsymbol{n}) = P(n_1, n_2) = P(n_1)P(n_2) = (1-\rho_1){\rho_1}^{n_1}(1-\rho_2){\rho_2}^{n_2} = K{\rho_1}^{n_1}{\rho_2}^{n_2}$$



Επαλήθευση για Αντιπροσωπευτικές Καταστάσεις:

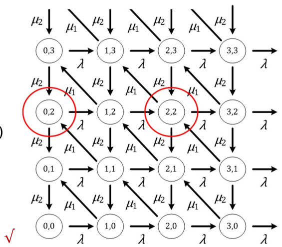
$\boldsymbol{n} = (2,2)$

$(\lambda + \mu_1 + \mu_2)P(2,2)$
$\quad =? \lambda P(1,2) + \mu_1 P(3,1) + \mu_2 P(2,3)$

$(\lambda + \mu_1 + \mu_2)K(\lambda/\mu_1)^2(\lambda/\mu_2)^2$
$=? \lambda K(\lambda/\mu_1)(\lambda/\mu_2)^2 + \mu_1 K(\lambda/\mu_1)^3(\lambda/\mu_2)$
$+ \mu_2 K(\lambda/\mu_1)^2(\lambda/\mu_2)^3 \quad \surd$

$\boldsymbol{n} = (0,2)$

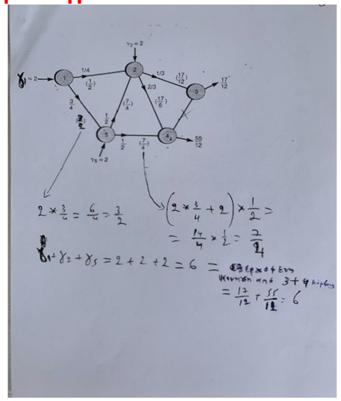
$(\lambda + \mu_2)P(0,2) =? \mu_1 P(1,1) + \mu_2 P(0,3)$
$(\lambda + \mu_2)K(\lambda/\mu_2)^2$
$=? \mu_1 K(\lambda/\mu_1)(\lambda/\mu_2) + \mu_2 K(\lambda/\mu_2)^3 \quad \surd$



ΑΝΟΙΚΤΑ ΔΙΚΤΥΑ ΕΚΘΕΤΙΚΩΝ ΟΥΡΩΝ (1/3)
Παραδοχές για Κατάσταση Δικτύου χωρίς Μνήμη (Markov)

- Έξοδος Ουράς M/M/1 – Θεώρημα ***Burke***
  - Οι αναχωρήσεις πελατών από σύστημα ***M/M/1*** αποτελούν διαδικασία Poisson
- Άθροιση – Διάσπαση διαδικασιών Poisson
  - Άθροιση (aggregation) ανεξαρτήτων ροών Poisson λ_1, λ_2: Poisson με μέσο ρυθμό $\lambda = \lambda_1 + \lambda_2$
  - Τυχαία Διάσπαση (random split, routing) ροής Poisson μέσου ρυθμού λ με πιθανότητες $p,\ q = 1 - p$:
    Παράγει διαδικασίες Poisson με ρυθμούς $p\lambda, (1-p)\lambda$

Παράδειγμα Ανοικτού Δικτύου Ουρών



ΑΝΟΙΚΤΑ ΔΙΚΤΥΑ ΕΚΘΕΤΙΚΩΝ ΟΥΡΩΝ (2/3)
Θεώρημα Jackson

Παραδοχές

- Ανοικτό δίκτυο M ***δικτυακών κόμβων εξυπηρέτησης κορμού*** (ουρών αναμονής) $\mathbf{Q}_i$, $i = 1,2,\ldots,M$ με εκθετικούς ρυθμούς εξυπηρέτησης μ_i
- Αφίξεις πελατών (πακέτων) από εξωτερικές πηγές (sources) άμεσα συνδεμένες στον δικτυακό κόμβο κορμού $\mathbf{Q}_s$ προς εξωτερικούς προορισμούς (destinations) άμεσα συνδεμένους στον δικτυακό κόμβο κορμού $\mathbf{Q}_d$: ***Ανεξάρτητες ροές Poisson*** μέσου ρυθμού γ_{sd} όπου $s, d \in \{1,2,\ldots,M\}$
  Συνολική εξωγενής ροή Poisson σε $\mathbf{Q}_s$: $\gamma_s = \sum_{d=1,d\neq s}^{M} \gamma_{sd}$, $s, d \in \{1,2,\ldots,M\}$
- Εσωτερική ***δρομολόγηση*** (routing) με τυχαίο τρόπο και πιθανότητα δρομολόγησης πελάτη από τον κόμβο κορμού (ουρά) $\mathbf{Q}_i$ στον κόμβο $\mathbf{Q}_j$: r_{ij}
- Τότε τον κόμβο εξυπηρέτησης $\mathbf{Q}_j$ διαπερνούν ροές με συνολικό μέσο ρυθμό

$$\lambda_j = \gamma_j + \sum_{i=1,i\neq j}^{M} r_{ij}\lambda_i\,, \qquad j = 1,2,\ldots,M$$

- Οι χρόνοι εξυπηρετήσεις πελατών όπως διαπερνούν το δίκτυο δεν διατηρούν την τιμή τους (έλλειψη μνήμης) αλλά αποκτούν χρόνο εξυπηρέτησης ανάλογα με την κατανομή του κάθε εξυπηρετητή (***Kleinrock's Independence Assumption***, επαληθευμένη με προσομοιώσεις σε δίκτυα με όχι απλοϊκή τοπολογία)

ΑΝΟΙΚΤΑ ΔΙΚΤΥΑ ΕΚΘΕΤΙΚΩΝ ΟΥΡΩΝ (3/3)
Θεώρημα Jackson

Αποτέλεσμα

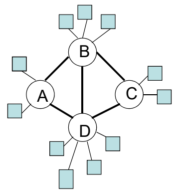
- Κατάσταση του δικτύου: Διάνυσμα $\boldsymbol{n} = (n_1, n_2, \ldots, n_M)$ αριθμού πελατών n_i στις ουρές (κόμβους κορμού) $\mathbf{Q}_i$
- Η ***Εργοδική Πιθανότητα*** των καταστάσεων $\boldsymbol{n}$ (αν υπάρχει) έχει μορφή γινομένου (product form) ***ανεξαρτήτων ουρών M/M/1***

$$P(\boldsymbol{n}) = P(n_1, n_2, \ldots, n_M) = P(n_1)P(n_2)\ldots P(n_M)$$
$$P(n_i) = (1-\rho_i)\rho_i^{\,n_i} \ \text{με}\ \ \rho_i = \lambda_i/\mu_i$$

  όπου λ_i ο συνολικός μέσος ρυθμός (***Poisson***) των πελατών που διαπερνούν τον κόμβο κορμού (ουρά) $\mathbf{Q}_i$ με ρυθμό εκθετικής εξυπηρέτησης μ_i
- Ουρά (κόμβος κορμού) συμφόρησης: Η $\mathbf{Q}_i$ με το μέγιστο ρ_i
- Μέσος αριθμός πελατών (πακέτων) στο δίκτυο: $\mathrm{E}(\boldsymbol{n}) = \sum_{i=1}^{M} \mathrm{E}(n_i) = \sum_{i=1}^{M} \frac{\rho_i}{1-\rho_i}$
- Μέση καθυστέρηση τυχαίου πακέτου από άκρο σε άκρο: $\mathrm{E}(T) = \mathrm{E}(\boldsymbol{n})/\boldsymbol{\gamma}$ (***τύπος Little***)
  όπου $\boldsymbol{\gamma}$ ο συνολικός μέσος ρυθμός πελατών (***Poisson***) που εισέρχονται στο δίκτυο από εξωτερικές πηγές (***network throughput***) $\boldsymbol{\gamma} = \sum_{s=1}^{M}\sum_{d=1,d\neq s}^{M} \gamma_{sd}$
- Μέση καθυστέρηση τυχαίου πακέτου από κόμβο s σε κόμβο d:
  $\mathrm{E}(T_{sd}) = \sum_{i-1}^{M} \delta_{sd}(i)\frac{1/\mu_i}{1-\rho_i}$ όπου $\delta_{sd}(i)$ το κλάσμα της ροής γ_{sd} που διαπερνά τον κόμβο $\mathbf{Q}_i$

ΕΦΑΡΜΟΓΗ: ΔΙΚΤΥΟ ΜΕΤΑΓΩΓΗΣ ΠΑΚΕΤΩΝ (1/2)

- Θεωρήστε ένα δίκτυο μεταγωγής πακέτων.
  - Όλες οι γραμμές (FDX) θεωρούνται χωρητικότητας $\mathrm{C}_i = \mathrm{C} = 10$ Gbits/sec. Το μέσο μήκος του πακέτου είναι $\mathrm{E(L)} = 1000$ bits (θεωρείστε εκθετική κατανομή).
  - Μεταξύ κόμβων θεωρείστε προσφερόμενους ρυθμούς πακέτων Poisson, με ίσους ρυθμούς r packets/sec (από άκρο σε άκρο).
  - Πακέτα από το A στο C και αντίστροφα δρομολογούνται εξίσου στους δύο ισότιμους δρόμους: (A-B-C) και (A-D-C). Τα πακέτα μεταξύ κόμβων κατευθείαν συνδεδεμένων (A-B), (A-D), (B-D), (B-C), (D-C) δρομολογούνται κατευθείαν.
- (A) Βρείτε το ρυθμό r ώστε η γραμμή συμφόρησης (με τη μέγιστη χρησιμοποίηση) να είναι 50%
- (B) Με το r του (A) βρείτε τη μέση καθυστέρηση ενός τυχαίου πακέτου στο δίκτυο (από άκρο σε άκρο)

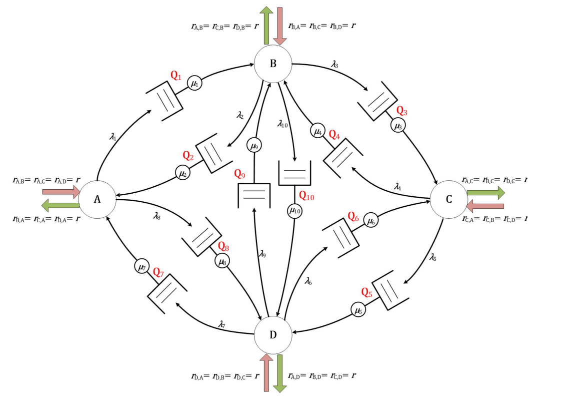


Κόμβος Δικτύου Κορμού
(Δρομολογητής Κορμού, Backbone Router, Packet Switch)

Κόμβος Εισόδου
(Η/Υ, Access Node, Customer Network - LAN)

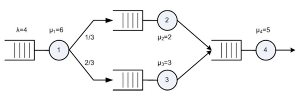
ΟΔΗΓΙΑ: Οι FDX γραμμές του δικτύου κορμού αναλύονται σε δύο ουρές με ροές πακέτων συνολικού μέσου ρυθμού λ_i (προκύπτει από την δρομολόγηση πακέτων) και μέσου ρυθμού εκθετικής εξυπηρέτησης $\mu_i = \mathrm{C}_i/\mathrm{E(L)}$. Το ανοικτό δίκτυο ουρών (**επόμενη διαφάνεια**) αναλύεται σαν **δίκτυο ανεξαρτήτων ουρών M/M/1** με το **Θεώρημα του Jackson**

ΕΦΑΡΜΟΓΗ: ΔΙΚΤΥΟ ΜΕΤΑΓΩΓΗΣ ΠΑΚΕΤΩΝ (2/2)



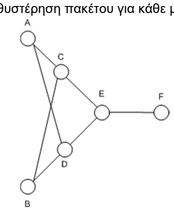
ΑΣΚΗΣΗ 1

- Το παρακάτω σχήμα (δίκτυο ουρών αναμονής) παριστά ένα τηλεπικοινωνιακό δίκτυο. Μια ροή κίνησης έντασης εισέρχεται στον κόμβο 1 και διασπάται τυχαία με πιθανότητα 1/3 προς τον κόμβο 2 και με πιθανότητα 2/3 προς τον κόμβο 3. Βρείτε τις εργοδικές κατανομές πιθανοτήτων του αριθμού πακέτων σε κάθε ουρά αναμονής. Βρείτε το μέσο αριθμό πακέτων σε κάθε ουρά και το μέσο χρόνο συστήματος που ακολουθούν τα πακέτα στις διαδρομές (υπορούς) 1-2-4 και 1-3-4. Κάθε σύνδεση μεταξύ διαδοχικών ουρών αναμονής μπορεί να θεωρηθεί ως μια ουρά *M/M/1*.
- Σε κάθε περίπτωση $\lambda_i < \mu_i, \quad i = 1,2,3,4$



ΑΣΚΗΣΗ 2

- Θεωρήστε το παρακάτω δίκτυο. Υπάρχουν 4 σύνοδοι (ροές πακέτων) ACE, ADE, BCEF, BDEF οι οποίες δημιουργούν κίνηση Poisson με ρυθμούς, 200, 400, 800 και 900 πακέτα ανά δευτερόλεπτο αντίστοιχα. Τα μήκη των πακέτων είναι εκθετικά κατανεμημένα με μέση τιμή 1000bits. Όλες οι γραμμές μετάδοσης έχουν χωρητικότητα 5Mbit/sec. Υποθέστε ότι η κάθε γραμμή μετάδοσης μπορεί να θεωρηθεί ως μια M/M/1 ουρά.
- A) Βρείτε το μέσο αριθμό πακέτων στο σύστημα και τη μέση καθυστέρηση ανά πακέτο (ανεξαρτήτως συνόδου).
- B) Βρείτε τη μέση καθυστέρηση πακέτου για κάθε μία σύνοδο.



ΑΣΚΗΣΗ 3

- Θεωρείστε δύο κανάλια επικοινωνίας, καθένα από τα οποία θα εξυπηρετεί μια ροή πακέτων, όπου όλα τα πακέτα έχουν τον ίδιο σταθερό χρόνο μετάδοσης *T* και τον ίδιο σταθερό χρόνο μεταξύ διαδοχικών αφίξεων, *R*>*T*. Θεωρήστε εναλλακτικά, ότι οι δύο σταθερές ροές συγχωνεύονται με τυχαίο συγχρονισμό έναρξης σε ένα κανάλι διπλής ταχύτητας. Δείξτε ότι ο μέσος χρόνος συστήματος (αναμονή + εξυπηρέτηση) ενός πακέτου θα μειωθεί από *T* σε μια τιμή μεταξύ *T*/2 και 3*T*/4.