

ΤΕΙ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Συστήματα Εξυπηρέτησης
Queueing Systems

Λιώδη Θεοδώρα
Παπανικολόπουλος Αργύρης

Εποπεύων Καθηγητής : Αναστασίου Αθανάσιος

2014

Περιεχόμενα

Περίληψη	6
Abstract	7
Εισαγωγή	8
Κεφάλαιο 1: Βασικές έννοιες από τη θεωρία πιθανοτήτων.....	10
1.1. Τυχαία μεταβλητή.....	10
1.2. Γεννήτρια συνάρτηση (Generating function)	10
1.3. Μετασχηματισμός Laplace-Stieltjes	11
1.4. Κατανομές πιθανότητας.....	12
1.4.1. Γεωμετρική κατανομή	12
1.4.2. Κατανομή Poisson	12
1.4.3. Εκθετική κατανομή.....	12
1.4.4. Κατανομή Erlang	15
1.4.5. Υπερεκθετική κατανομή	16
1.4.6. Κατανομή Cox	17
1.4.7. Άλλες κατανομές	18
1.5. Ταξινόμηση στοχαστικών διαδικασιών	19
1.5.1 Στάσιμες διαδικασίες	19
1.5.2 Ανεξάρτητες διαδικασίες	20
1.5.3 Διαδικασίες Markov(Μαρκοβιανές).....	20
1.5.4 Διαδικασίες γεννήσεων-θανάτων	20
1.5.5 Διαδικασίες Semi-Markov(Ημιμαρκοβιανές).....	21
1.5.6 Τυχαίοι περίπατοι	21
1.5.7 Ανανεωτικές διαδικασίες.....	21
1.5.8. Εργοδικές διαδικασίες	22
1.6. Διαδικασία Poisson	22
Κεφάλαιο 2: Βασικό μοντέλο ουρών αναμονής και βασικές σχέσεις	26

2.1. Μοντέλα ουρών αναμονής.....	26
2.2. Ο συμβολισμός του Kendall	28
2.3. Ρυθμός απασχόλησης.....	29
2.4. Μετρικές απόδοσης	29
2.5. Ο νόμος του Little	30
Κεφάλαιο 3: Μοντέλα ουρών αναμονής.....	32
3.1. Ουρά M/M/1	32
3.1.1. Χρονο-εξαρτημένη συμπεριφορά	32
3.1.2. Περιοριστική συμπεριφορά	32
3.1.3. Μέσες μετρικές απόδοσης	34
3.1.4. Προτεραιότητες.....	34
3.1.5. Απασχολημένη περίοδος (Busy period)	36
3.2. Ουρά M/M/c	36
3.2.1. Πιθανότητες ισορροπίας	37
3.2.2. Μέσο μήκος ουράς και μέσος χρόνος αναμονής.....	38
3.2.3. Ουρές Tandem	39
3.3. Ουρά M/M/∞	39
3.4. Ουρά M/M/1/K	40
3.5. Ουρά M/M/c/c.....	41
3.6. Ουρά M/M/1//M	42
3.7. Ουρά M/Er/1	43
3.7.1. Δυο εναλλακτικές περιγραφές των σταδίων	44
3.7.2. Κατανομή ισορροπίας.....	44
3.7.3. Μέσος χρόνος αναμονής.....	46
3.7.4. Κατανομή του χρόνου αναμονής.....	47
3.8. Ουρά M/G/1	48
3.8.1. Περιοριστική κατανομή	48

3.8.2. Κατανομή αναχωρήσεων	49
3.8.3. Κατανομή του χρόνου παραμονής.....	51
3.8.4. Κατανομή του χρόνου αναμονής	52
3.8.5. Εξίσωση του Lindley	52
3.8.6. Προσέγγιση της μέσης τιμής.....	53
3.8.7. Υπολειπόμενος χρόνος εξυπηρέτησης.....	54
3.8.8. Διακύμανση του χρόνου αναμονής.....	54
3.8.9. Κατανομή της απασχολημένης περιόδου	55
3.9. Ουρά G/M/1	55
3.9.1. Κατανομή αφίξεων	56
3.9.2. Κατανομή του χρόνου παραμονής.....	58
3.9.3. Μέσος χρόνος παραμονής.....	58
Κεφάλαιο 4: Εφαρμογές των συστημάτων εξυπηρέτησης	59
4.1. Χρονοπρογραμματισμός υπολογιστών	60
4.2. Ουρές σε τράπεζα	61
4.2.1. Εισαγωγικά	61
4.2.2. Υποθέσεις και λεπτομέρειες του μοντέλου	61
4.2.3. Σύμβολα	62
4.2.4. Υπολογισμοί	62
4.3. Ουρές που δημιουργούνται κατά την ώρα αιχμής	63
4.3.1. Εισαγωγικά	63
4.3.2. Σύμβολα	63
4.3.3. Μοντελοποίηση της ουράς.....	63
4.4. Γραφείο ταξινόμησης γραμμάτων	65
4.4.1. Εισαγωγικά	65
4.4.2. Υποθέσεις και απλοποιήσεις.....	65
4.4.3. Σύμβολα	66

4.4.4. Τοποθέτηση των εργαζομένων	66
4.4.5. Περίληψη των αποτελεσμάτων.....	67
4.5. Ουρές που δημιουργούνται σε διασταύρωση δρόμων.....	67
4.5.1. Εισαγωγικά	67
4.5.2. Υποθέσεις.....	68
4.5.3. Σύμβολα	68
4.5.4. Μοντελοποίηση της ροής της κίνησης	70
Κεφάλαιο 5: Υλοποίηση τραπεζικού συστήματος εξυπηρέτησης.....	72
5.1. Περιγραφή.....	72
5.2. Κλάσεις	72
5.2.1. BankQueue.....	72
5.2.2. Constants.....	74
5.2.3. ServiceWindow	74
5.2.4. Ticket	77
5.2.5. TicketType	79
Κεφάλαιο 6: Συμπεράσματα	81
Βιβλιογραφία	82
Ελληνική Βιβλιογραφία	82
Ξένη Βιβλιογραφία	82

Περίληψη

Η θεωρία των συστημάτων αναμονής (ή εξυπηρέτησης) είναι ένας τομέας των θετικών επιστημών με πολλές και σημαντικές εφαρμογές σε διάφορες περιοχές των επιστημών αλλά και της καθημερινότητας. Η βάση για τη θεωρία αναμονής είναι η θεωρία πιθανοτήτων, ιδιαίτερα οι κατανομές πιθανότητας και οι διαδικασίες Poisson. Το βασικό μοντέλο ουράς αναμονής πάνω στο οποίο στηρίζονται όλα τα άλλα αποτελείται από έναν πελάτη, έναν εξυπηρετητή, με εκθετικούς χρόνους αφίξεων και εξυπηρέτησεων. Τα πιο σημαντικά μοντέλα ουρών είναι: $M/M/1$, $M/M/c$, $M/M/\infty$, $M/M/1/K$, $M/M/c/c$, $M/M/1//M$, $M/Er/1$, $M/G/1$ και $G/M/1$ σύμφωνα με το συμβολισμό του Kendall, γνωστός για το έργο του σχετικά με τις πιθανότητες, τη στατιστική σχηματική ανάλυση και τη θεωρία αναμονής. Τα κύρια στοιχεία που μελετώνται σε αυτά τα μοντέλα είναι ο ρυθμός απασχόλησης, διάφορες μετρικές απόδοσης όπως ο μέσος χρόνος καθυστέρησης και το μήκος της ουράς που συνδέονται με το νόμο του Little. Αυτός είναι ένας γενικός λειτουργικός νόμος που εφαρμόζεται σε όλα τα μοντέλα συστημάτων ανεξαρτήτως της διαδικασίας αφίξεως των συμβάντων (όπως κλήσεις σε τηλεφωνικό κέντρο, πελατών σε μια ουρά αναμονής κλπ), του μηχανισμού εξυπηρέτησης και του είδους της ουράς αναμονής. Τα συστήματα εξυπηρέτησης βρίσκουν τις κύριες εφαρμογές τους στην επιστήμη των υπολογιστών, στις ουρές τραπεζών ή οποιουδήποτε άλλου οργανισμού, στην κίνηση των δρόμων και την οργάνωση και ταξινόμηση γραμμμάτων ή πακέτων.

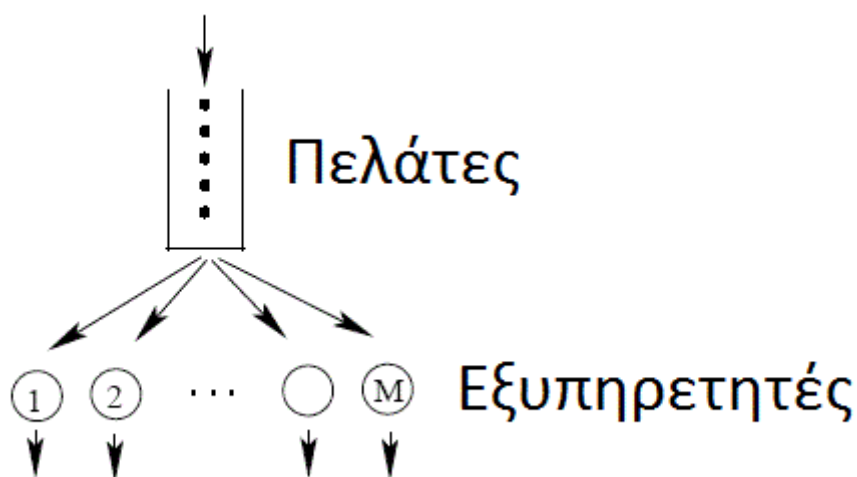
Abstract

Queuing theory is a field of science with various and important applications in a vast number of areas of science and everyday life. The basis for the queuing theory is probability theory, particularly the probability distributions and the Poisson procedures. The basic queue model on which all the other models are based consists of a client, a server with exponential arrivals and service times. The most important models are: $M/M/1$, $M/M/c$, $M/M/\infty$, $M/M/1/K$, $M/M/c/c$, $M/M/1/M$, $M/Er/1$, $M/G/1$ and $G/M/1$ according to Kendall's notation, known for his work on probability, statistical shape analysis and queuing theory. The main elements considered in these models are the throughput, various performance metrics such as average delay time and the length of the queue which are associated with Little's law, a general functional rule which applies to all system models regardless of the arrival process of events (such as calls in a call center, customers in a queue line etc), the service facility and the type of queue. The queuing systems find their main applications in the areas of computer science, at bank queues or any other organization, in the traffic of roads and in organizing and sorting letters or packages.

Εισαγωγή

Αντικείμενο της παρούσας εργασίας είναι η μελέτη των θεμελιωδών συστημάτων εξυπηρέτησης. Τα συστήματα εξυπηρέτησης, ή αλλιώς τα συστήματα αναμονής, βρίσκονται σε πολλούς τομείς της καθημερινότητας, αφού διέπουν όλες τις διαδικασίες που περιλαμβάνουν χρόνο αναμονής. Είναι σημαντικό να χρόνος αυτός να μπορεί να προβλεφθεί και να υπολογιστεί, συνεπώς χρειάζονται μοντέλα και τεχνικές για την ανάλυσή του.

Μελετώνται τα βασικά και πιο διαδεδομένα μοντέλα συστημάτων εξυπηρέτησης. Δίνεται έμφαση στις μεθόδους ανάλυσης αυτών των μοντέλων, καθώς και στις εφαρμογές των μοντέλων αναμονής. Σημαντικές περιοχές εφαρμογής των μοντέλων αναμονής είναι τα συστήματα παραγωγής, οι μεταφορές, τα συστήματα μετοχών, τα τηλεπικοινωνιακά συστήματα και τα συστήματα επεξεργασίας πληροφοριών. Τα συστήματα αναμονής είναι ιδιαίτερα χρήσιμα για το σχεδιασμό των συστημάτων αυτών όσον αφορά τη διάρθρωση, τις χωρητικότητες και τον έλεγχο.



Εικόνα 1 Απλό μοντέλο συστήματος αναμονής

Η οργάνωση της εργασίας έχει ως εξής:

Κεφάλαιο 1 : Παρουσιάζονται κάποιες βασικές έννοιες από τη θεωρία των πιθανοτήτων οι οποίες είναι απαραίτητες για την κατανόηση των ουρών αναμονής. Κυριαρχεί η έννοια της τυχαίας μεταβλητής, των γεννητριών συναρτήσεων και των κατανομών πιθανότητας. Αναφέρεται επίσης ο μετασχηματισμός Laplace και η διαδικασία Poisson.

Κεφάλαιο 2 : Αναλύεται το βασικό μοντέλο ουρών αναμονής, πάνω στο οποίο βασίζονται όλα τα μοντέλα που χρησιμοποιούνται σήμερα. Παρουσιάζονται επίσης κάποιες βασικές έννοιες και συναρτήσεις που εμφανίζονται στα συστήματα εξυπηρέτησης.

Κεφάλαιο 3 : Γίνεται εκτενής αναφορά στα συνηθέστερα μοντέλα ουρών αναμονής, τα πιο σημαντικά εκ των οποίων είναι η ουρά $M/M/1$, $M/M/c$, $M/Et/1$, $M/G/1$ και $G/M/1$.

Κεφάλαιο 4 : Στο κεφάλαιο αυτό παρουσιάζονται οι τομείς στους οποίους βρίσκουν εφαρμογή τα συστήματα εξυπηρέτησης καθώς και κάποια ειδικευμένα παραδείγματα.

Κεφάλαιο 5: Παρουσιάζεται ο κώδικας υλοποίησης ενός τραπεζικού συστήματος εξυπηρέτησης.

Κεφάλαιο 6 : Παρουσιάζονται τα συμπεράσματα που προκύπτουν από την προηγούμενη ανάλυση.

Κεφάλαιο 1: Βασικές έννοιες από τη θεωρία πιθανοτήτων

Στο κεφάλαιο αυτό αναφέρονται όλες οι έννοιες της θεωρίας πιθανοτήτων που χρησιμοποιούνται στα επόμενα κεφάλαια. Καθιερώνεται το υπόβαθρο πάνω στο οποίο θα στηριχθεί η ανάλυση των συστημάτων εξυπηρέτησης.

1.1. Τυχαία μεταβλητή

Οι τυχαίες μεταβλητές συμβολίζονται με κεφαλαία γράμματα όπως X , Y κλπ. Η αναμενόμενη τιμή ή μέσος της X συμβολίζεται με $E(X)$ και η διακύμανση με $\sigma^2(X)$, όπου $\sigma(X)$ είναι η τυπική απόκλιση της μεταβλητής X (Κουτρουβάλη, 1999).

Μια σημαντική ποσότητα είναι συντελεστής διακύμανσης της θετικής τυχαίας μεταβλητής X που ορίζεται ως:

$$c_x = \frac{\sigma(X)}{E(X)} \quad (2.1.1)$$

Ο συντελεστής διακύμανσης είναι ένα μέτρο (χωρίς διαστάσεις) της ποικιλότητας της τυχαίας μεταβλητής X .

1.2. Γεννήτρια συνάρτηση (Generating function)

Έστω ότι X είναι μια μη-αρνητική διακριτή τυχαία μεταβλητή με $P(X=n)=p(n)$, $n=0,1,2,\dots$. Τότε η γεννήτρια συνάρτηση της μεταβλητής X ορίζεται ως (Χρυσ αφίνου, 2004):

$$P_x(z) = E(z^X) = \sum_{n=0}^{\infty} p(n)z^n \quad (2.2.1)$$

Να σημειωθεί ότι ισχύει:

$$|P_x(z)| \leq 1, \text{ για κάθε } |z| \leq 1 \quad (2.2.2)$$

Επιπρόσθετα:

$$P_x(0) = p(0), P_x(1) = 1, P'_x(1) = E(X) \quad (2.2.3)$$

και γενικά:

$$P_x^{(k)}(1) = E(X(X-1)\dots(X-k+1)) \quad (2.2.4)$$

όπου ο όρος (k) δηλώνει την k -οστή παράγωγο.

Για τη γεννήτρια συνάρτηση του αθροίσματος $Z=X+Y$ δύο ανεξάρτητων τυχαίων μεταβλητών X και Y , ισχύει ότι:

$$P_Z(z) = P_X(x) \cdot P_Y(y) \quad (2.2.5)$$

Όταν η μεταβλητή Z έχει πιθανότητα q να ισούται με το X και πιθανότητα $1-q$ να ισούται με το Y , τότε:

$$P_Z(z) = qP_X(x) + (1 - q)P_Y(y) \quad (2.2.6)$$

1.3. Μετασχηματισμός Laplace-Stieltjes

Ο μετασχηματισμός Laplace-Stieltjes μιας μη-αρνητικής τυχαίας μεταβλητής X με συνάρτηση κατανομής F ορίζεται ως:

$$\tilde{X}(s) = E(e^{-sX}) = \int_{x=0}^{\infty} e^{-sx} dF(x), s \geq 0 \quad (2.3.1)$$

Όταν η τυχαία μεταβλητή έχει πυκνότητα f , τότε ο μετασχηματισμός απλοποιείται σε:

$$\tilde{X}(s) = E(e^{-sX}) = \int_{x=0}^{\infty} e^{-sx} f(x) dx, s \geq 0$$

Να σημειωθεί ότι

$$|\tilde{X}(s)| \leq 1, \text{ για κάθε } s \geq 0 \quad (2.3.2)$$

Ακόμα:

$$\tilde{X}(0) = 1, \tilde{X}'(0) = -E(X), \tilde{X}^{(k)}(0) = (-1)^k E(X^k) \quad (2.3.3)$$

Για το μετασχηματισμό του αθροίσματος $Z=X+Y$ δύο ανεξάρτητων τυχαίων μεταβλητών X και Y ισχύει ότι (Κούτρα, 2004):

$$\tilde{Z}(s) = \tilde{X}(s) \cdot \tilde{Y}(s) \quad (2.3.4)$$

Όταν η μεταβλητή Z έχει πιθανότητα q να ισούται με το X και πιθανότητα $1-q$ να ισούται με το Y , τότε:

$$\tilde{Z}(s) = q\tilde{X}(s) + (1 - q)\tilde{Y}(s) \quad (2.3.5)$$

1.4. Κατανομές πιθανότητας

Στην παράγραφο αυτή παρουσιάζονται μερικές από τις πιο γνωστές κατανομές πιθανότητας, οι οποίες βρίσκουν εφαρμογή σε πολλές περιπτώσεις.

1.4.1. Γεωμετρική κατανομή

Μια γεωμετρική τυχαία μεταβλητή X με παράμετρο p έχει κατανομή πιθανότητας (Δάρας, Σύψας, 2003):

$$P(X = n) = (1 - p)p^n, n = 0, 1, 2, \dots \quad (2.4.1)$$

Για αυτή την κατανομή έχουμε:

$$P_X(z) = \frac{1-p}{1-pz}, E(X) = \frac{p}{1-p}, \sigma^2(X) = \frac{p}{(1-p)^2}, c_X^2 = \frac{1}{p} \quad (2.4.2)$$

1.4.2. Κατανομή Poisson

Μια τυχαία μεταβλητή Poisson X με παράμετρο μ έχει πιθανότητα κατανομής (Χρυσσαφίνου, 2004):

$$P(X = n) = \frac{\mu^n}{n!} e^{-\mu}, n = 0, 1, 2, \dots \quad (2.4.3)$$

Για την κατανομή Poisson ισχύει ότι:

$$P_X(z) = e^{-\mu(1-z)}, E(X) = \sigma^2(X) = \mu, c_X^2 = \frac{1}{\mu} \quad (2.4.4)$$

1.4.3. Εκθετική κατανομή

Η πυκνότητα μιας εκθετικής κατανομής με παράμετρο μ δίνεται από τον τύπο (Κουτροβέλη, 1999):

$$f(t) = \mu e^{-\mu t}, t > 0 \quad (2.4.5)$$

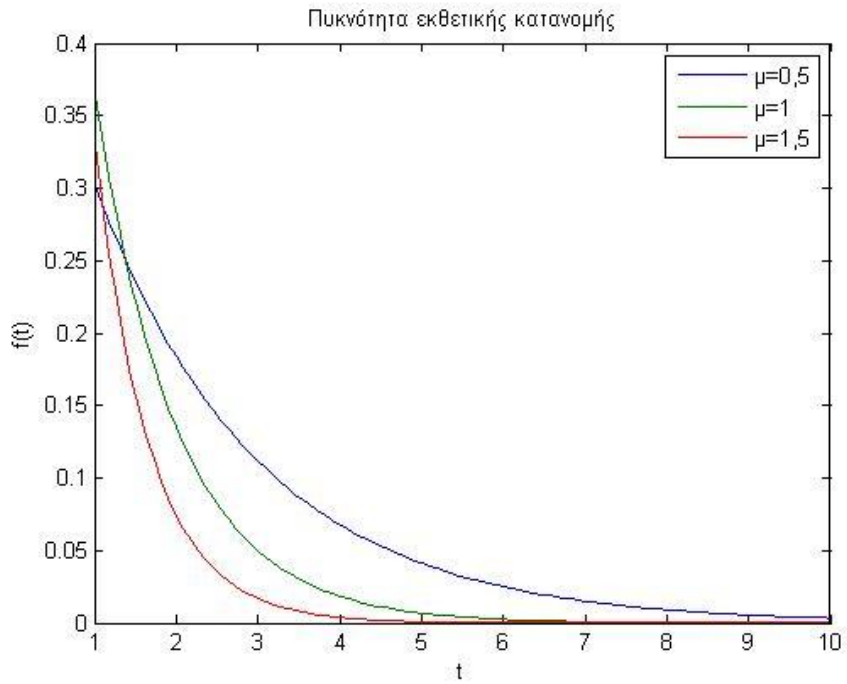
Η συνάρτηση κατανομής είναι:

$$F(t) = 1 - e^{-\mu t}, t \geq 0 \quad (2.4.6)$$

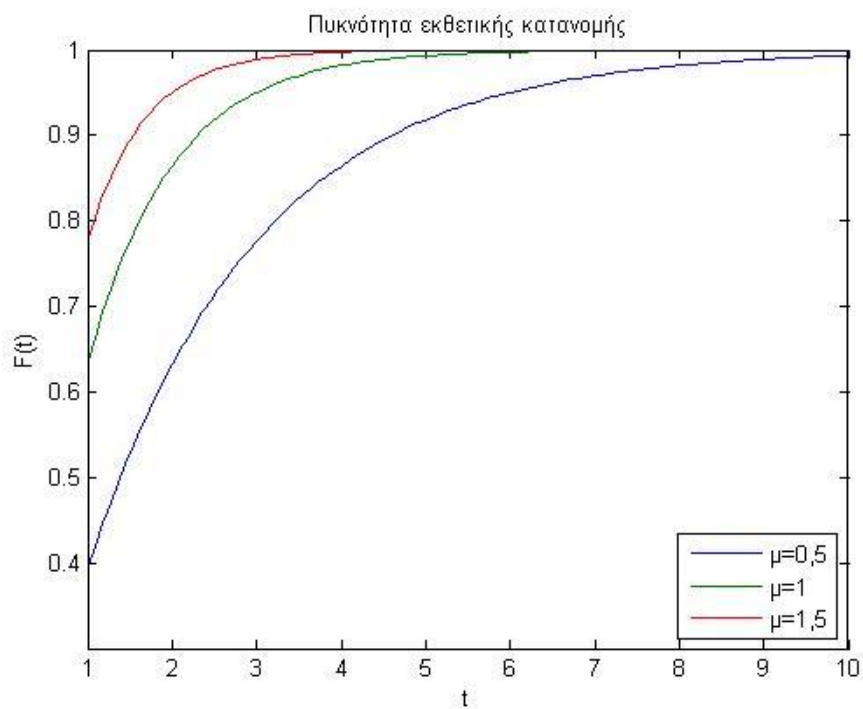
Για αυτή την κατανομή έχουμε:

$$\tilde{X}(s) = \frac{\mu}{\mu + s}, E(X) = \frac{1}{\mu}, \sigma^2(X) = \frac{1}{\mu^2}, c_X = 1 \quad (2.4.7)$$

Παρακάτω φαίνεται η συνάρτηση πυκνότητας πιθανότητας και η συνάρτηση κατανομής πιθανότητας της εκθετικής κατανομής με παράμετρο $\lambda = \mu$.



Εικόνα 2 Συνάρτηση πυκνότητας πιθανότητας εκθετικής κατανομής για τρεις ενδεικτικές τιμές του μ



Εικόνα 3 Συνάρτηση κατανομής πιθανότητας εκθετικής κατανομής για τρεις ενδεικτικές τιμές του μ^1

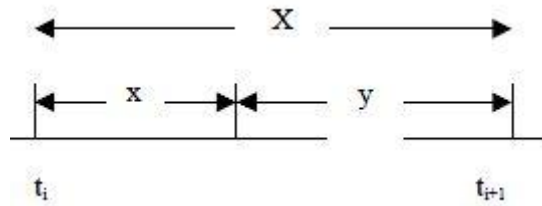
¹ Τα σχήματα δημιουργήθηκαν με το πρόγραμμα Matlab.

Ιδιότητα έλλειψης μνήμης (memory-less property) ή ιδιότητα αμνησίας

Η σπουδαιότερη ιδιότητα της εκθετικής κατανομής είναι η έλλειψη μνήμης (Κούτρας, 2004). Μια τυχαία μεταβλητή λέγεται ότι δεν έχει μνήμη (memoryless), εάν:

$$Pr[X > s + t | X > t] = Pr[X > s], \forall s, t > 0 \quad (2.4.8)$$

Εύκολα αποδεικνύεται ότι αυτό ισχύει για την εκθετική κατανομή. Έστω t_i η στιγμή του i -στού γεγονότος και έστω ότι έχει παρέλθει διάστημα x πριν συμβεί το επόμενο γεγονός, όπως φαίνεται στην παρακάτω εικόνα:



Εικόνα 4 Ιδιότητα έλλειψης μνήμης

Ενδιαφερόμαστε για την πιθανότητα το διάστημα που υπολείπεται μέχρι το επόμενο γεγονός να είναι μεγαλύτερο από y , δεδομένου ότι έχει ήδη παρέλθει διάστημα x από το τελευταίο γεγονός. Αν X ο χρόνος μεταξύ γεγονότων, θα έχουμε σύμφωνα με τον ορισμό της πιθανότητας υπό συνθήκη:

$$\begin{aligned} P(X > x + y | X > x) &= \frac{P(X > x + y, X > x)}{P(X > x)} = \frac{P(X > x + y)}{P(X > x)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda y} \\ &= P(X > y) \end{aligned}$$

δηλαδή η υπό συνθήκη κατανομή του υπολειπόμενου διαστήματος είναι ανεξάρτητη του x και είναι ίδια με την κατανομή του X . Με άλλα λόγια η κατανομή του χρόνου μέχρι το επόμενο γεγονός δεν εξαρτάται από το πότε συνέβη το τελευταίο γεγονός. Αποδεικνύεται ότι η εκθετική κατανομή είναι η μόνη συνεχής κατανομή με την ιδιότητα έλλειψης μνήμης.

Απόδειξη:

Εάν μια τυχαία μεταβλητή X είναι εκθετικά κατανομημένη, τότε:

$$P[X \leq s] = 1 - e^{-\lambda s} \quad \text{ή} \quad P[X > s] = e^{-\lambda s}$$

και

$$P(X > x + t | X > t) = \frac{P(X > x + t, X > t)}{P(X > t)} = \frac{P(X > x + t)}{P(X > t)}$$

Για την απόδειξη της μοναδικότητας, έστω:

$$F(x) = P[X > x]$$

τότε

$$P[X > s + t, X > t] = P[X > t] \cdot P[X > s]$$

ή

$$F(s + t) = F(t) \cdot F(s)$$

Αποδεικνύεται ότι η μόνη (μετρήσιμη) λύση της συναρτησιακής συνάρτησης αυτής είναι η:

$$F(t) = e^{-\lambda t}$$

που είναι η συνάρτηση κατανομής της εκθετικής τυχαίας μεταβλητής.

1.4.4. Κατανομή Erlang

Μια τυχαία μεταβλητή X έχει κατανομή Erlang- k ($k=1,2, \dots$) με μέσο k/μ αν η X είναι το άθροισμα k ανεξάρτητων τυχαίων μεταβλητών X_1, \dots, X_k που έχουν κοινή εκθετική κατανομή με μέσο $1/\mu$. Ο συνήθης συμβολισμός είναι $E_k(\mu)$ ή απλά E_k .

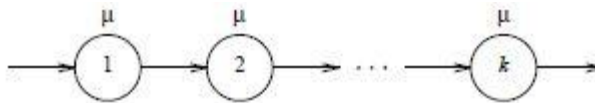
Η πυκνότητα μιας κατανομής E_k δίνεται από:

$$f(t) = \mu \frac{(\mu t)^{k-1}}{(k-1)!} e^{-\mu t}, t > 0 \quad (2.4.9)$$

Η συνάρτηση κατανομής ισούται με:

$$F(t) = 1 - \sum_{j=0}^{k-1} \frac{(\mu t)^j}{j!} e^{-\mu t}, t \geq 0 \quad (2.4.10)$$

Η παράμετρος μ ονομάζεται παράμετρος κλίμακας και η παράμετρος k ονομάζεται παράμετρος σχήματος. Ένα διάγραμμα φάσης μιας κατανομής $E_k(\mu)$ φαίνεται στο παρακάτω σχήμα:



Εικόνα 5 Διάγραμμα φάσης για την κατανομή Erlang- k με παράμετρο κλίμακας μ

Σύμφωνα με την παράσταση αυτή, ένας πελάτης θα πρέπει να διασχίσει k στάδια εξυπηρέτησης, καθένα από τα οποία είναι κατανομημένο εκθετικά με παράμετρο μ . Κάθε στιγμή μόνο ένας πελάτης μπορεί να βρίσκεται σε κάποιο από τα k στάδια της εξυπηρέτησης. Σύμφωνα με την ερμηνεία αυτή, αν έχουμε αφίξεις Poisson (σύστημα $M/Et/1$), η συμπεριφορά του συστήματος μπορεί να περιγραφεί από μια αλυσίδα Markov συνεχούς χρόνου. Η κατάσταση της αλυσίδας Markov θα παριστάνεται από το ζεύγος (n, j) όπου n ο αριθμός πελατών στο σύστημα και j ο αριθμός του σταδίου στο οποίο

βρίσκεται ο πελάτης που εξυπηρετείται. Μια ισοδύναμη περιγραφή θα προέκυπτε, αν θεωρούσαμε ότι κάθε πελάτης που φθάνει στο σύστημα ισοδυναμεί με k στάδια εξυπηρέτησης. Έτσι θα μπορούσαμε αντί για την προηγούμενη διδιάστατη περιγραφή να θεωρήσουμε σαν κατάσταση του συστήματος τον αριθμό σταδίων εξυπηρέτησης που απομένουν κάθε στιγμή

Ο μέσος, η διακύμανση και ο συντελεστής διακύμανσης στο τετράγωνο είναι:

$$E(X) = \frac{k}{\mu}, \sigma^2(X) = \frac{k}{\mu^2}, c_X^2 = \frac{1}{k} \quad (2.4.11)$$

Ο μετασχηματισμός Laplace-Stieltjes είναι:

$$\tilde{X}(s) = \left(\frac{\mu}{\mu + s} \right)^k \quad (2.4.12)$$

1.4.5. Υπερεκθετική κατανομή

Μια τυχαία μεταβλητή X είναι κατανομημένη υπερεκθετικά αν η X έχει πιθανότητα $p_i, i = 1, \dots, k$ και είναι μια εκθετική τυχαία μεταβλητή X_i με μέσο $1/\mu_i$. Για αυτή την τυχαία μεταβλητή χρησιμοποιείται ο συμβολισμός $H_k(p_1, \dots, p_k; \mu_1, \dots, \mu_k)$ ή απλούστερα H_k . Η συνάρτηση πυκνότητας πιθανότητας δίνεται από τη σχέση:

$$f(t) = \sum_{i=1}^k p_i \mu_i e^{-\mu_i t}, t > 0 \quad (2.4.13)$$

και ο μέσος ισούται με

$$E(X) = \sum_{i=1}^k \frac{p_i}{\mu_i} \quad (2.4.14)$$

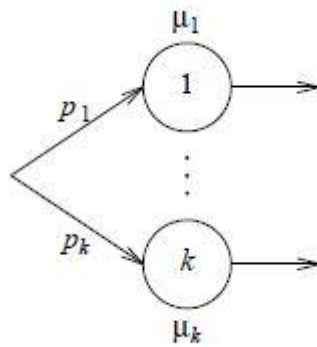
Ο μετασχηματισμός Laplace-Stieltjes είναι:

$$\tilde{X}(s) = \sum_{i=1}^k \frac{p_i \mu_i}{\mu_i + s} \quad (2.4.15)$$

Ο συντελεστής διακύμανσης c_X αυτής της κατανομής είναι πάντα μεγαλύτερος ή ίσος του 1.

Και στην περίπτωση αυτή, ο χρόνος εξυπηρέτησης μπορεί να παρασταθεί με τη βοήθεια k εκθετικών σταδίων, μόνο που εδώ τα στάδια δεν είναι στη σειρά αλλά παράλληλα, και κάθε φορά επιλέγεται κάποιο από τα k στάδια σύμφωνα με τις πιθανότητες p_i . Κάθε στιγμή μόνο ένας πελάτης μπορεί να βρίσκεται σε κάποιο από τα k στάδια.

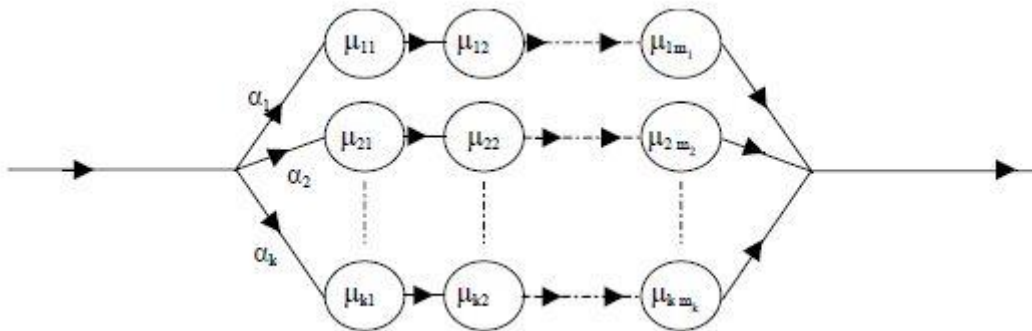
Ένα διάγραμμα φάσης της κατανομής H_k δίνεται στο ακόλουθο σχήμα:



Εικόνα 6 Διάγραμμα φάσης για την υπερκεθτική κατανομή

1.4.6. Κατανομή Cox

Συνδυασμός των προηγούμενων περιπτώσεων μας οδηγεί στην γενικευμένη μέθοδο των σταδίων, σύμφωνα με την οποία η αντίστοιχη κατανομή παριστάνεται από k παράλληλες γραμμές, κάθε μία από τις οποίες περιλαμβάνει ένα αριθμό σταδίων στη σειρά, όπως φαίνεται στο παρακάτω σχήμα (Δάρας, Σύψας, 2003):



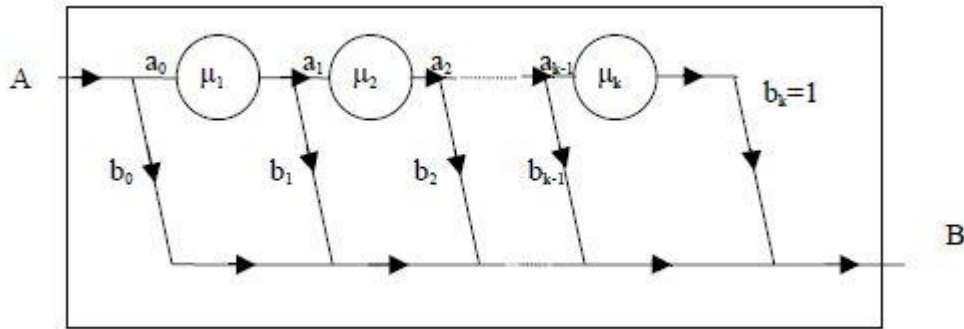
Εικόνα 7 Επιλογή ενός από k στάδια εξυπηρέτησης με m_k στάδια εξυπηρέτησης έκαστο

Ο μετασχηματισμός Laplace για την κατανομή αυτή θα είναι:

$$\Phi(s) = \sum_{i=1}^k p_i \prod_{j=1}^{m_k} \frac{\mu_{ij}}{\mu_{ij} + s} \quad (2.4.16)$$

Οι κατανομές αυτές ανήκουν σε μια γενική κατηγορία κατανομών, οι οποίες είναι γνωστές σαν κατανομές Cox και επιτρέπουν την ανάλυση των αντιστοίχων συστημάτων αναμονής με βάση τη θεωρία των διαδικασιών Markov. Οι κατανομές αυτές έχουν ιδιαίτερη χρησιμότητα στη μελέτη των δικτύων αναμονής.

Θεωρούμε την παράσταση μιας κατανομής Cox με τη βοήθεια k εκθετικών σταδίων, όπως φαίνεται στο παρακάτω σχήμα:



Εικόνα 8 Παράσταση κατανομής Cox

Υποθέτουμε ότι κάθε στιγμή μόνο ένας πελάτης βρίσκεται ανάμεσα στα σημεία A και B. Πριν από κάθε στάδιο $i+1$, $i=0, 1, \dots, k-1$, ο πελάτης επιλέγει αν θα συνεχίσει τη διάσχιση ή θα σταματήσει σύμφωνα με τις αντίστοιχες πιθανότητες a_i και b_i , όπου $a_i+b_i=1$. Μετά το στάδιο k φεύγει με πιθανότητα $b_k=1$.

Ο μετασχηματισμός Laplace της κατανομής Cox δίνεται από τη σχέση:

$$\Phi(s) = b_0 + \sum_{i=1}^k A_i b_i \prod_{j=1}^i \frac{\mu_j}{\mu_j + s} \quad (2.4.17)$$

όπου:

$$A_i = a_0 a_1 \dots a_{i-1} \quad i = 1, 2, \dots, k$$

είναι η πιθανότητα να φθάσει ο πελάτης στο στάδιο i . Το δεξιό μέλος της (2.29) είναι ρητή συνάρτηση του s : μπορεί να γραφεί με τη μορφή $P(s)/Q(s)$ όπου $P(s)$ και $Q(s)$ είναι πολυώνυμα και ο βαθμός του $Q(s)$ είναι μεγαλύτερος ή ίσος από το βαθμό του $P(s)$. Επίσης, όλες οι ρίζες του $Q(s)$ είναι πραγματικές και ισχύει $\Phi(0)=1$. Αντίστροφα, κάθε κατανομή της οποίας ο μετασχηματισμός Laplace ικανοποιεί τις πιο πάνω συνθήκες μπορεί να παρασταθεί με εκθετικά στάδια όπως στην παραπάνω εικόνα. Κάθε γραμμικός συνδυασμός κατανομών Cox είναι επίσης κατανομή Cox. Επιπλέον, οποιαδήποτε κατανομή πιθανότητας μπορεί να προσεγγισθεί με οσοδήποτε ακρίβεια από μια κατανομή Cox.

1.4.7. Άλλες κατανομές

Τελειώνοντας, αναφέρουμε δύο ειδικές περιπτώσεις μαρκοβιανών συστημάτων αναμονής με ιδιαίτερο ενδιαφέρον, η μελέτη των οποίων χρησιμοποιεί αναλυτικές μεθόδους ανάλογες με αυτές που χρησιμοποιούνται στην παράσταση με εκθετικά στάδια. Πρόκειται για συστήματα με *ομαδικές αφίξεις* ή *ομαδική εξυπηρέτηση*.

1.5. Ταξινόμηση στοχαστικών διαδικασιών

Η ταξινόμηση των στοχαστικών διαδικασιών εξαρτάται από τρεις παράγοντες, οι οποίοι είναι (Κουτρουβέλη, 1999):

1. Ο χώρος καταστάσεων, δηλαδή το σύνολο των δυνατών τιμών που μπορεί να πάρουν οι τυχαίες μεταβλητές $X(t)$. Ο χώρος των καταστάσεων είναι διακριτός, αν είναι πεπερασμένος ή απαριθμητός. Μια διαδικασία διακριτών καταστάσεων αναφέρεται συχνά σαν αλυσίδα. Διαφορετικά, ο χώρος καταστάσεων είναι συνεχής, αν αποτελείται από ένα πεπερασμένο ή άπειρο συνεχές διάστημα της ευθείας των πραγματικών αριθμών (ή από ένα σύνολο τέτοιων διαστημάτων).

2. Η παράμετρος του χρόνου (ή δείκτης), η οποία χαρακτηρίζεται από το σύνολο των επιτρεπτών τιμών του χρόνου για τις οποίες ορίζεται η στοχαστική διαδικασία. Όπως και ο χώρος καταστάσεων, το σύνολο αυτό μπορεί να είναι διακριτό ή συνεχές, οπότε αναφερόμαστε σε διαδικασίες διακριτής παραμέτρου ή διαδικασίες συνεχούς παραμέτρου.

3. Οι στατιστικές εξαρτήσεις μεταξύ των τυχαίων μεταβλητών $X(t)$ για διαφορετικές τιμές της παραμέτρου t , οι οποίες περιγράφονται από την από κοινού συνάρτηση κατανομής πιθανότητας $\hat{F}_{\hat{X}}(\hat{x}; \hat{t})$ των τυχαίων μεταβλητών $\hat{X} = (X(t_1), X(t_2), \dots, X(t_n))$ για όλα τα $\hat{x} = (x_1, x_2, \dots, x_n)$ και $\hat{t} = (t_1, t_2, \dots, t_n)$ και όλες τις τιμές του n .

Όπως αναφέρθηκε ήδη, ο καθορισμός αυτός είναι δύσκολος. Θα περιγράψουμε στην συνέχεια μερικούς συνηθισμένους τύπους στοχαστικών διαδικασιών, που χαρακτηρίζονται από διάφορα είδη σχέσεων εξάρτησης μεταξύ των τυχαίων μεταβλητών.

1.5.1 Στάσιμες διαδικασίες

Μια στοχαστική διαδικασία $X(t)$ ονομάζεται *στάσιμη* (*stationary*) όταν όλες οι συναρτήσεις $\hat{F}_{\hat{X}}(\hat{x}; \hat{t})$ μένουν αμετάβλητες σε μετατοπίσεις στο χρόνο, δηλαδή ισχύει:

$$\hat{F}_{\hat{X}}(\hat{x}; \hat{t} + \tau) = \hat{F}_{\hat{X}}(\hat{x}; \hat{t}) \quad (2.5.1)$$

όπου τ σταθερά και $\hat{t} + \tau = (t_1 + \tau, t_2 + \tau, \dots, t_n + \tau)$.

Αν μια διαδικασία είναι στατική θα ισχύει:

$$\bar{X}(t) = \bar{X} \quad (2.5.2)$$

και

$$R_{XX}(t_1, t_2) = R_{XX}(t_2 - t_1) \quad (2.5.3)$$

δηλαδή η μέση τιμή $\bar{X}(t)$ είναι ανεξάρτητη του t και η συνάρτηση αυτοσυσχέτισης $R_{XX}(t_1, t_2)$ εξαρτάται μόνο από τη διαφορά $\tau=t_2-t_1$.

Μια στοχαστική διαδικασία $X(t)$ ονομάζεται *στάσιμη με την ευρεία έννοια* (*wide-sense stationary*) όταν η πρώτη και δεύτερη ροπή αυτής είναι ανεξάρτητες του χρόνου, δηλαδή εάν το $E[X(t)]$ είναι ανεξάρτητο του χρόνου t και το $E[X(t)X(t+\tau)]$ εξαρτάται μονάχα από το τ . Όλες οι στάσιμες στοχαστικές διαδικασίες είναι στάσιμες με την ευρεία έννοια, ενώ το αντίστροφο δεν ισχύει.

1.5.2 Ανεξάρτητες διαδικασίες

Η απλούστερη περίπτωση στοχαστικής διαδικασίας είναι όταν οι τυχαίες μεταβλητές που αντιστοιχούν σε διαφορετικές τιμές της παραμέτρου t είναι ανεξάρτητες μεταξύ τους, οπότε θα ισχύει:

$$f_{\hat{X}}(\hat{x}; \hat{t}) = f_{X_1}(x_1; t_1) \dots f_{X_n}(x_n; t_n) \quad (2.5.4)$$

Μια τέτοια διαδικασία στερείται δομής και αποτελεί ακραία περίπτωση, η οποία για συνεχή παράμετρο αναφέρεται σαν λευκός θόρυβος.

1.5.3 Διαδικασίες Markov(Μαρκοβιανές)

Μια διαδικασία Markov με διακριτό χώρο καταστάσεων ονομάζεται *αλυσίδα Markov*. Ένα σύνολο από τυχαίες μεταβλητές $\{X_n\}$ αποτελούν μια αλυσίδα Markov όταν η πιθανότητα η επόμενη τιμή (κατάσταση) να είναι ίση με x_{n+1} εξαρτάται μονάχα από την παρούσα τιμή (κατάσταση) x_n και όχι από οποιαδήποτε άλλη τιμή του παρελθόντος (Χρυσσαφίνου, 2004).

Η ιδιότητα αυτή είναι γνωστή σαν ιδιότητα αμνησίας και περιορίζει τη γενικότητα των διαδικασιών Markov. Η μελέτη των διαδικασιών αυτών, όμως, είναι βασική για τη θεωρία αναμονής .

1.5.4 Διαδικασίες γεννήσεων-θανάτων

Οι διαδικασίες γεννήσεων-θανάτων (*birth-death processes*) αποτελούν μια σπουδαία κλάση αλυσίδων Markov (διακριτού ή συνεχούς χρόνου). Η ιδιαίτερη συνθήκη η οποία ισχύει για αυτές είναι ότι μεταβάσεις μεταξύ καταστάσεων λαμβάνουν χώρα

μονάχα μεταξύ γειτονικών (*neighboring*) καταστάσεων. Δηλαδή στην περίπτωση αλυσίδας Markov διακριτού χρόνου εάν $X_n=i$, τότε $X_{n+1}=i-1, i$ ή $i+1$.

1.5.5 Διαδικασίες Semi-Markov(Ημιμαρκοβιανές)

Στις διαδικασίες Markov η ιδιότητα έλλειψης μνήμης ισχύει σε οποιαδήποτε χρονική στιγμή. Η συνθήκη αυτή επιβάλλει ότι το χρονικό διάστημα μεταξύ διαδοχικών αλλαγών κατάστασης ακολουθεί μια κατανομή πιθανότητας που εξασφαλίζει την έλλειψη μνήμης. Η κατανομή αυτή πρέπει να είναι η γεωμετρική κατανομή για διακριτό χρόνο ή η εκθετική κατανομή για συνεχή χρόνο.

Αν ο παραπάνω περιορισμός δεν ισχύει (ο χρόνος αυτός δεν ακολουθεί τις παραπάνω κατανομές) οδηγούμαστε σε μια διαδικασία Semi-Markov (ο χρόνος παραμονής σε κάποια κατάσταση μπορεί να ακολουθεί οποιαδήποτε κατανομή). Στις χρονικές στιγμές αλλαγής κατάστασης η διαδικασία συμπεριφέρεται σαν μια κοινή διαδικασία Markov, οπότε αναφερόμαστε στην ενσωματωμένη διαδικασία (ή αλυσίδα) Markov (Embedded Markov chain). Οι διαδικασίες Markov αποτελούν υποσύνολο των διαδικασιών Semi-Markov.

1.5.6 Τυχαίοι περίπατοι

Μια ακολουθία τυχαίων μεταβλητών $\{S_n\}$ ονομάζεται *τυχαίος περίπατος* (*random walk*) αν ισχύει:

$$S_n = X_1 + X_2 + \dots + X_n, n = 1, 2, \dots \quad (2.5.5)$$

Όπου $S_0=0$ και X_1, X_2, \dots είναι μια ακολουθία τυχαίων μεταβλητών ανεξάρτητων και με την ίδια κατανομή πιθανότητας. Ένας τυχαίος περίπατος θα μπορούσε να θεωρηθεί σαν η κίνηση ενός σωματιδίου σε ένα διακριτό χώρο καταστάσεων, όπου κάθε φορά η επόμενη θέση καθορίζεται από την προηγούμενη συν μια τυχαία μεταβλητή. Ο δείκτης n μετρά τον αριθμό των αλλαγών κατάστασης για την στοχαστική διαδικασία. Οι τυχαίοι περίπατοι είναι υποσύνολο των διαδικασιών Semi-Markov.

1.5.7 Ανανεωτικές διαδικασίες

Οι *ανανεωτικές διαδικασίες* (*renewal processes*) μπορούν να θεωρηθούν ειδική περίπτωση των τυχαίων περιπάτων. Οι διαδικασίες αυτές περιγράφουν τον αριθμό των αλλαγών κατάστασης (μεταβάσεις) σαν συνάρτηση του χρόνου. Έτσι, στην εξίσωση:

$$S_n = X_1 + X_2 + \dots + X_n, n = 1, 2, \dots$$

αν οι τυχαίες μεταβλητές $\{X_n\}$ παριστάνουν τους χρόνους μεταξύ μεταβάσεων, τότε η τυχαία μεταβλητή S_n παριστάνει τον χρόνο στον οποίο έγινε η n μετάβαση. Η διαφορά με τους τυχαίους περιπάτους είναι ότι για αυτούς η τυχαία μεταβλητή S_n παριστάνει την κατάσταση της διαδικασίας, ενώ ο χρόνος μεταξύ μεταβάσεων είναι κάποια άλλη τυχαία μεταβλητή.

1.5.8. Εργοδικές διαδικασίες

Η εργοδική θεωρία (ergodic theory) είναι ένας τομέας των μαθηματικών που εξετάζει δυναμικά συστήματα με ένα αμετάβλητο μέτρο. Η αρχική της ανάπτυξη προήλθε από προβλήματα της στατιστικής φυσικής. Το βασικό ενδιαφέρον αυτής της θεωρίας είναι η συμπεριφορά δυναμικών συστημάτων που επιτρέπεται να λειτουργούν για πολύ χρόνο.

Οι εργοδικές διαδικασίες ορίστηκαν στα πλαίσια της εργοδικής θεωρίας. Για αυτές ισχύει ότι ο μέσος χρόνος είναι ο ίδιος σχεδόν όλα τα αρχικά σημεία. Στατιστικά, το σύστημα αυτό που εξελίσσεται για πολύ μεγάλο χρονικό διάστημα «ξεχνά» την αρχική του κατάσταση.

1.6. Διαδικασία Poisson

Έστω $N(t)$ ο αριθμός των αφίξεων στο διάστημα $[0, t]$ για μία διαδικασία Poisson με ρυθμό λ , δηλαδή ο χρόνος ανάμεσα σε διαδοχικές αφίξεις είναι εκθετικά κατανομημένος με παράμετρο λ και ανεξάρτητος του παρελθόντος. Τότε το $N(t)$ έχει κατανομή Poisson με παράμετρο λt έτσι ώστε (Κούτρας, 2004):

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0, 1, 2, \dots \quad (2.6.1)$$

Ο μέσος, η διακύμανση και ο συντελεστής διακύμανσης στο τετράγωνο είναι:

$$E(N(t)) = \lambda t, \sigma^2(N(t)) = \lambda t, c_x^2 = \frac{1}{\lambda t} \quad (2.6.2)$$

Μια συνάρτηση f λέμε ότι είναι $o(h)$ εάν

$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0 \quad (2.6.3)$$

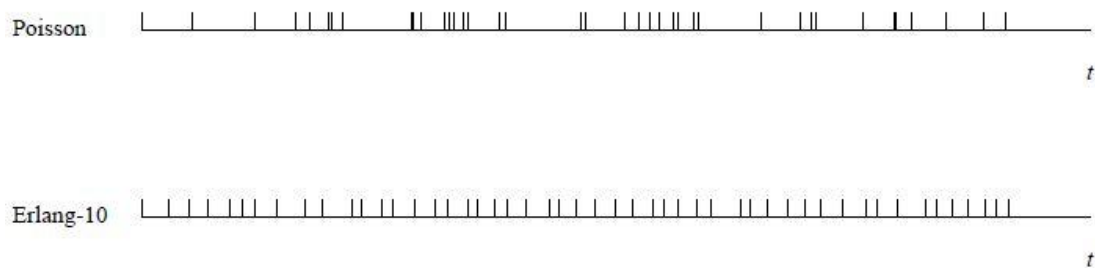
Η πιθανότητα άφιξης σε ένα συγκεκριμένο χρονικό διάστημα δίνεται από τον τύπο:

$$P(\text{arrival in } (t, t + \Delta t]) = \lambda \Delta t + o(\Delta t), (\Delta t \rightarrow 0) \quad (2.5.4)$$

Συνεπώς για μικρό Δt έχουμε:

$$P(\text{arrival in } (t, t + \Delta t]) \approx \lambda \Delta t \quad (2.6.5)$$

Άρα για μικρό χρονικό διάστημα παρεμβολής Δt το γεγονός μιας άφιξης είναι εξίσου πιθανό. Με άλλα λόγια, οι αφίξεις Poisson γίνονται πλήρως τυχαία στο χρόνο. Στο παρακάτω σχήμα φαίνεται μια απεικόνιση μιας διαδικασίας άφιξης Poisson και μιας διαδικασίας άφιξης Erlang-10. Και οι δύο έχουν ρυθμό 1. Η εικόνα δείχνει ότι οι αφίξεις Erlang είναι πολύ περισσότερο ισοκατανεμημένες στο χρόνο από τις αφίξεις Poisson.



Εικόνα 9 Μια απεικόνιση των αφίξεων Poisson και Erlang-10 με ρυθμό 1

Η διαδικασία Poisson είναι μια εξαιρετικά χρήσιμη διαδικασία για μοντελοποίηση πολλών πρακτικών εφαρμογών, όπως το μοντέλο των διεργασιών άφιξης για μοντέλα αναμονής ή διεργασιών ζήτησης για συστήματα απογραφής. Έχει αποδειχθεί εμπειρικά ότι πολλές στοχαστικές διεργασίες μπορούν να προσεγγιστούν ικανοποιητικά με μία διαδικασία Poisson (Χρυσάφινου, 2004).

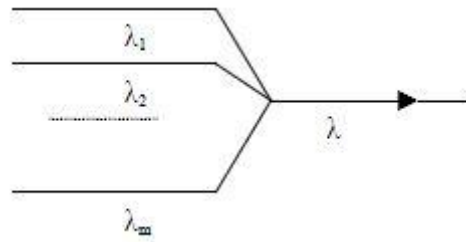
Η διαδικασία Poisson έχει δύο σημαντικές ιδιότητες:

1. Συνένωση

Έστω ότι $N_1(t)$ και $N_2(t)$ είναι δύο ανεξάρτητες διαδικασίες Poisson με ρυθμούς λ_1 και λ_2 αντίστοιχα. Τότε το άθροισμα των δύο διαδικασιών είναι επίσης διαδικασία Poisson με ρυθμό $\lambda_1 + \lambda_2$.

Η ιδιότητα μπορεί να γενικευθεί και για m διαδικασίες. Ας θεωρήσουμε m ανεξάρτητες τυχαίες μεταβλητές εκθετικά κατανεμημένες με παραμέτρους $\lambda_i, i=1,2,\dots,m$ τότε το \min των μεταβλητών αυτών ακολουθεί επίσης εκθετική κατανομή με παράμετρο $\lambda=\lambda_1+\lambda_2+\dots+\lambda_m$. Θεωρούμε την υπέρθεση m ανεξαρτήτων διαδικασιών Poisson με ρυθμούς $\lambda_i, i=1,2,\dots,m$, τότε το διάστημα από μια τυχαία χρονική στιγμή μέχρι το επόμενο γεγονός θα ισοδυναμεί με το \min m ανεξάρτητων τυχαίων μεταβλητών εκθετικά κατανεμημένων με τις αντίστοιχες παραμέτρους λ_i . Συμπεραίνουμε ότι η διαδικασία που

προκύπτει από την υπέρθεση ανεξάρτητων διαδικασιών Poisson είναι επίσης διαδικασία Poisson με ρυθμό το άθροισμα των ρυθμών των επί μέρους διαδικασιών.



Εικόνα 10 Άθροισμα ανεξαρτήτων διαδικασιών Poisson

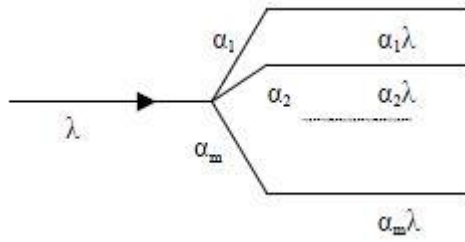
2. Διαχωρισμός

Έστω ότι $N(t)$ είναι μια διαδικασία Poisson με ρυθμό λ και ότι κάθε άφιξη σημειώνεται με πιθανότητα p ανεξάρτητη από όλες τις άλλες αφίξεις. Έστω $N_1(t)$ και $N_2(t)$ ο αριθμός των σημειωμένων και μη αφίξεων αντίστοιχα στο διάστημα $[0,t]$. Τότε οι $N_1(t)$ και $N_2(t)$ είναι και οι δύο διαδικασίες Poisson με ρυθμούς λp και $\lambda(1-p)$ αντίστοιχα. Αυτές οι δύο διαδικασίες είναι ανεξάρτητες.

Η διάσπαση πραγματοποιείται με μια ακολουθία πειραμάτων Bernoulli: κάθε γεγονός της διαδικασίας N ανατίθεται στη διαδικασία N_1 με πιθανότητα a_1 και στην N_2 με πιθανότητα a_2 ($a_1+a_2=1$). Η από κοινού κατανομή πιθανότητας των $N_1(t)$, $N_2(t)$ θα είναι:

$$\begin{aligned} P(N_1(t) = n_1, N_2(t) = n_2) &= P(N_1(t) = n_1, N_2(t) = n_2 \mid N(t) = n_1 + n_2) \cdot P(N(t) = n_1 + n_2) \\ &= \frac{(n_1 + n_2)!}{n_1! n_2!} a_1^{n_1} a_2^{n_2} \frac{(\lambda t)^{n_1+n_2}}{(n_1 + n_2)!} e^{-\lambda t} \\ &= \frac{(a_1 \lambda t)^{n_1}}{n_1!} e^{-a_1 \lambda t} \frac{(a_2 \lambda t)^{n_2}}{n_2!} e^{-a_2 \lambda t} \end{aligned}$$

δηλαδή οι διαδικασίες που προκύπτουν από τη διάσπαση είναι επίσης Poisson με ρυθμούς $a_1 \lambda$ και $a_2 \lambda$ και επιπλέον ανεξάρτητες μεταξύ τους. Το αποτέλεσμα αυτό γενικεύεται εύκολα για διάσπαση σε οποιονδήποτε αριθμό επιμέρους διεργασιών.



Εικόνα 11 Διάσπαση διαδικασίας Poisson με πείραμα Bernoulli

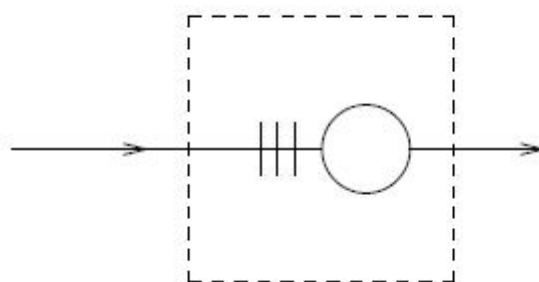
Συμπερασματικά, οι διαδικασίες Poisson παραμένουν διαδικασίες Poisson μετά από συνένωση και διαχωρισμό.

Κεφάλαιο 2: Βασικό μοντέλο ουρών αναμονής και βασικές σχέσεις

Το κεφάλαιο αυτό αναφέρεται στο βασικό μοντέλο ουράς αναμονής, στον συμβολισμό του Kendall και κάποιες βασικές σχέσεις των συστημάτων εξυπηρέτησης.

2.1. Μοντέλα ουρών αναμονής

Το βασικό μοντέλο ουρών αναμονής φαίνεται στο παρακάτω σχήμα. Μπορεί να χρησιμοποιηθεί για μοντελοποίηση μηχανών, φορέων επεξεργασίας παραγγελιών ή εξοπλισμού για επεξεργασία πληροφοριών στις τηλεπικοινωνίες (Ford, 1980).



Εικόνα 12 Βασικό μοντέλο ουρών αναμονής

Μεταξύ άλλων, ένα μοντέλο χαρακτηρίζεται από:

- Το ρυθμό άφιξης των πελατών

Συνήθως υποθέτουμε ότι οι χρόνοι μεταξύ των αφίξεων είναι ανεξάρτητοι και έχουν μια κοινή κατανομή. Σε πολλές πρακτικές περιπτώσεις οι πελάτες φθάνουν στο σύστημα σύμφωνα με μία ροή Poisson, δηλαδή με εκθετικούς χρόνους μεταξύ των αφίξεων. Οι πελάτες μπορεί να φθάνουν ένας ένας ή πολλοί μαζί. Ένα παράδειγμα πολλαπλών αφίξεων είναι ένα τελωνείο σε σύνορα όπου πρέπει να ελεγχθούν τα ταξιδιωτικά έγγραφα των επιβατών ενός λεωφορείου.

- Τη συμπεριφορά των πελατών

Οι πελάτες μπορεί να είναι υπομονετικοί και να θέλουν να περιμένουν, ακόμα και για πολύ χρόνο. Μπορεί όμως να είναι ανυπόμονοι και να φεύγουν μετά από κάποιο χρονικό διάστημα. Για παράδειγμα, σε τηλεφωνικά κέντρα, οι πελάτες θα κατεβάσουν το ακουστικό τους όταν βρίσκονται πολλή ώρα σε αναμονή πριν γίνει διαθέσιμος κάποιος εξυπηρετητής, και πιθανότατα να προσπαθήσουν πάλι αργότερα.

- Τους χρόνους εξυπηρέτησης

Συχνά υποθέτουμε ότι οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι και μοναδικά κατανομημένοι, και ότι είναι ανεξάρτητοι των χρόνων ανάμεσα στις αφίξεις. Για παράδειγμα, ο χρόνος εξυπηρέτησης μπορεί να είναι ντετερμινιστικός ή εκθετικά κατανομημένος. Υπάρχει επίσης η πιθανότητα οι χρόνοι εξυπηρέτησης να εξαρτώνται από το μήκος της ουράς. Για παράδειγμα, ο ρυθμός επεξεργασίας των μηχανών σε ένα σύστημα παραγωγής μπορεί να αυξηθεί όταν ο αριθμός των εργασιών που περιμένουν να εκτελεστούν γίνει πολύ μεγάλος.

- Την πειθαρχία στην υπηρεσία

Οι πελάτες μπορεί να εξυπηρετούνται ένας ένας ή πολλοί μαζί. Υπάρχουν πολλές πιθανότητες για τη σειρά με την οποία ξεκινούν να εξυπηρετούνται. Ενδεικτικά:

- Ο πρώτος που φθάνει εξυπηρετείται (ουρά FCFS), εξαρτάται δηλαδή από τη σειρά άφιξης
- Τυχαία σειρά
- Ο τελευταίος που φθάνει εξυπηρετείται (στοίβα)
- Προτεραιότητες
- Διαμοιρασμός πόρων, όπως σε υπολογιστές που διαμοιράζουν εξίσου την επεξεργαστική δύναμη πάνω σε όλες τις εργασίες του συστήματος

- Την χωρητικότητα της υπηρεσίας

Μπορεί να υπάρχει ένας μοναδικός εξυπηρετητής ή μια ομάδα εξυπηρετητών που βοηθούν τους πελάτες.

- Το χώρο αναμονής

Υπάρχουν περιορισμοί σε ότι αφορά τον αριθμό των πελατών στο σύστημα. Για παράδειγμα, σε δίκτυα επικοινωνίας δεδομένων, μόνο πεπερασμένα κελιά μπορούν να αποθηκευτούν σε μία αλλαγή. Ο καθορισμός του ικανοποιητικού μεγέθους των καταχωρητών είναι ένα σημαντικό ζήτημα στο σχεδιασμό αυτών των δικτύων.

- Τον τρόπο εξυπηρέτησης

Παραθέτονται μερικοί από τους πιο συνηθισμένους τρόπους εξυπηρέτησης:

- *FIFO (First In First Out) ή FCFS (First Come First Served)*: Οι πελάτες εξυπηρετούνται σύμφωνα με την σειρά άφιξής τους.
- *LIFO (Last In First Out) ή LCFS (Last Come First Served)*: Κάθε φορά εξυπηρετείται ο πελάτης με τον πιο πρόσφατο χρόνο άφιξης.

- *FIRO (First In Random Out)*: Ισχύει τυχαία σειρά εξυπηρέτησης των πελατών.
- *Χρονοδρομολόγηση με προτεραιότητες (Priority Scheduling)*: Οι πελάτες χωρίζονται σε κατηγορίες με διαφορετικές προτεραιότητες. Διακρίνουμε δύο γενικούς τύπους προτεραιοτήτων:
 - § *Απλή προτεραιότητα ή προτεραιότητα χωρίς διακοπή (non-preemptive)*: μετά το τέλος εξυπηρέτησης επιλέγεται για την επόμενη εξυπηρέτηση ο πελάτης με την υψηλότερη προτεραιότητα (μεταξύ πελατών με ίση προτεραιότητα ακολουθείται ο κανόνας FCFS).
 - § *Απόλυτη προτεραιότητα ή προτεραιότητα με διακοπή (preemptive)*: όταν ένας πελάτης που φθάνει στο σύστημα βρίσκει ένα πελάτη με χαμηλότερη προτεραιότητα να εξυπηρετείται, το διακόπτει και αρχίζει η δική του εξυπηρέτηση.
 - § *(R-R) Round Robin*: Είναι ένας από τους πιο διαδεδομένους αλγόριθμους χρονοδρομολόγησης για συστήματα καταμερισμού χρόνου (time-sharing). Οι πελάτες εξυπηρετούνται σε διάταξη FCFS εφόσον ο χρόνος εξυπηρέτησής τους δεν ξεπερνά ένα σταθερό χρονικό διάστημα. Όταν ο χρόνος εξυπηρέτησής τους φθάσει το διάστημα αυτό, ο πελάτης διακόπτεται και τοποθετείται στο τέλος της ουράς. Η διαδικασία επαναλαμβάνεται για όλους τους πελάτες.

2.2. Ο συμβολισμός του Kendall

Ο Kendall εισήγαγε ένα συνοπτικό συμβολισμό για να χαρακτηρίσει το εύρος αυτών των μοντέλων. Είναι ένας τριμερής κωδικός $\alpha/\beta/\gamma$. Το πρώτο γράμμα (α) χαρακτηρίζει την κατανομή του χρόνου μεταξύ των αφίξεων και το δεύτερο γράμμα (β) την κατανομή του χρόνου εξυπηρέτησης. Για παράδειγμα, για μία γενική κατανομή χρησιμοποιείται το γράμμα G, για την εκθετική κατανομή χρησιμοποιείται το M (από τη φράση “Memory less”) και το γράμμα D για ντετερμινιστικούς

χρόνους. Το τρίτο γράμμα (γ) προσδιορίζει τον αριθμό των εξυπηρετητών. Μερικά παραδείγματα είναι: $M/M/1$, $M/M/c$, $M/G/1$, $G/M/1$ και $M/D/1$. Ο συμβολισμός μπορεί να επεκταθεί με ένα επιπλέον γράμμα που καλύπτει άλλα μοντέλα ουρών αναμονής. Για παράδειγμα, ένα σύστημα με εκθετικούς χρόνους αφίξεων και εξυπηρέτησης, έναν εξυπηρετητή και χωρητικότητα του χώρου αναμονής N πελάτες (συμπεριλαμβανομένου και του ενός που εξυπηρετείται) συμβολίζεται με τον κωδικό τεσσάρων γραμμάτων $M/M/1/N$ (Ευσταθίου, Τσίτσος, 2011).

Στο βασικό μοντέλο, φθάνει ένας πελάτης τη φορά στο σύστημα και επιτρέπεται πάντα να εισέρχεται σε αυτό, υπάρχει πάντα χώρος, δεν υπάρχουν κανόνες προτεραιότητας και οι πελάτες εξυπηρετούνται με σειρά άφιξης.

2.3. Ρυθμός απασχόλησης

Σε ένα σύστημα με ένα μοναδικό εξυπηρετητή $G/G/1$ με ρυθμό άφιξης λ και μέσο χρόνο εξυπηρέτησης $E(B)$ το ποσό της εργασίας που φθάνει ανά μονάδα χρόνου ισούται με $\lambda * E(B)$. Ο εξυπηρετητής μπορεί να διαχειρίζεται μία μονάδα εργασίας ανά μονάδα χρόνου. Για να αποφευχθεί η αύξηση της ουράς προς το άπειρο, πρέπει να απαιτείται ότι $\lambda * E(B) < 1$. Το μέσο μήκος ουράς απειρίζεται όταν $\lambda * E(B) = 1$, εκτός από τα συστήματα $D/D/1$, δηλαδή αυτά που δεν περιλαμβάνουν καμία τυχειότητα.

Συνηθίζεται να χρησιμοποιείται ο συμβολισμός:

$$\rho = \lambda * E(B) \quad (3.3.1)$$

Αν $\rho < 1$, τότε το ρ ονομάζεται *ρυθμός εξυπηρέτησης* ή *χρησιμοποίηση* του εξυπηρετητή, επειδή είναι το κλάσμα του χρόνου στο οποίο ο εξυπηρετητής εργάζεται.

Σε ένα σύστημα $G/G/c$ με πολλούς εξυπηρετητές πρέπει να απαιτηθεί ότι $\lambda * E(B) < c$. Σε αυτή την περίπτωση ο ρυθμός εξυπηρέτησης είναι:

$$\rho = \frac{\lambda * E(B)}{c} \quad (3.3.2)$$

2.4. Μετρικές απόδοσης

Σχετικές μετρικές απόδοσης στην ανάλυση ουρών αναμονής είναι:

- Η κατανομή του χρόνου αναμονής και ο χρόνος παραμονής του πελάτη. Ο χρόνος παραμονής είναι ο χρόνος αναμονής συν το χρόνο εξυπηρέτησης.
- Η κατανομή του αριθμού των πελατών στο σύστημα, συμπεριλαμβανομένων ή μη αυτών που βρίσκονται στους εξυπηρετητές.
- Η κατανομή του χρόνου εργασίας στο σύστημα. Αυτή είναι το άθροισμα των χρόνων εξυπηρέτησης των πελατών εν αναμονή και ο υπολειπόμενος χρόνος εξυπηρέτησης ενός πελάτη που εξυπηρετείται.
- Η κατανομή των απασχολημένων περιόδων του εξυπηρετητή. Αυτή είναι μια χρονική περίοδος στην οποία ο εξυπηρετητής δουλεύει ασταμάτητα.

Συγκεκριμένα, ενδιαφερόμαστε στις μέσες μετρικές απόδοσης, όπως ο μέσος χρόνος αναμονής και παραμονής.

Ας θεωρήσουμε την ουρά $G/G/c$. Έστω ότι η τυχαία μεταβλητή $L(t)$ δηλώνει τον αριθμό των πελατών στο σύστημα τη χρονική στιγμή t και έστω ότι S_n δηλώνει το χρόνο παραμονής του n -οστού πελάτη στο σύστημα. Με την υπόθεση ότι ο ρυθμός εξυπηρέτησης ανά εξυπηρετητή είναι μικρότερος από 1, μπορεί να αποδειχθεί ότι αυτές οι τυχαίες μεταβλητές έχουν μια περιορισμένη κατανομή καθώς $t \rightarrow \infty$ και $n \rightarrow \infty$. Αυτές οι κατανομές είναι ανεξάρτητες από την αρχική κατάσταση του συστήματος.

Έστω ότι οι τυχαίες μεταβλητές L και S έχουν περιορισμένες κατανομές $L(t)$ και S_n , αντίστοιχα. Τότε:

$$p_k = P(L = k) = \lim_{t \rightarrow \infty} P(L(t) = k), F_S(x) = P(S \leq x) = \lim_{n \rightarrow \infty} P(S_n \leq x) \quad (3.4.1)$$

Η πιθανότητα p_k μπορεί να ερμηνευθεί ως το κλάσμα του χρόνου όταν k πελάτες βρίσκονται στο σύστημα και η $F_S(x)$ δίνει την πιθανότητα ο χρόνος παραμονής ενός μοναδιαίου πελάτη που εισέρχεται στο σύστημα να μην είναι μεγαλύτερη από x μονάδες χρόνου. Ισχύει επίσης με πιθανότητα 1 ότι:

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_{x=0}^t L(x) dx = E(L), \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n S_n = E(S) \quad (3.4.2)$$

Συνεπώς, ο τελικός μέσος αριθμός πελατών στο σύστημα και ο τελικός χρόνος παραμονής είναι ίσοι με $E(L)$ και $E(S)$ αντίστοιχα.

2.5. Ο νόμος του Little

Ο νόμος του Little δίνει μια πολύ σημαντική σχέση ανάμεσα στο $E(L)$, το μέσο αριθμό πελατών στο σύστημα, το $E(S)$, το μέσο χρόνο παραμονής στο σύστημα και

το λ , το μέσο αριθμό πελατών που εισέρχονται στο σύστημα ανά μονάδα χρόνου. Ο νόμος του Little δηλώνει ότι:

$$E(L) = \lambda E(S) \quad (3.5.1)$$

Εδώ γίνεται η υπόθεση ότι η χωρητικότητα του συστήματος επαρκεί για να εξυπηρετηθούν οι πελάτες, δηλαδή ο αριθμός των πελατών στο σύστημα δεν αυξάνεται μέχρι το άπειρο.

Διαισθητικά, αυτό το αποτέλεσμα μπορεί να γίνει κατανοητό ως εξής. Ας υποθέσουμε ότι όλοι οι πελάτες πληρώνουν 1\$ ανά μονάδα χρόνου ενώ είναι στο σύστημα. Αυτά τα χρήματα μπορούν να αποκτηθούν με δύο τρόπους. Η πρώτη πιθανότητα είναι να αφήσουμε τους πελάτες να πληρώνουν «συνεχόμενα» στο χρόνο. Τότε το μέσο κέρδος που αποκτάται από το σύστημα ισούται με $E(L)$ \$ ανά μονάδα χρόνου. Η δεύτερη πιθανότητα είναι να αφήσουμε τους πελάτες να πληρώσουν 1\$ ανά μονάδα χρόνου για τη διαμονή τους στο σύστημα όταν φεύγουν από αυτό. Στην ισορροπία, ο μέσος αριθμός των πελατών που φεύγουν από το σύστημα ανά μονάδα χρόνου είναι ίσος με τον μέσο αριθμό πελατών που εισέρχονται σε αυτό. Έτσι το σύστημα κερδίζει ένα μέσο κέρδος $\lambda E(S)$ \$ ανά μονάδα χρόνου. Προφανώς, το σύστημα κερδίζει το ίδιο και στις δύο περιπτώσεις (Παπαδοπούλου, 2009).

Για να δείξουμε τη χρήση του νόμου του Little θεωρούμε το βασικό μοντέλο της εικόνας 10 με έναν εξυπηρετητή. Για αυτό το μοντέλο μπορούν να παραχθούν σχέσεις μεταξύ διαφόρων μετρικών απόδοσης με την εφαρμογή του νόμου σε κατάλληλα ορισμένα υποσυστήματα. Η εφαρμογή του νόμου στο σύστημα που αποτελείται από ουρά και εξυπηρετητή παράγει τη σχέση (3.5.1). Η εφαρμογή του νόμου στην ουρά (εκτός του εξυπηρετητή) παράγει μια σχέση μεταξύ του μήκους της ουράς L^q και του χρόνου αναμονής W , την:

$$E(L^q) = \lambda E(W) \quad (3.5.2)$$

Τέλος, όταν εφαρμόζεται ο νόμος του Little μόνο στον εξυπηρετητή προκύπτει:

$$\rho = \lambda E(B)$$

όπου ρ είναι ο μέσος αριθμός πελατών στον εξυπηρετητή (που είναι ο ίδιος με το κλάσμα του χρόνου στο οποίο ο εξυπηρετητής είναι απασχολημένος) και $E(B)$ είναι ο μέσος χρόνος εξυπηρέτησης.

Κεφάλαιο 3: Μοντέλα ουρών αναμονής

3.1. Ουρά M/M/1

Σε αυτή την παράγραφο αναλύεται το μοντέλο με εκθετικούς χρόνους άφιξης με μέσο $1/\lambda$, εκθετικούς χρόνους εξυπηρέτησης με μέσο $1/\mu$ με ένα μοναδικό εξυπηρετητή. Οι πελάτες εξυπηρετούνται με βάση τη σειρά άφιξης. Απαιτείται να ισχύει (Adan, 2001):

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

αφού, σε διαφορετική περίπτωση, το μήκος της ουράς θα απειριζόταν όπως δείχθηκε στο κεφάλαιο 2. Η ποσότητα ρ είναι το κλάσμα του χρόνου στο οποίο ο εξυπηρετητής απασχολείται.

3.1.1. Χρονο-εξαρτημένη συμπεριφορά

Η εκθετική κατανομή επιτρέπει μια πολύ απλή επιγραφή της κατάστασης του συστήματος τη χρονική στιγμή t , εξαρτώμενη από τον αριθμό των πελατών στο σύστημα, δηλαδή των πελατών που περιμένουν στην ουρά και αυτού που εξυπηρετείται. Δεν υπάρχει η ανάγκη να θυμόμαστε πότε έφτασε ο τελευταίος πελάτης ούτε πρέπει να γίνεται καταγραφή αυτού του γεγονότος. Αφού η εκθετική κατανομή δεν έχει μνήμη (memory less), αυτή η πληροφορία δεν περιλαμβάνει κάποια πρόβλεψη για το μέλλον (Yue et al, 2009).

Έπεται από υπολογισμούς ότι ισχύει το ακόλουθο σύνολο διαφορικών εξισώσεων για την πιθανότητα $p_n(t)$ να υπάρχουν n πελάτες στο σύστημα τη χρονική στιγμή t :

$$\begin{aligned} p'_0(t) &= -\lambda p_0(t) + \mu p_1(t) & (4.1.1) \\ p'_n(t) &= \lambda p_{n-1}(t) - (\lambda + \mu)p_n(t) + \mu p_{n+1}(t), n = 1, 2, \dots \end{aligned}$$

3.1.2. Περιοριστική συμπεριφορά

Όσο $t \rightarrow \infty$ τότε $p'_n(t) \rightarrow 0$ και $p_n(t) \rightarrow p_n$. Συνεπώς από τις σχέσεις (4.1) προκύπτει ότι οι περιοριστικές πιθανότητες (ή πιθανότητες ισορροπίας) p_n ικανοποιούν τις εξισώσεις (Adan, 2001):

$$0 = -\lambda p_0 + \mu p_1 \quad (4.1.2)$$

$$0 = \lambda p_{n-1} - (\lambda + \mu)p_n + \mu p_{n+1}, n = 1, 2, \dots \quad (4.1.3)$$

Προφανώς, για τις πιθανότητες p_n ισχύει επίσης η σχέση:

$$\sum_{n=0}^{\infty} p_n = 1 \quad (4.1.4)$$

η οποία ονομάζεται σχέση κανονικοποίησης. Μπορούμε επίσης να εξάγουμε τις εξισώσεις (4.4.2) και (4.4.3) απευθείας από ένα διάγραμμα ροής, το οποίο φαίνεται στο παρακάτω σχήμα:



Εικόνα 13 Διάγραμμα ροής για το μοντέλο M/M/1

Τα βέλη δείχνουν τις πιθανές μεταβάσεις. Ο ρυθμός με τον οποίο πραγματοποιείται μία μετάβαση είναι λ για μια άφιξη από το n στο $n+1$ στάδιο και μ για μία αναχώρηση από το $n+1$ στο n στάδιο. Ο αριθμός των μεταβάσεων ανά μονάδα χρόνου από το n στο $n+1$, ο οποίος ονομάζεται επίσης ροή, είναι ίσος με p_n , το κλάσμα του χρόνου για το οποίο το σύστημα είναι στην κατάσταση n , επί λ , τον ρυθμό με τον οποίο πραγματοποιούνται αφίξεις όταν το σύστημα είναι στην κατάσταση n .

Με εφαρμογή στις γενικές εξισώσεις ισορροπίας βρίσκουμε:

$$p_n = p_0 \left(\frac{\lambda}{\mu}\right)^n, n \geq 0 \quad (4.1.5)$$

και

$$p_0 = \left(\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n\right)^{-1} \quad (4.1.6)$$

εφόσον ισχύει η σχέση:

$$S = \left(\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n\right) < \infty \quad (4.1.7)$$

δηλαδή το S είναι φραγμένο.

Με χρήση του ρυθμού απασχόλησης ρ , η πιθανότητα να υπάρχουν n πελάτες στο σύστημα είναι:

$$p_n = (1 - \rho)\rho^n, n \geq 0 \quad (4.1.8)$$

δηλαδή ο αριθμός των πελατών στο σύστημα ακολουθεί γεωμετρική κατανομή.

3.1.3. Μέσες μετρικές απόδοσης

Από τις πιθανότητες ισορροπίας μπορούν να προκύψουν εκφράσεις για το μέσο αριθμό των πελατών στο σύστημα και το μέσο χρόνο που βρίσκονται σε αυτό.

$$E(L) = \sum_{n=0}^{\infty} np_n = \frac{\rho}{1-\rho} \quad (4.1.9)$$

και με εφαρμογή του νόμου του Little:

$$E(S) = \frac{\frac{1}{\mu}}{1-\rho} \quad (4.1.10)$$

Συνδυάζοντας τις παραπάνω σχέσεις προκύπτει η σχέση άφιξης, η οποία είναι:

$$E(S) = E(L) \frac{1}{\mu} + \frac{1}{\mu} \quad (4.1.11)$$

Μαζί με το νόμο του Little προκύπτει η έκφραση (4.1.6). Αυτή η προσέγγιση ονομάζεται *προσέγγιση της μέσης τιμής* (Παπαδοπούλου, 2009).

Ο μέσος αριθμός πελατών στην ουρά, $E(L^q)$ μπορεί να βρεθεί από το $E(L)$ αφαιρώντας το μέσο αριθμό των πελατών στην υπηρεσία, έτσι ώστε:

$$E(L^q) = E(L) - \rho = \frac{\rho^2}{1-\rho} \quad (4.1.12)$$

Ο μέσος χρόνος αναμονής $E(W)$ προκύπτει από το $E(S)$ αφαιρώντας το μέσο χρόνο εξυπηρέτησης έτσι ώστε:

$$E(W) = E(S) - \frac{1}{\mu} = \frac{\frac{\rho}{\mu}}{1-\rho} \quad (4.1.13)$$

3.1.4. Προτεραιότητες

Σε αυτή την παράγραφο θεωρούμε ένα σύστημα M/M/1 με διαφορετικούς τύπους πελατών που χρειάζονται διαφορετική προτεραιότητα εξυπηρέτησης. Ας θεωρήσουμε δύο διαφορετικούς τύπους 1 και 2, χωρίς βλάβη της γενικότητας. Ο τύπος 1 και ο τύπος 2 φθάνει στο σύστημα σύμφωνα με ανεξάρτητες διαδικασίες Poisson με ρυθμό λ_1 και λ_2 αντίστοιχα. Ο χρόνος εξυπηρέτησης όλων των πελατών είναι εκθετικά κατανομημένος με τον ίδιο μέσο $1/\mu$ (Μάγκλαρης, Παπαβασιλείου, 2011). Υποθέτουμε ότι

$$\rho_1 + \rho_2 < 1 \quad (4.1.14)$$

όπου $\rho_i = \frac{\lambda_i}{\mu}$ είναι ο ρυθμός απασχόλησης του τύπου i .

Διακοπή (preemptive) προτεραιότητα

Σε αυτό τον τύπο προτεραιότητας θεωρούμε ότι ο τύπος 1 των πελατών έχει απόλυτη προτεραιότητα έναντι του τύπου 2.

Έστω η τυχαία μεταβλητή L_i που δηλώνει τον αριθμό των πελατών τύπου i στο σύστημα και S_i ο χρόνος παραμονής του τύπου αυτού.

Για τον τύπο 1, ο τύπος 2 δεν υπάρχει. Συνεπώς έπεται ότι:

$$E(S_1) = \frac{\frac{1}{\mu}}{1 - \rho_1}, \quad E(L_1) = \frac{\rho_1}{1 - \rho_1} \quad (4.1.15)$$

Εφόσον οι χρόνοι εξυπηρέτησης όλων των πελατών είναι εκθετικά κατανομημένοι με τον ίδιο μέσο, ο συνολικός αριθμός πελατών στο σύστημα δεν εξαρτάται από τη σειρά με την οποία εξυπηρετούνται οι πελάτες. Έτσι αυτός ο αριθμός είναι ο ίδιος αν στο σύστημα όλοι οι πελάτες εξυπηρετούνταν με βάση τη σειρά άφιξης. Έτσι:

$$E(L_1) + E(L_2) = \frac{\rho_1 + \rho_2}{1 - \rho_1 - \rho_2} \quad (4.1.16)$$

Εισάγοντας τις σχέσεις (4.11) βρίσκουμε για τον τύπο 2 των πελατών:

$$E(L_2) = \frac{\rho_2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} \quad (4.1.17)$$

και εφαρμόζοντας το νόμο του Little έχουμε:

$$E(S_2) = \frac{1/\mu}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} \quad (4.1.18)$$

Μη-διακοπή πιθανότητα

Σε αυτό τον τύπο πιθανότητας, ο τύπος 1 έχει σχεδόν πλήρη προτεραιότητα έναντι του τύπου 2. Η διαφορά με τη διακοπή πιθανότητα είναι ότι σε αυτή την περίπτωση οι πελάτες τύπου 1 δεν μπορούν να διακόψουν την εξυπηρέτηση των πελατών τύπου 2 (Adan, 2001).

Για το μέσο χρόνο παραμονής του τύπου 1 πελατών βρίσκουμε:

$$E(S_1) = E(L_1) \frac{1}{\mu} + \frac{1}{\mu} + \rho_2 \frac{1}{\mu} \quad (4.1.19)$$

Εφαρμόζοντας το νόμο του Little έχουμε:

$$E(S_1) = \frac{1 - \rho_2}{1 - \rho_1} \frac{1}{\mu}, \quad E(L_1) = \frac{(1 + \rho_2)\rho_1}{1 - \rho_1} \quad (4.1.20)$$

Για τον τύπο 2 πελατών, με εφαρμογή της σχέσης (4.4.12) προκύπτει ότι:

$$E(L_2) = \frac{(1 - \rho_1(1 - \rho_1 - \rho_2))\rho_2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} \quad (4.1.21)$$

και με εφαρμογή του νόμου του Little:

$$E(S_2) = \frac{\frac{(1 - \rho_1(1 - \rho_1 - \rho_2))\rho_2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)}}{\mu} \quad (4.1.22)$$

3.1.5. Απασχολημένη περίοδος (Busy period)

Στη ζωή ενός εξυπηρετητή μπορούμε να διακρίνουμε κύκλους. Ένας κύκλος είναι ο χρόνος που βρίσκεται ανάμεσα σε δύο συνεχόμενες αφίξεις που βρίσκουν ένα άδειο σύστημα. Ένας κύκλος ξεκινά με μία απασχολημένη περίοδο BP κατά την οποία ο εξυπηρετητής βοηθά πελάτες, ακολουθούμενη από μία ανενεργή (idle) περίοδο IP στην οποία το σύστημα είναι άδειο (Γεωργίου, 2011).

Η ανενεργή περίοδος είναι εκθετικά κατανομημένη με μέσο $1/\lambda$, λόγω της ιδιότητας έλλειψης μνήμης της εκθετικής κατανομής.

Είναι προφανές ότι η μέση απασχολημένη περίοδος διαιρούμενη με τη μέση διάρκεια κύκλου ισούται με το κλάσμα του χρόνου στο οποίο ο εξυπηρετητής δουλεύει, δηλαδή:

$$\frac{E(BP)}{E(BP) + E(IP)} = \frac{E(BP)}{E(BP) + \frac{1}{\lambda}} = \rho \quad (4.1.23)$$

Άρα:

$$E(BP) = \frac{\frac{1}{\mu}}{1 - \rho} \quad (4.1.24)$$

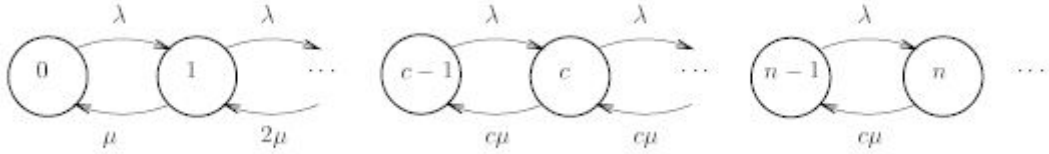
3.2. Ουρά M/M/c

Σε αυτή την παράγραφο αναλύεται το μοντέλο με εκθετικούς χρόνους αφίξεων με μέσο $1/\lambda$, εκθετικούς χρόνους εξυπηρέτησης με μέσο $1/\mu$ και c παράλληλους όμοιους εξυπηρετητές (Adan, 2001). Οι πελάτες εξυπηρετούνται με σειρά αφίξεως. Υποθέτουμε ότι ο ρυθμός απασχόλησης ανά εξυπηρετητή είναι

$$\rho = \frac{\lambda}{c\mu} < 1 \quad (4.2.1)$$

3.2.1. Πιθανότητες ισορροπίας

Η κατάσταση του συστήματος χαρακτηρίζεται πλήρως από τον αριθμό των πελατών στο σύστημα (Γεωργίου, 2011). Έστω p_n η πιθανότητα ισορροπίας ότι n πελάτες βρίσκονται στο σύστημα. Όμοια με το μοντέλο M/M/1 μπορούμε να εξάγουμε τις εξισώσεις ισορροπίας για τις πιθανότητες p_n από το παρακάτω διάγραμμα ροής:



Εικόνα 14 Διάγραμμα ροής για το μοντέλο M/M/c

Αντί να εξισώσουμε τη ροή εντός και εκτός από μία μοναδική κατάσταση n , λαμβάνουμε απλούστερες εξισώσεις με την εξίσωση της ροής εντός και εκτός του συνόλου των καταστάσεων $\{0, 1, \dots, n-1\}$. Αυτό δίνει ότι:

$$\lambda p_{n-1} = \min(n, c) \mu p_n = \begin{cases} n\mu, & n = 1, 2, \dots, c \\ c\mu, & n = c, \dots \end{cases} \quad (4.2.2)$$

Με επανάληψη προκύπτει ότι:

$$p_n = \begin{cases} \frac{(c\rho)^n}{n!} p_0, & n = 0, \dots, c \\ p_0 \frac{c^n \rho^n}{c!}, & n = c + 1, \dots \end{cases} \quad (4.2.3)$$

όπου

$$\rho = \frac{\lambda}{c\mu} < 1$$

Η ποσότητα αυτή εκφράζει το αναμενόμενο ποσοστό απασχολημένων μονάδων εξυπηρέτησης ή ισοδύναμα το βαθμό χρησιμοποίησης κάθε μονάδας εξυπηρέτησης.

Η πιθανότητα p_0 προκύπτει από κανονικοποίηση και είναι:

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!} * \frac{1}{1-\rho} \right)^{-1} \quad (4.2.4)$$

Μια σημαντική ποσότητα είναι η πιθανότητα ότι μια εργασία πρέπει να περιμένει. Συμβολίζουμε αυτή την πιθανότητα P_W και αναφέρεται συχνά ως *πιθανότητα καθυστέρησης*. Προκύπτει ότι είναι:

$$\begin{aligned}
\Pi_W &= p_c + p_{c+1} + p_{c+2} + \dots = \frac{p_c}{1 - \rho} \\
&= \frac{(cp)^c}{c!} \left((1 - \rho) \sum_{n=0}^{c-1} \frac{(cp)^n}{n!} + \frac{(cp)^c}{c!} \right)^{-1} \quad (4.2.5)
\end{aligned}$$

Οι εκφράσεις για τα διάφορα μέτρα επιδόσεων του συστήματος M/M/c είναι σχετικά πολύπλοκες.

Ένα σύστημα M/M/c έχει κατώτερες επιδόσεις από ένα σύστημα M/M/1 με ισοδύναμο ρυθμό εξυπηρέτησης, δηλαδή με ρυθμό εξυπηρέτησης $c\mu$. Αυτό συμβαίνει γιατί στο σύστημα M/M/c δεν χρησιμοποιούνται πάντα όλες οι δυνατότητές του (όταν υπάρχουν λιγότεροι από c πελάτες, κάποιες μονάδες εξυπηρέτησης δεν εργάζονται).

Η έξοδος από ένα σύστημα M/M/c είναι διαδικασία Poisson.

3.2.2. Μέσο μήκος ουράς και μέσος χρόνος αναμονής

Από τις πιθανότητες ισορροπίας προκύπτει άμεσα το μέσο μήκος ουράς (Adan, 2001):

$$E(L^q) = \sum_{n=0}^{\infty} np_{c+n} = \frac{p_c}{1 - \rho} \sum_{n=0}^{\infty} n(1 - \rho)\rho^n = \Pi_W \frac{\rho}{1 - \rho} \quad (4.2.6)$$

και από το νόμο του Little έχουμε:

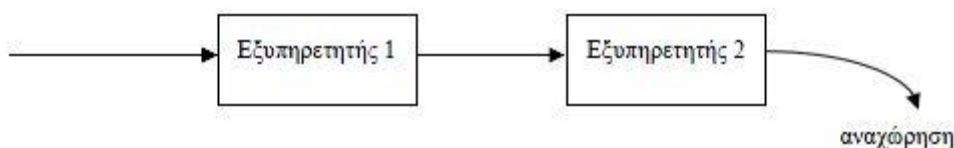
$$E(W) = \Pi_W \frac{1}{1 - \rho} \frac{1}{c\mu} \quad (4.2.7)$$

Οι παραπάνω σχέσεις για $E(L^q)$ και $E(W)$ μπορούν επίσης να βρεθούν χρησιμοποιώντας την τεχνική μέσης τιμής. Αν δεν είναι όλοι οι εξυπηρετητές απασχολημένοι στην άφιξη, ο χρόνος αναμονής είναι μηδέν. Αν είναι όλοι οι εξυπηρετητές απασχολημένοι και υπάρχουν μηδέν ή περισσότεροι πελάτες που αναμένουν, τότε ένας νέος πελάτης που φθάνει στο σύστημα πρέπει πρώτα να περιμένει μέχρι την πρώτη αναχώρηση και τότε συνεχίζει να περιμένει για όσες αναχωρήσεις όσοι ήταν οι πελάτες που περίμεναν κατά την άφιξη. Ο χρόνος ενδο-αναχώρησης είναι ο ελάχιστος των c εκθετικών χρόνων εξυπηρέτησης με μέσο $1/\mu$ και έτσι είναι εκθετικός με μέσο $1/c\mu$. Συνεπώς λαμβάνουμε ότι:

$$E(W) = \Pi_W \frac{1}{c\mu} + E(L^q) \frac{1}{c\mu} \quad (4.2.8)$$

3.2.3. Ουρές Tandem

Η αντιστρεψιμότητα στο χρόνο της ουράς M/M/c έχει σημαντικές συνέπειες στη θεωρία αναμονής. Για παράδειγμα θεωρούμε ένα σύστημα δύο εξυπηρετητών στο οποίο οι πελάτες φθάνουν με αφίξεις Poisson ρυθμού λ στον πρώτο εξυπηρετητή. Μετά την εξυπηρέτησή τους από τον πρώτο εξυπηρετητή, πηγαίνουν στη δεύτερη ουρά του δεύτερου εξυπηρετητή. Θεωρούμε άπειρο χώρο αναμονής και στους δύο εξυπηρετητές. Κάθε εξυπηρετητής εξυπηρετεί έναν πελάτη με ρυθμό μ_i , $i=1,2$. Ένα σύστημα αναμονής όπως αυτό ονομάζεται *tandem* ή *sequential system* (Adan, 2001).



Εικόνα 15 Ουρά Tandem

Δεδομένου ότι η έξοδος από τον πρώτο εξυπηρετητή αποτελεί διαδικασία Poisson, συνεπάγεται ότι ο δεύτερος εξυπηρετητής αποτελεί επίσης σύστημα M/M/1.

Σε μια εργοδική ουρά M/M/1 στη μόνιμη κατάσταση:

- ο αριθμός των πελατών που βρίσκονται στο σύστημα είναι ανεξάρτητος της σειράς των χρόνων αναχώρησης στο παρελθόν,
- ο χρόνος παραμονής στο σύστημα (χρόνος αναμονής στην ουρά και εξυπηρέτησης) ενός πελάτη είναι ανεξάρτητος της διαδικασίας αναχωρήσεων πριν την αναχώρησή του.

Για μια εργοδική ουρά tandem στη μόνιμη κατάσταση:

- οι αριθμοί των πελατών που βρίσκονται στον εξυπηρετητή 1 και τον εξυπηρετητή 2 είναι ανεξάρτητοι και:
- $P(n \text{ πελάτες στον } 1, m \text{ πελάτες στο } 2) = \left(\frac{\lambda}{\mu_1}\right)^n \left(1 - \frac{\lambda}{\mu_1}\right) \left(\frac{\lambda}{\mu_2}\right)^m \left(1 - \frac{\lambda}{\mu_2}\right)$ ο χρόνος αναμονής ενός πελάτη στον εξυπηρετητή 1 είναι ανεξάρτητος του χρόνου αναμονής στον εξυπηρετητή 2.

3.3. Ουρά M/M/∞

Στο σύστημα αυτό μπορούμε να θεωρήσουμε είτε ότι υπάρχει μια μονάδα εξυπηρέτησης, η οποία επιταχύνει τον ρυθμό της γραμμικά όσο έρχονται

περισσότεροι πελάτες, είτε ότι διατίθεται πάντα μια καινούργια μονάδα, για κάθε πελάτη που φθάνει στο σύστημα. Εδώ, ο αριθμός των πελατών στο σύστημα ισοδυναμεί με τον αριθμό πελατών που εξυπηρετούνται, εφ' όσον δεν υπάρχει αναμονή (Μάγκλαρης, Παπαβασιλείου, 2011).



Εικόνα 16 Διάγραμμα ροής για το μοντέλο M/M/∞

Από το παραπάνω διάγραμμα ροής, εξισώνοντας τη ροή από την κατάσταση n-1 στη n και τη ροή από τη n στη n-1 έχουμε ότι:

$$p_{n-1}\lambda = p_n n\mu$$

Συνεπώς:

$$p_n = \frac{\lambda}{n\mu} p_{n-1} = \frac{\rho}{n} p_{n-1} = \frac{\rho^2}{n(n-1)} p_{n-2} = \dots = \frac{\rho^n}{n!} p_0$$

Αφού το άθροισμα των πιθανοτήτων p_n είναι ένα, προκύπτει ότι:

$$p_0^{-1} = \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = e^{\rho}$$

Συνοψίζοντας, βρήκαμε ότι η στατική κατανομή υπάρχει πάντα και δίνεται από τη σχέση:

$$p_n = \frac{\rho^n}{n!} e^{-\rho}, \quad n = 0, 1, 2, \dots \quad (4.3.1)$$

όπου $\rho = \lambda/\mu$. Παρατηρούμε ότι πρόκειται για κατανομή Poisson. Ο μέσος αριθμός πελατών στο σύστημα είναι:

$$E(L) = \rho = \frac{\lambda}{\mu} \quad (4.3.2)$$

και ο μέσος χρόνος απόκρισης είναι προφανώς $1/\mu$ ίσος με το μέσο χρόνο εξυπηρέτησης.

3.4. Ουρά M/M/1/K

Το σύστημα M/M/1/K αποτελεί ένα ακριβέστερο μοντέλο πραγματικών συστημάτων σε σχέση με μοντέλα απείρου χώρου αναμονής, στο οποίο ένας πελάτης μπορεί να

εξυπηρετείται και το πολύ $K-1$ να περιμένουν. Υποθέτουμε ότι οι πελάτες που φθάνοντας βρίσκουν K πελάτες στο σύστημα φεύγουν και “χάνονται” (Γεωργίου, 2011).

Οι ρυθμοί γεννήσεων-θανάτων θα είναι:

$$\lambda_n = \begin{cases} \lambda, n < K \\ 0, n \geq K \end{cases} \quad (4.4.1)$$

$$\mu_n = \begin{cases} \mu, n \leq K \\ 0, n > K \end{cases} \quad (4.4.2)$$

Παρατηρούμε ότι πρόκειται για μια αμείωτη αλυσίδα Markov με πεπερασμένο αριθμό καταστάσεων, η οποία είναι πάντα εργοδική, και επομένως υπάρχει πάντα η στατική κατανομή πιθανότητας:

$$p_n = p_0 \rho^n, n = 0, 1, 2, \dots, K \quad (4.4.3)$$

όπου $\rho = \lambda/\mu$. Εύκολα βρίσκουμε ότι:

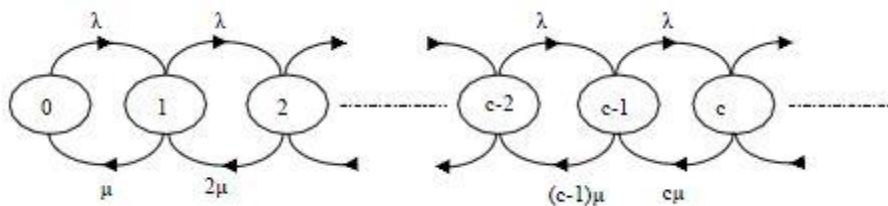
$$p_0 = \frac{1 - \rho}{1 - \rho^{K+1}}, \rho \neq 1 \quad (4.4.4)$$

Για την περίπτωση $\lambda = \mu$ ή $\rho = 1$ ισχύει:

$$p_n = \frac{1}{K + 1}, n = 0, 1, \dots, K \quad (4.4.5)$$

3.5. Ουρά M/M/c/c

Στο σύστημα αυτό γίνονται δεκτοί μόνο οι πελάτες που βρίσκουν διαθέσιμη μονάδα εξυπηρέτησης, διαφορετικά οι πελάτες χάνονται. Κλασσική εφαρμογή ενός τέτοιου μοντέλου είναι το τηλεφωνικό κέντρο με c γραμμές εξυπηρέτησης (Yue et al, 2009).



Εικόνα 17 Σύστημα αναμονής c εξυπηρετητών με απώλειες

Για το σύστημα αυτό οι ρυθμοί γεννήσεων-θανάτων είναι:

$$\lambda_n = \begin{cases} \lambda, n < c \\ 0, n \geq c \end{cases} \quad (4.5.1)$$

$$\mu_n = \begin{cases} n\mu, n \leq c \\ 0, n > c \end{cases} \quad (4.5.2)$$

Από το διάγραμμα ροής προκύπτει ότι:

$$p_{n-1} \lambda = p_n n \mu, n = 1, 2, \dots, c$$

Και στην περίπτωση αυτή, η αλυσίδα Markov είναι πάντα εργοδική, οπότε η στατική κατανομή πιθανότητας βρίσκεται:

$$p_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n! \sum_{i=0}^c \frac{\left(\frac{\lambda}{\mu}\right)^i}{i!}} = p_0 \frac{\rho^n}{n!}, n = 0, 1, 2, \dots, c \quad (4.5.3)$$

θέτοντας, όπως συνήθως $\rho = \frac{\lambda}{\mu}$. Η πιθανότητα p_0 δίνεται από τη σχέση:

$$p_0 = \left(\sum_{i=0}^c \frac{\rho^i}{i!} \right)^{-1} \quad (4.5.4)$$

Ιδιαίτερο ενδιαφέρον παρουσιάζει η πιθανότητα p_c να είναι απασχολημένες όλες οι μονάδες εξυπηρέτησης, η οποία ισοδυναμεί με το ποσοστό των πελατών που χάνονται:

$$p_c = \frac{\frac{\rho^c}{c!}}{1 + \rho + \frac{\rho^2}{2!} + \dots + \frac{\rho^c}{c!}} \quad (4.5.5)$$

Η σχέση (4.5.5) είναι γνωστή στην τηλεφωνία σαν ο τύπος “απώλειας” του Erlang (*Loss formula*) και υπολογίστηκε για πρώτη φορά από τον A.K. Erlang το 1917.

Η επονομαζόμενη πιθανότητα εμπλοκής (blocking probability) δίνεται από τον τύπο:

$$B(c, \rho) = p_c = \frac{\frac{\rho^c}{c!}}{\sum_{i=0}^c \frac{\rho^i}{i!}} \quad (4.5.6)$$

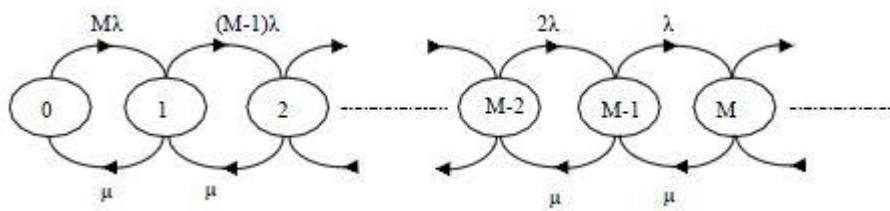
Από το νόμο του Little προκύπτει ότι:

$$E(L) = \rho(1 - B(c, \rho))$$

3.6. Ουρά M/M/1//M

Στο σύστημα αναμονής αυτό θεωρούμε ότι ο πληθυσμός από τον οποίο προέρχονται οι πελάτες είναι πεπερασμένος. Υπάρχουν δηλαδή μονάχα M πελάτες του συστήματος. Ένας πελάτης θα βρίσκεται είτε στο σύστημα (εξυπηρετούμενος ή σε αναμονή) είτε εκτός συστήματος (φθάνοντας στο σύστημα). Στη δεύτερη περίπτωση ο χρόνος που απαιτείται μέχρι την άφιξη είναι μια τυχαία μεταβλητή κατανεμημένη εκθετικά με μέση τιμή $1/\lambda$ (Morgan, 1999).

Όλοι οι πελάτες δρουν ανεξάρτητα ο ένας από τον άλλο. Όταν υπάρχουν k πελάτες στο σύστημα (αναμονή και εξυπηρέτηση), υπάρχουν $M-k$ πελάτες, οι οποίοι “φθάνουν” στο σύστημα, και κατά συνέπεια ο συνολικός μέσος ρυθμός αφίξεων στην κατάσταση αυτή είναι $\lambda(M-k)$. Το σύστημα ρυθμίζει τις επιδόσεις του μόνο του. Δηλαδή σε περίπτωση που πολλοί πελάτες περιμένουν προκειμένου να εξυπηρετηθούν, ο ρυθμός αφίξεων μειώνεται ανάλογα, ελαττώνοντας το φαινόμενο της συμφόρησης.



Εικόνα 18 Σύστημα αναμονής M/M/1/M

Οι παράμετροι της διαδικασίας γεννήσεων-θανάτων είναι:

$$\lambda_k = \begin{cases} \lambda(M-k), & 0 \leq k \leq M \\ 0, & \text{αλλιού} \end{cases} \quad (4.6.1)$$

$$\mu_k = \mu, k = 1, 2, \dots \quad (4.6.2)$$

Το σύστημα είναι εργοδικό.

3.7. Ουρά M/Er/1

Πριν την ανάλυση της ουράς M/G/1, μελετάται η ουρά M/Er/1. Η κατανομή Erlang μπορεί να χρησιμοποιηθεί για να μοντελοποιήσει χρόνους εξυπηρέτησης με χαμηλό συντελεστή διακύμανσης (μικρότερο του ενός), αλλά μπορεί να προκύψει και φυσικά. Για παράδειγμα, αν μια εργασία πρέπει να περάσει, από στάδιο σε στάδιο, από μία σειρά r ανεξάρτητων σταδίων παραγωγής, όπου κάθε στάδιο έχει εκθετικά κατανομημένο χρόνο. Η ανάλυση της ουράς M/Er/1 είναι παρόμοια με αυτή της ουράς M/M/1 (Adan, 2001).

Θεωρούμε μια ουρά με ένα μοναδικό εξυπηρετητή. Οι πελάτες φθάνουν σε αυτή σύμφωνα με διαδικασία Poisson με ρυθμό λ και εξυπηρετούνται με σειρά άφιξης. Οι χρόνοι εξυπηρέτησης είναι Erlang- r κατανομημένοι με μέσο r/μ .

Για σταθερότητα απαιτείται ο ρυθμός εξυπηρέτησης

$$\rho = \lambda \frac{r}{\mu} \quad (4.7.1)$$

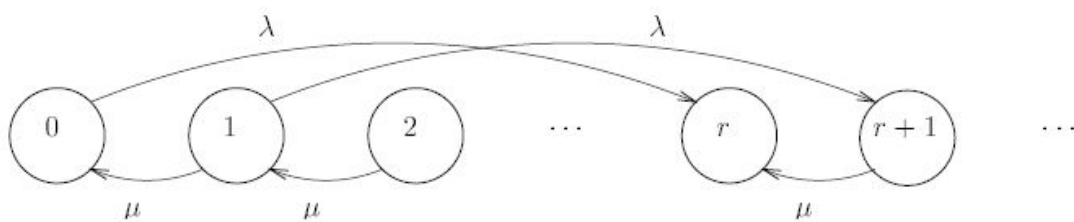
να είναι μικρότερος του ενός.

3.7.1. Δυο εναλλακτικές περιγραφές των σταδίων

Ο φυσικός τρόπος για να περιγραφεί η κατάσταση ενός μη κενού συστήματος είναι με το ζευγάρι (k,l) όπου το k δηλώνει τον αριθμό των πελατών στο σύστημα και l τον υπολειπόμενο αριθμό φάσεων για τον πελάτη στο σύστημα. Αυτή είναι εμφανώς μια δισδιάστατη περιγραφή. Ένας εναλλακτικός τρόπος για να περιγραφεί η κατάσταση είναι με τη μέτρηση του συνολικού αριθμού των μη ολοκληρωμένων φάσεων εργασίας στο σύστημα. Υπάρχει μια ένα προς ένα αντιστοίχιση ανάμεσα σε αυτό τον αριθμό και το ζεύγος (k,l) . Ο αριθμός των μη ολοκληρωμένων φάσεων εργασίας στο σύστημα ισούται με $(k-1)r + l$ (για τον πελάτη στο σύστημα υπάρχουν l φάσεις εργασίας αντί για r). Θα ακολουθηθεί η μονοδιάστατη περιγραφή των σταδίων (Ford, 1980).

3.7.2. Κατανομή ισορροπίας

Από τη μονοδιάστατη περιγραφή προκύπτει το ακόλουθο διάγραμμα ροής:



Εικόνα 19 Μονοδιάστατο διάγραμμα ροής για το μοντέλο M/Er/1

Έστω p_n η πιθανότητα ισορροπίας των n φάσεων εργασίας στο σύστημα. Με την εξίσωση της ροής έξω από την κατάσταση n και της ροής προς την κατάσταση n προκύπτει το παρακάτω σύνολο εξισώσεων ισορροπίας για την p_n :

$$p_0 \lambda = p_1 \mu \quad (4.7.2)$$

$$p_n (\lambda + \mu) = p_{n+1} \mu, n = 1, \dots, r-1 \quad (4.7.3)$$

$$p_n (\lambda + \mu) = p_{n-r} \lambda + p_{n+1} \mu, n = r, r+1, r+2, \dots \quad (4.7.4)$$

Αυτές οι εξισώσεις μπορούν να λυθούν ως εξής. Αρχικά αναζητούμε λύσεις της (4.7.4) της μορφής:

$$p_n = x^n, n = 0, 1, 2, \dots \quad (4.7.5)$$

και στη συνέχεια κατασκευάζουμε ένα γραμμικό συνδυασμό αυτών των λύσεων που ικανοποιούν τις οριακές εξισώσεις (4.7.2) και (4.7.3) και την εξίσωση κανονικοποίησης:

$$\sum_{n=0}^{\infty} p_n = 1$$

Μια εναλλακτική προσέγγιση επίλυσης είναι με τη χρήση συναρτήσεων παραγωγής. Η αντικατάσταση της (4.7.5) στην (4.7.4) και στη συνέχεια η διαίρεση με την κοινή δύναμη x^{n-r} δίνει την πολυωνυμική εξίσωση:

$$(\lambda + \mu)x^r = \lambda + \mu x^{r+1} \quad (4.7.6)$$

Η μία ρίζα είναι η $x=1$, αλλά δεν είναι χρήσιμη, γιατί πρέπει να υπάρχει η δυνατότητα κανονικοποίησης της λύσης των εξισώσεων ισορροπίας. Εάν η συνθήκη (4.7.1) ισχύει, μπορεί να αποδειχθεί ότι η εξίσωση (4.7.6) έχει ακριβώς r διακριτές ρίζες x με $|x| < 1$, έστω x_1, \dots, x_r . Θεωρούμε τώρα το γραμμικό συνδυασμό:

$$p_n = \sum_{k=1}^r c_k x_k^n, n = 0, 1, 2, \dots \quad (4.7.7)$$

Για κάθε επιλογή των συντελεστών c_k αυτός ο γραμμικός συνδυασμός ικανοποιεί την (4.7.4). Αυτή η ελευθερία επιλογής χρησιμοποιείται για να ικανοποιήσει τις εξισώσεις (4.7.2) και (4.7.3) και την εξίσωση κανονικοποίησης. Σημειώνεται ότι, αφού οι εξισώσεις ισορροπίας είναι εξαρτημένες, η εξίσωση (4.7.2) (ή κάποια από τις εξισώσεις (4.7.3)) μπορεί να παραλειφθεί. Η αντικατάσταση της (4.7.7) σε αυτές στις εξισώσεις δίνει ένα σύνολο r γραμμικών συνδυασμών για r διαφορετικούς συντελεστές. Μπορεί να αποδειχθεί ότι αυτό το σύνολο εξισώσεων έχει μοναδική λύση, που δίνεται από:

$$c_k = \frac{1 - \rho}{\prod_{j \neq k} \left(1 - \frac{x_j}{x_k}\right)}, k = 1, \dots, r$$

όπου ο δείκτης j κυμαίνεται από 1 έως r . Αυτό ολοκληρώνει τον καθορισμό των εξισώσεων ισορροπίας p_n . Το σημαντικό συμπέρασμα είναι ότι για την ουρά $M/E_r/1$ οι πιθανότητες ισορροπίας μπορούν να εκφραστούν ως μείγμα r γεωμετρικών κατανομών.

Από την κατανομή του αριθμού των φάσεων στο σύστημα μπορεί να βρεθεί η κατανομή του αριθμού των πελατών στο σύστημα. Έστω q_i η πιθανότητα για i πελάτες στο σύστημα.

Ισχύει ότι $q_0 = p_0$ και για $i \geq 1$:

$$\begin{aligned}
 q_i &= \sum_{n=(i-1)r+1}^{ir} p_n \\
 &= \sum_{n=(i-1)r+1}^{ir} \sum_{k=1}^r c_k x_k^n \\
 &= \sum_{k=1}^r \sum_{n=(i-1)r+1}^{ir} c_k x_k^n \\
 &= \sum_{k=1}^r c_k (x_k^{-r+1} + x_k^{-r+2} + \dots + 1) (x_k^r)^i
 \end{aligned}$$

Συνεπώς, οι πιθανότητες μήκους της ουράς μπορούν επίσης να εκφραστούν ως μείγμα r γεωμετρικών κατανομών.

3.7.3. Μέσος χρόνος αναμονής

Έστω L^f η τυχαία μεταβλητή που δηλώνει τον αριθμό των φάσεων εργασίας στο σύστημα. Τότε ισχύει:

$$E(W) = E(L^f) \frac{1}{\mu}$$

και από την (4.7.7) προκύπτει:

$$\begin{aligned}
 E(L^f) &= \sum_{n=0}^{\infty} n p_n \\
 &= \sum_{n=0}^{\infty} \sum_{k=1}^r c_k n x_k^n \\
 &= \sum_{k=1}^r \sum_{n=0}^{\infty} c_k n x_k^n
 \end{aligned}$$

$$= \sum_{k=1}^r \frac{c_k x_k}{(1 - x_k)^2}$$

Από το νόμο του Little μπορεί επίσης να προκύψει ο μέσος αριθμός πελατών που περιμένουν σε μία ουρά, $E(L^q)$.

Ένας πελάτης που φθάνει στο σύστημα πρέπει να περιμένει για τους πελάτες στην ουρά και, εάν ο εξυπηρετητής είναι απασχολημένος, και για τον πελάτη που εξυπηρετείται. Ο μέσος αριθμός των πελατών που περιμένουν στο σύστημα είναι ίσος με $E(L^q)$ και η πιθανότητα ότι ο εξυπηρετητής είναι απασχολημένος κατά την άφιξη είναι ρ , δηλαδή το κλάσμα του χρόνου για το οποίο είναι απασχολημένος. Συνεπώς:

$$E(W) = E(L^q) \frac{r}{\mu} + \rho E(R) \quad (4.7.8)$$

όπου η τυχαία μεταβλητή R δηλώνει τον υπολειπόμενο χρόνο εξυπηρέτησης του πελάτη στο σύστημα. Αν ο εξυπηρετητής είναι απασχολημένος κατά την άφιξη, τότε με πιθανότητα $1/r$ θα είναι απασχολημένος με την πρώτη φάση του χρόνου εξυπηρέτησης, επίσης με πιθανότητα $1/r$ θα είναι απασχολημένος με τη δεύτερη φάση, κ.ο.κ. Έτσι ο μέσος υπολειπόμενος χρόνος εξυπηρέτησης $E(R)$ είναι ίσος με:

$$\begin{aligned} E(R) &= \frac{1}{r} \frac{r}{\mu} + \frac{1}{r} \frac{r-1}{\mu} + \dots + \frac{1}{r} \frac{1}{\mu} \\ &= \frac{r+1}{2} \frac{1}{\mu} \quad (4.7.9) \end{aligned}$$

Η αντικατάσταση αυτής της σχέσης στην (4.7.8) δίνει:

$$E(W) = E(L^q) \frac{r}{\mu} + \rho \frac{r+1}{2} \frac{1}{\mu}$$

Μαζί με το νόμο του Little, ο οποίος δηλώνει ότι:

$$E(L^q) = \lambda E(W)$$

προκύπτει:

$$E(W) = \frac{\rho}{1-\rho} \frac{r+1}{2} \frac{1}{\mu}$$

3.7.4. Κατανομή του χρόνου αναμονής

Ο χρόνος αναμονής μπορεί να εκφραστεί ως:

$$W = \sum_{i=1}^{L^f} B_i$$

όπου B_i είναι το ποσό εργασίας στο k -οστό στάδιο. Συνεπώς, οι τυχαίες μεταβλητές B_i είναι ανεξάρτητες και εκθετικά κατανομημένες με μέσο $1/\mu$. Με τη συνθήκη του L^f και χρησιμοποιώντας ότι οι L^f και B_i είναι ανεξάρτητες, προκύπτει ότι (Yue et al, 2009):

$$\begin{aligned}
 P(W > t) &= \sum_{n=1}^{\infty} P\left(\sum_{i=1}^n B_i > t\right) p_n \\
 &= \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \frac{(\mu t)^i}{i!} e^{-\mu t} \sum_{k=1}^r c_k x_k^n \\
 &= \sum_{k=1}^r c_k \sum_{i=0}^{\infty} \sum_{n=i+1}^{\infty} \frac{(\mu t)^i}{i!} e^{-\mu t} x_k^n \\
 &= \sum_{k=1}^r \frac{c_k x_k}{1 - x_k} \sum_{i=0}^{\infty} \frac{(\mu x_k t)^i}{i!} e^{-\mu t} \\
 &= \sum_{k=1}^r \frac{c_k}{1 - x_k} x_k e^{-\mu(1-x_k)t}, t \geq 0
 \end{aligned}$$

Να σημειωθεί ότι αυτή η κατανομή είναι μια γενίκευση της κατανομής του μοντέλου M/M/1, και για την ακρίβεια είναι ένα μείγμα εκθετικών.

3.8. Ουρά M/G/1

Στην ουρά M/G/1 οι πελάτες φθάνουν σύμφωνα με διαδικασία Poisson με ρυθμό λ και εξυπηρετούνται σύμφωνα με τη σειρά άφιξης. Οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι και όμοια κατανομημένοι με συνάρτηση κατανομής $F_B(\cdot)$ και πυκνότητα $f_B(\cdot)$. Για σταθερότητα πρέπει να απαιτήσουμε ο ρυθμός απασχόλησης (Adan, 2001)

$$\rho = \lambda E(B) \quad (4.8.1)$$

να είναι μικρότερος του ενός.

3.8.1. Περιοριστική κατανομή

Η κατάσταση της ουράς M/G/1 μπορεί να περιγραφεί από το ζεύγος (n, x) όπου το n δηλώνει τον αριθμό των πελατών στο σύστημα και το x το χρόνο εξυπηρέτησης που έχει ήδη καταναλώσει ο πελάτης που εξυπηρετείται. Χρειάζεται λοιπόν μια δισδιάστατη περιγραφή καταστάσεων. Η πρώτη διάσταση είναι διακριτή, αλλά η δεύτερη είναι συνεχής και αυτό δυσκολεύει την ανάλυση. Ωστόσο, αν εξετάσουμε το

σύστημα ακριβώς μετά τις αναχωρήσεις, τότε η περιγραφή των καταστάσεων μπορεί να απλοποιηθεί μόνο σε n , διότι $x=0$ για τους νέους πελάτες (αν υπάρχουν) στην υπηρεσία. L_k^d είναι ο αριθμός των πελατών που μένουν πίσω από τον k -οστό πελάτη που αναχωρεί (Ford, 1980). Θα καθορίσουμε την περιοριστική κατανομή:

$$d_n = \lim_{k \rightarrow \infty} P(L_k^d = n)$$

Η πιθανότητα d_n μπορεί να ερμηνευθεί ως το κλάσμα των πελατών που αφήνουν πίσω τους n πελάτες. Η περιοριστική κατανομή είναι χρησιμότερο να ορίζεται ως:

$$p_n = \lim_{t \rightarrow \infty} P(L(t) = n)$$

όπου $L(t)$ είναι ο αριθμός των πελατών στο σύστημα τη στιγμή t . Η πιθανότητα p_n μπορεί να ερμηνευθεί ως το κλάσμα του χρόνου για το οποίο υπάρχουν n πελάτες στο σύστημα. Από αυτή την κατανομή μπορεί να υπολογισθεί ο μέσος αριθμός πελατών στο σύστημα. Μια ακόμα πιο σημαντική κατανομή είναι η περιοριστική κατανομή του αριθμού των πελατών στο σύστημα που φαίνονται από έναν πελάτη που φθάνει, δηλαδή:

$$\alpha_n = \lim_{k \rightarrow \infty} P(L_k^a = n)$$

όπου L_k^a είναι ο αριθμός των πελατών στο σύστημα ακριβώς πριν φθάσει ο k -οστός πελάτης. Με αυτή την κατανομή υπολογίζεται η κατανομή του χρόνου παραμονής.

Μπορεί να αποδειχθεί ότι, για την ουρά $M/G/1$, οι αφίξεις, οι αναχωρήσεις και οι εξωτερικοί παρατηρητές βλέπουν την ίδια κατανομή του αριθμού των πελατών στο σύστημα, δηλαδή για κάθε n ισχύει:

$$\alpha_n = d_n = p_n$$

3.8.2. Κατανομή αναχωρήσεων

Συμβολίζουμε με L_k^d τον αριθμό των πελατών που μένουν πίσω από τον k -πελάτη που αναχωρεί. Ο αριθμός των πελατών που μένουν πίσω από τον $k+1$ -οστό πελάτη είναι προφανώς ίσος με τον αριθμό των πελατών που βρίσκονται στο σύστημα όταν αναχώρησε ο k -οστός πελάτης πλην ενός (αφού ο $k+1$ -οστός πελάτης αναχώρησε ο ίδιος) συν τον αριθμό των πελατών που έφθασαν στο σύστημα κατά το χρόνο εξυπηρέτησής του. Ο τελευταίος αριθμός συμβολίζεται με την τυχαία μεταβλητή A_{k+1} (Adan, 2001). Έτσι έχουμε:

$$L_{k+1}^d = L_k^d - 1 + A_{k+1}$$

ο οποίος είναι έγκυρος αν $L_k^d > 0$.

Τώρα ορίζουμε τις πιθανότητες μετάβασης:

$$p_{i,j} = P(L_{k+1}^d = j | L_k^d = i)$$

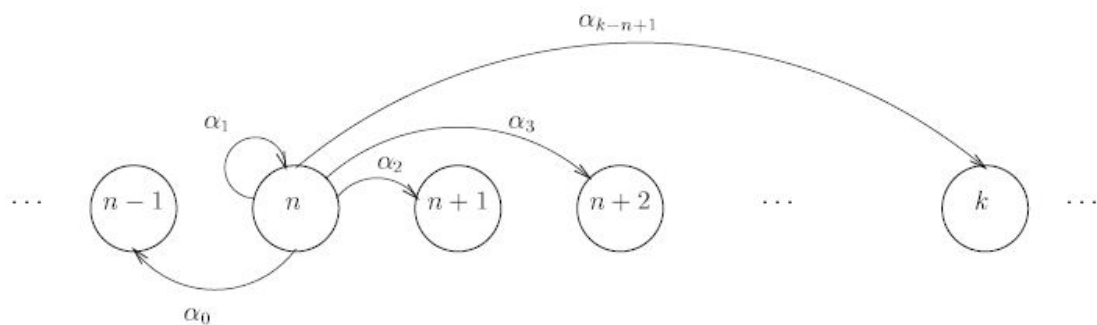
Προφανώς $p_{i,j} = 0$ για κάθε $j < i - 1$ και $p_{i,j}$ για $j \geq i - 1$ δηλώνει την πιθανότητα ότι ακριβώς $j - i + 1$ πελάτες έφθασαν κατά το χρόνο εξυπηρέτησης του $k+1$ -οστού πελάτη. Αυτό ισχύει για $i > 0$. Στην κατάσταση 0 ο k -οστός πελάτης αφήνει πίσω του ένα άδειο σύστημα και τότε η $p_{0,j}$ δίνει την πιθανότητα ότι κατά τη διάρκεια του χρόνου εξυπηρέτησης του $k+1$ -οστού πελάτη ακριβώς j πελάτες έφθασαν. Συνεπώς ο πίνακας P των πιθανοτήτων μετάβασης παίρνει τη μορφή:

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ \mathbf{0} & a_0 & a_1 & a_2 & \dots \\ \mathbf{0} & \mathbf{0} & a_0 & a_1 & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & a_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

όπου a_n δηλώνει την πιθανότητα ότι κατά τη διάρκεια του χρόνου εξυπηρέτησης φθάνουν ακριβώς n πελάτες. Για να υπολογίσουμε το a_n σημειώνουμε ότι δεδομένης της διάρκειας του χρόνου εξυπηρέτησης, έστω t , ο αριθμός των πελατών που φθάνουν κατά τη διάρκεια του χρόνου εξυπηρέτησης είναι κατανομημένος σύμφωνα με Poisson με παράμετρο λt . Συνεπώς ισχύει:

$$a_n = \int_{t=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} f_B(t) dt \quad (4.8.2)$$

Το διάγραμμα πιθανοτήτων μετάβασης φαίνεται στο ακόλουθο σχήμα:



Εικόνα 20 Διάγραμμα πιθανοτήτων μετάβασης για το μοντέλο M/G/1

Οι περιοριστικές πιθανότητες d_n , που γνωρίζουμε ότι είναι ίσες με a_n , ικανοποιούν τις εξισώσεις ισορροπίας:

$$d_n = d_{n+1}a_0 + d_n a_1 + \dots + d_1 a_n + d_0 a_n$$

$$= \sum_{k=0}^n d_{n+1-k} a_k + d_0 a_n, n = 0, 1, \dots \quad (4.8.3)$$

Για να λύσουμε τις εξισώσεις ισορροπίας πρέπει να χρησιμοποιήσουμε την προσέγγιση των παραγόμενων συναρτήσεων. Εισάγονται οι πιθανοτικές παραγόμενες συναρτήσεις:

$$P_{L^d}(z) = \sum_{n=0}^{\infty} d_n z^n$$

και

$$P_A(z) = \sum_{n=0}^{\infty} a_n z^n$$

οι οποίες ορίζονται για κάθε $z \leq 1$. Πολλαπλασιάζοντας την (4.8.3) με z^n και αθροίζοντας για όλα τα n οδηγούμαστε σε:

$$\begin{aligned} P_{L^d}(z) &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^n d_{n+1-k} a_k + d_0 a_n \right) z^n \\ &= z^{-1} P_A(z) (P_{L^d}(z) - d_0) + d_0 P_A(z) \end{aligned}$$

Συνεπώς βρίσκουμε:

$$P_{L^d}(z) = \frac{d_0 P_A(z) (1 - z^{-1})}{1 - z^{-1} P_A(z)}$$

Για να βρεθεί η πιθανότητα d_0 σημειώνουμε ότι $d_0 = p_0$, το οποίο είναι το κλάσμα του χρόνου για το οποίο το σύστημα είναι άδειο. Συνεπώς $d_0 = p_0 = 1 - \rho$. Έτσι, πολλαπλασιάζοντας τον αριθμητή με τον παρονομαστή με $-z$ έχουμε:

$$P_{L^d}(z) = \frac{(1 - \rho) P_A(z) (1 - z)}{P_A(z) - z} \quad (4.8.4)$$

Χρησιμοποιώντας την (4.8.2) η παραγόμενη συνάρτηση $P_A(z)$ μπορεί να ξαναγραφτεί ως:

$$P_A(z) = \tilde{B}(\lambda - \lambda z) \quad (4.8.5)$$

Η αντικατάσταση της (4.8.5) στην (4.8.4) δίνει τελικά:

$$P_{L^d}(z) = \frac{(1 - \rho) \tilde{B}(\lambda - \lambda z) (1 - z)}{\tilde{B}(\lambda - \lambda z) - z} \quad (4.8.6)$$

3.8.3. Κατανομή του χρόνου παραμονής

Ας θεωρήσουμε έναν πελάτη που φθάνει στο σύστημα σε ισορροπία. Συμβολίζουμε το συνολικό χρόνο που ο πελάτης μένει στο σύστημα με την τυχαία μεταβλητή S με

συνάρτηση κατανομής $F_S(\cdot)$ και συνάρτηση πυκνότητας πιθανότητας $f_S(\cdot)$. Η κατανομή του αριθμού των πελατών που αφήνονται πίσω κατά την αναχώρηση του πελάτη είναι ίση με $\{d_n\}_{n=0}^{\infty}$. Θεωρώντας ένα σύστημα με ουρά FCFS² είναι εμφανές ότι όλοι οι πελάτες που αφήνονται πίσω είναι ακριβώς αυτοί που έφτασαν κατά την παραμονή του πελάτη στο σύστημα (Yue et al, 2009). Συνεπώς έχουμε:

$$d_n = \int_{t=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} f_S(t) dt$$

Παρόμοια με την (4.8.5) βρίσκουμε ότι:

$$P_{Ld}(z) = \tilde{S}(\lambda - \lambda z)$$

Η αντικατάσταση αυτής της σχέσης στην (4.8.6) δίνει:

$$\tilde{S}(\lambda - \lambda z) = \frac{(1 - \rho)\tilde{B}(\lambda - \lambda z)(1 - z)}{\tilde{B}(\lambda - \lambda z) - z}$$

Κάνοντας αλλαγή μεταβλητής $s = \lambda - \lambda z$ έχουμε τελικά ότι:

$$\tilde{S}(s) = \frac{(1 - \rho)\tilde{B}(s) s}{\lambda\tilde{B}(s) + s - \lambda} \quad (4.8.7)$$

Αυτή η σχέση είναι γνωστή και ως *Σχέση Pollaczek-Khinchin*.

3.8.4. Κατανομή του χρόνου αναμονής

Ισχύει ότι το S , ο χρόνος παραμονής του πελάτη στο σύστημα, είναι το άθροισμα του W , του χρόνου αναμονής του, και του B , του χρόνου εξυπηρέτησης, όπου τα W και B είναι ανεξάρτητα. Αφού ο μετασχηματισμός του αθροίσματος δυο τυχαίων μεταβλητών είναι το γινόμενο των μετασχηματισμών αυτών των δυο τυχαίων μεταβλητών έχουμε:

$$\tilde{S}(s) = \tilde{W}(s) \cdot \tilde{B}(s) \quad (4.8.8)$$

Μαζί με την (4.8.7) προκύπτει ότι:

$$\tilde{W}(s) = \frac{(1 - \rho)s}{\lambda\tilde{B}(s) + s - \lambda} \quad (4.8.9)$$

η οποία είναι η Τρίτη μορφή της *Σχέσης Pollaczek-Khinchin*.

3.8.5. Εξίσωση του Lindley

Συμβολίζουμε με A_n τον ενδιάμεσο χρόνο μεταξύ των αφίξεων του n -οστού και του $n+1$ -οστού πελάτη. Επιπλέον, W_n και B_n είναι ο χρόνος αναμονής και εξυπηρέτησης αντίστοιχα. Ισχύει ότι:

² First come-first served

$$W_{n+1} = \max(W_n + B_n - A_n, 0) \quad (4.8.10)$$

Αυτή η εξίσωση αναφέρεται ως *εξίσωση του Lindley*. Τείνοντας το n στο άπειρο στη σχέση (4.8.10) προκύπτει η περιοριστική μορφή:

$$W = \max(W + B - A, 0) \quad (4.8.11)$$

Με

$$S = W + B$$

η παραπάνω εξίσωση μπορεί να γραφεί ως:

$$W = \max(S - A, 0) \quad (4.8.12)$$

Ο χρόνος μεταξύ των αφίξεων A είναι εκθετικά κατανομημένος με παράμετρο λ . Επίσης, οι τυχαίες μεταβλητές S και A είναι ανεξάρτητες. Δεδομένων αυτών μπορεί να υπολογιστεί ο μετασχηματισμός του W απευθείας από τη σχέση (4.8.12), ο οποίος είναι:

$$\tilde{W}(s) = \frac{s}{s - \lambda} \tilde{S}(\lambda) - \frac{\lambda}{s - \lambda} \tilde{S}(s)$$

Με αντικατάσταση της (7.8.8) προκύπτει:

$$\tilde{W}(s) = \frac{\tilde{S}(\lambda)s}{\lambda\tilde{B}(s) + s - \lambda} \quad (4.8.13)$$

3.8.6. Προσέγγιση της μέσης τιμής

Η τυχαία μεταβλητή R δείχνει τον υπολειπόμενο χρόνο εξυπηρέτησης και το L^q τον αριθμό των πελατών που περιμένουν στην ουρά. Συνεπώς:

$$E(W) = E(L^q)E(B) + \rho E(R)$$

και με το νόμο του Little

$$E(L^q) = \lambda E(W)$$

Έτσι προκύπτει ότι:

$$E(W) = \frac{\rho E(R)}{1 - \rho} \quad (4.8.14)$$

Ο τύπος (4.8.14) αναφέρεται συχνά ως Τύπος των *Pollaczek-Khinchin για τη μέση τιμή*. Ισχύει ότι:

$$E(R) = \frac{E(B^2)}{2E(B)} \quad (4.8.15)$$

που μπορεί να γραφτεί και στη μορφή:

$$E(R) = \frac{\sigma_B^2 + E(B)^2}{2E(B)} = \frac{1}{2}(c_B^2 + 1)E(B) \quad (4.8.16)$$

3.8.7. Υπολειπόμενος χρόνος εξυπηρέτησης

Έστω ότι ο πελάτης φθάνει όταν ο εξυπηρετητής είναι απασχολημένος. Συμβολίζουμε το συνολικό χρόνο του πελάτη που εξυπηρετείται με X . $f_X(\cdot)$ είναι η συνάρτηση πυκνότητας πιθανότητας του X . Η βασική παρατήρηση για να βρεθεί η συνάρτηση αυτή είναι ότι είναι πολύ πιο πιθανό ο πελάτης να φθάσει σε μεγάλο χρόνο εξυπηρέτησης παρά σε μικρό. Έτσι η πιθανότητα ότι το X είναι μήκος x πρέπει να είναι ανάλογη στο μήκος x όπως και στη συχνότητα τέτοιων χρόνων εξυπηρέτησης, που είναι $f_B(x)dx$ (Adan, 2001). Συνεπώς μπορούμε να γράψουμε:

$$P(x \leq X \leq x + dx) = f_X(x)dx = C_X f_B(x)dx$$

όπου C είναι μια σταθερά για να κανονικοποιείται η πυκνότητα. Ισχύει:

$$C^{-1} = \int_{x=0}^{\infty} x f_B(x) dx = E(B)$$

Συνεπώς:

$$f_X(x) = \frac{x f_B(x)}{E(B)}$$

3.8.8. Διακύμανση του χρόνου αναμονής

Για να λάβουμε πληροφορίες για τη διακύμανση του χρόνου αναμονής μπορούμε να χρησιμοποιήσουμε τον τύπο (4.8.9) για το μετασχηματισμό Laplace του χρόνου αναμονής.

Αρχικά ξαναγράφουμε τον τύπο (4.8.9) ως εξής:

$$\tilde{W}(s) = \frac{1 - \rho}{1 - \rho \left(\frac{1 - \tilde{B}(s)}{sE(B)} \right)}$$

Ο όρος μεταξύ των παρενθέσεων είναι ο μετασχηματισμός του R , αφού ισχύει:

$$\tilde{R}(s) = \frac{1 - \tilde{B}(s)}{sE(B)}$$

Συνεπώς:

$$\tilde{W}(s) = \frac{1 - \rho}{1 - \rho \tilde{R}(s)} \quad (4.8.17)$$

Με παραγοντοποίηση της (4.8.17) και αντικατάσταση $s=0$ βρίσκουμε τον τύπο (4.8.14), δηλαδή:

$$E(W) = \frac{\rho E(R)}{1 - \rho}$$

Με παραγοντοποίηση της (4.8.17) δύο φορές και αντικαθιστώντας $s=0$ βρίσκουμε:

$$E(W^2) = 2(E(W))^2 + \frac{\rho E(R^2)}{1 - \rho}$$

Συνεπώς, για τη διακύμανση του χρόνου αναμονής ισχύει:

$$\sigma^2(W) = E(W^2) - E(W)^2 = \frac{\rho}{(1 - \rho)^2} \left(\rho(E(R))^2 + (1 - \rho)E(R^2) \right)$$

3.8.9. Κατανομή της απασχολημένης περιόδου

Η μέση διάρκεια της απασχολημένης περιόδου στην ουρά M/G/1 μπορεί να καθοριστεί με τον ίδιο τρόπο όπως και στην ουρά M/M/1. Συνεπώς έχουμε:

$$E(BP) = \frac{E(B)}{1 - \rho} \quad (4.8.18)$$

3.9. Ουρά G/M/1

Σε αυτό το κεφάλαιο μελετάται η ουρά G/M/1, η οποία αποτελεί το δυαδικό ισοδύναμο της ουράς M/G/1. Σε αυτό το σύστημα, οι πελάτες φθάνουν ένας ένας με χρόνους ανάμεσα στις αφίξεις όμοιους και εκθετικά κατανεμημένους σύμφωνα με μια μοναδιαία συνάρτηση κατανομής $F_A(\cdot)$ και πυκνότητας $f_A(\cdot)$. Ο μέσος χρόνος ανάμεσα στις αφίξεις είναι ίσος με $1/\lambda$. Οι χρόνοι εξυπηρέτησης είναι εκθετικά κατανεμημένοι με μέσο $1/\mu$. Για σταθερότητα απαιτούμε ο ρυθμός απασχόλησης

$$\rho = \frac{\lambda}{\mu}$$

να είναι μικρότερος του ενός.

Η κατάσταση της ουράς G/M/1 μπορεί να περιγραφεί από το ζευγάρι (n, x) , όπου το n δηλώνει τον αριθμό των πελατών στο σύστημα και το x το χρόνο που πέρασε από την τελευταία άφιξη. Συνεπώς, χρειαζόμαστε μια διδιάστατη περιγραφή. Όμοια όμως με την ουρά M/G/1, η περιγραφή των καταστάσεων μπορεί να είναι πολύ πιο απλή σε συγκεκριμένα σημεία στο χρόνο. Αν κοιτάξουμε στο σύστημα τις στιγμές των αφίξεων, η περιγραφή των καταστάσεων μπορεί να απλοποιηθεί μόνο στο n , διότι $x=0$ κατά την άφιξη. Συμβολίζουμε με L_k^a τον αριθμό των πελατών στο σύστημα ακριβώς πριν την άφιξη του k -οστού πελάτη (Adan, 2001). Η περιοριστική κατανομή είναι:

$$\alpha_n = \lim_{k \rightarrow \infty} P(L_k^a = n)$$

3.9.1. Κατανομή αφίξεων

Ορίζοντας την τυχαία μεταβλητή D_{k+1} ως τον αριθμό των πελατών που εξυπηρετούνται ανάμεσα στην άφιξη του k -οστού και του $k+1$ -οστού πελάτη, ισχύει ότι:

$$L_{k+1}^a = L_k^a + \mathbf{1} - D_{k+1}$$

Από την παραπάνω εξίσωση προκύπτει ότι η ακολουθία $\{L_k^a\}_{k=0}^{\infty}$ σχηματίζει μία αλυσίδα Markov. Αυτή η αλυσίδα ονομάζεται *ενταγμένη αλυσίδα Markov G/M/1*.

Πρέπει να υπολογιστούν οι σχετικές πιθανότητες μετάβασης:

$$p_{i,j} = P(L_{k+1}^a = j | L_k^a = i)$$

Ισχύει ότι $p_{i,j} = \mathbf{0}$ για κάθε $j > i + 1$ και $p_{i,j}$ για κάθε $j \leq i + 1$ είναι ίσο με την πιθανότητα ότι ακριβώς $i+1-j$ πελάτες εξυπηρετούνται κατά το ενδιάμεσο διάστημα αφίξεων του $k+1$ -οστού πελάτη. Συνεπώς ο πίνακας P των πιθανοτήτων μετάβασης παίρνει τη μορφή:

$$P = \begin{pmatrix} p_{0,0} & \beta_0 & \mathbf{0} & \mathbf{0} & \cdots \\ p_{1,0} & \beta_1 & \beta_0 & \mathbf{0} & \cdots \\ p_{2,0} & \beta_2 & \beta_1 & \beta_0 & \mathbf{0} \\ p_{3,0} & \beta_3 & \beta_2 & \beta_1 & \beta_0 \\ \vdots & & & & \ddots \end{pmatrix}$$

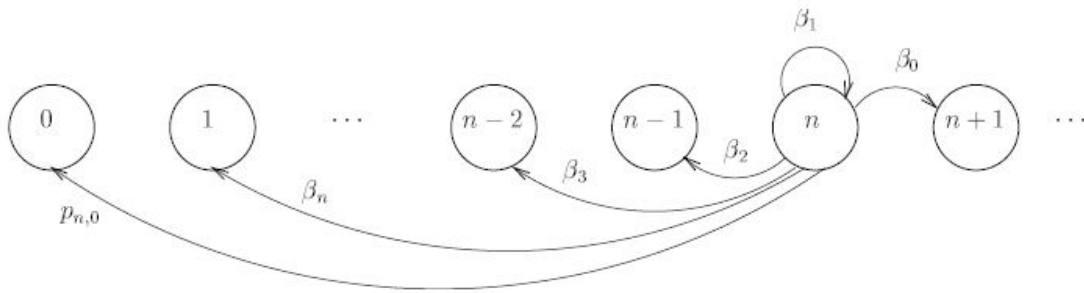
όπου β_i δηλώνει την πιθανότητα εξυπηρέτησης i πελατών κατά την διάρκεια ανάμεσα σε αφίξεις, δεδομένου ότι ο εξυπηρετητής παραμένει απασχολημένος αυτή την περίοδο. Ισχύει ότι:

$$\beta_i = \int_{t=0}^{\infty} \frac{(\mu t)^i}{i!} e^{-\mu t} f_A(t) dt \quad (4.9.1)$$

Αφού οι πιθανότητες μετάβασης από την κατάσταση j αν προστεθούν δίνουν ένα, προκύπτει ότι:

$$p_{i,0} = \mathbf{1} - \sum_{j=0}^i \beta_j = \sum_{j=i+1}^{\infty} \beta_j$$

Το διάγραμμα πιθανοτήτων μετάβασης φαίνεται στο ακόλουθο σχήμα:



Εικόνα 21 Διάγραμμα πιθανοτήτων μετάβασης για το μοντέλο G/M/1

Οι περιοριστικές πιθανότητες α_n ικανοποιούν τις εξισώσεις ισορροπίας:

$$\begin{aligned} \alpha_0 &= \alpha_0 p_{0,0} + \alpha_1 p_{1,0} + \alpha_2 p_{2,0} + \dots \\ &= \sum_{t=0}^{\infty} \alpha_t p_{t,0} \quad (4.9.2) \end{aligned}$$

$$\begin{aligned} \alpha_n &= \alpha_{n-1} \beta_0 + \alpha_n \beta_1 + \alpha_{n+1} \beta_2 + \dots \\ &= \sum_{i=0}^{\infty} \alpha_{n-1} \beta_i \quad (4.9.3) \end{aligned}$$

Για να βρεθεί η λύση στις εξισώσεις ισορροπίας προκύπτει ότι η μέθοδος των παραγόμενων συναρτήσεων δεν μπορεί να εφαρμοστεί. Για το λόγο αυτό, υιοθετείται η άμεση προσέγγιση προσπαθώντας να βρεθούν λύσεις της μορφής:

$$\alpha_n = \sigma^n, n = 0, 1, 2, \dots \quad (4.9.4)$$

Η αντικατάσταση αυτής της μορφής στην (4.9.3) και διαιρώντας με την κοινή δύναμη σ^{n-1} έχουμε:

$$\sigma = \sum_{i=0}^{\infty} \sigma^i \beta_i$$

Αντικαθιστώντας τον τύπο του β_i από την (4.9.1) προκύπτει ότι:

$$\begin{aligned} \sigma &= \sum_{i=0}^{\infty} \sigma^i \int_{t=0}^{\infty} \frac{(\mu t)^i}{i!} e^{-\mu t} f_A(t) dt \\ &= \int_{t=0}^{\infty} e^{-(\mu - \mu \sigma)t} f_A(t) dt \end{aligned}$$

Το τελευταίο ολοκλήρωμα αναγνωρίζεται ως ο μετασχηματισμός Laplace-Stieltjes του χρόνου μεταξύ των αφίξεων. Συνεπώς καταλήγουμε στην ακόλουθη εξίσωση:

$$\sigma = \tilde{A}(\mu - \mu \sigma) \quad (4.9.5)$$

Απευθείας φαίνεται ότι $\sigma = 1$ είναι μια ρίζα της εξίσωσης (4.9.5), αφού $\tilde{A}(\mathbf{0}) = \mathbf{1}$. Όμως αυτή η ρίζα δεν είναι χρήσιμη, διότι πρέπει να μπορούμε να κανονικοποιήσουμε τη λύση των εξισώσεων ισορροπίας. Μπορεί να αποδειχθεί ότι όσο $\rho < 1$ η εξίσωση (4.9.5) έχει μια μοναδική ρίζα σ στο διάστημα $\mathbf{0} < \sigma < \mathbf{1}$, και αυτή είναι η ρίζα που αναζητούμε. Να σημειωθεί ότι απομένουσα εξίσωση (4.9.2) ικανοποιείται επίσης από την (4.9.4) αφού οι εξισώσεις ισορροπίας είναι εξαρτημένες. Τελικά, κανονικοποιούμε τη λύση (4.9.4) και προκύπτει:

$$\alpha_n = (\mathbf{1} - \sigma)\sigma^n, n = \mathbf{0}, \mathbf{1}, \mathbf{2}, \dots \quad (4.9.6)$$

Συνεπώς μπορούμε να καταλήξουμε ότι η κατανομή του μήκους της ουράς που μπορεί να βρεθεί ακριβώς πριν την άφιξη ενός πελάτη είναι *γεωμετρική* με παράμετρο σ , όπου σ είναι η μοναδική ρίζα της εξίσωσης (4.9.5) στο διάστημα $(0,1)$.

3.9.2. Κατανομή του χρόνου παραμονής

Αφού η κατανομή των αφίξεων είναι γεωμετρική, είναι εύκολο να καθοριστεί η κατανομή του χρόνου παραμονής. Για την ακρίβεια, η ανάλυση είναι παρόμοια με αυτή της ουράς M/M/1. Με πιθανότητα α_n ένας πελάτης που φθάνει στο σύστημα βρίσκει n πελάτες ήδη σε αυτό. Τότε ο χρόνος παραμονής του είναι το άθροισμα των $n+1$ εκθετικά κατανεμημένων χρόνων εξυπηρέτησης, κάθε ένας με μέσο $1/\mu$. Συνεπώς:

$$\begin{aligned} \tilde{S}(s) &= E(e^{-sS}) \\ &= \sum_{n=0}^{\infty} \alpha_n \left(\frac{\mu}{\mu + s} \right)^{n+1} \\ &= \frac{\mu(\mathbf{1} - \sigma)}{\mu(\mathbf{1} - \sigma) + s} \end{aligned}$$

3.9.3. Μέσος χρόνος παραμονής

Για έναν πελάτη που φθάνει στο σύστημα έχουμε:

$$E(S) = E(L^a) \frac{1}{\mu} + \frac{1}{\mu} \quad (4.9.7)$$

όπου η τυχαία μεταβλητή L^a συμβολίζει τον αριθμό των πελατών στο σύστημα που βρίσκει ο πελάτης κατά την άφιξη. Σύμφωνα με το νόμο του Little ισχύει ότι:

$$E(L) = \lambda E(S) \quad (4.9.8)$$

Επειδή δεν υπάρχουν αφίξεις Poisson ισχύει ότι:

$$E(L^a) \neq E(L)$$

Συνεπώς, η προσέγγιση της μέσης τιμής δεν μπορεί να εφαρμοστεί εδώ, διότι καταλήγουμε με δύο εξισώσεις για τρεις αγνώστους. Επιπλέον πληροφορίες αντλούνται από τη σχέση (4.9.6), δίνοντας:

$$E(L^a) = \sum_{n=0}^{\infty} n a_n = \sum_{n=0}^{\infty} n(1-\sigma)\sigma^n = \frac{\sigma}{1-\sigma}$$

Τότε προκύπτει από τις (4.9.7) και (4.9.8) ότι:

$$E(S) = \frac{\sigma}{1-\sigma} \frac{1}{\mu} + \frac{1}{\mu} = \frac{1}{(1-\sigma)\mu}$$

και

$$E(L) = \frac{\lambda}{(1-\sigma)\mu} = \frac{\rho}{1-\sigma}$$

Κεφάλαιο 4: Εφαρμογές των συστημάτων εξυπηρέτησης

Τα συστήματα εξυπηρέτησης έχουν πολλές και σημαντικές εφαρμογές σε πολλούς τομείς (Morgan, 1999). Επιγραμματικά, κάποιες από τις εφαρμογές αυτές είναι:

- Εμπορικά συστήματα εξυπηρέτησης
 - Εξυπηρέτηση εξωτερικών πελατών
 - § πρατήρια βενζίνης
 - § ταμεία τράπεζας
 - § ταμεία κοινωφελών οργανισμών
 - § καταστήματα
 - § αυτόματοι πωλητές
 - § ATM
 - § ταχυδρομεία
 - § supermarket
- Μεταφορικά μέσα εξυπηρέτησης
 - Τα μεταφορικά μέσα ως πελάτες
 - § σταθμοί διοδίων
 - § σηματοδότες
 - § πλοία σε προβλήτες
 - § φορτηγά που αναμένουν για φορτοεκφόρτωση
 - § αεροσκάφη που αναμένουν για απογείωση ή προσγείωση
 - Τα μεταφορικά μέσα ως θέσεις εξυπηρέτησης
 - § επιβάτες που αναμένουν ταξί
 - § λεωφορεία
 - § πυροσβεστικά οχήματα
 - § ασθενοφόρα

- § ανελκυστήρες
 - § αεροπορικές μεταφορές
- Εμποροβιομηχανικά συστήματα εξυπηρέτησης
 - ο Εξυπηρέτηση εσωτερικών πελατών
 - § διακίνηση υλικών σε παραγωγικές διαδικασίες
 - § συντήρηση μηχανών
 - § σταθμοί επιθεώρησης
 - § συστήματα H/Y
 - § συστήματα εξυπηρέτησης εργαζομένων (αποθήκες εξοπλισμού, εστιατόρια, κάρτες πρόσβασης κλπ)
 - § συστήματα όπου μηχανές αποτελούν θέσεις εξυπηρέτησης
- Κοινωνικά συστήματα εξυπηρέτησης
 - ο Εξυπηρέτηση ευρύτερου κοινωνικού συνόλου άμεσα ή έμμεσα
 - § σύστημα δικαιοσύνης
 - § νομοθετικό σύστημα
 - § συστήματα υγείας
 - § κοινωνικές υπηρεσίες (στεγαστική πρόνοια, υποστήριξη ανέργων κλπ)
 - § παιδεία
 - § δημόσια έργα

Στο κεφάλαιο αυτό θα αναλυθούν κάποιες αντιπροσωπευτικές εφαρμογές των συστημάτων εξυπηρέτησης.

4.1. Χρονοπρογραμματισμός υπολογιστών

Υπάρχουν πολλές διαφορετικές στρατηγικές που χρησιμοποιούνται για τον χρονοπρογραμματισμό των ηλεκτρονικών υπολογιστών. Σε αυτή την παράγραφο αναφέρεται μια απλή στρατηγική, χωρίς περιθώρια για προτεραιότητα ανάμεσα στις διεργασίες (Yue et al., 2009). Αυτός είναι ο *αλγόριθμος της μαζικής επεξεργασίας* (*batch processing algorithm*) και έχει τις ακόλουθες ιδιότητες:

- Κάθε διεργασία λαμβάνει ακριβώς όσο χρόνο χρειάζεται για να ολοκληρωθεί
- Υπάρχει μια μοναδική ουρά FIFO που αναμένει επεξεργασία
- Νέες διεργασίες τοποθετούνται στο τέλος της ουράς

Αυτό το μοντέλο μπορεί να θεωρηθεί ως μια περίπτωση ουράς M/G/1.

Είναι εμφανές ότι ο μέσος χρόνος απόκρισης του συστήματος για μια διεργασία που χρειάζεται x δευτερόλεπτα για να ολοκληρωθεί, $T(x)$, θα είναι ο χρόνος που περνάει αναμένοντας στην ουρά, συν το χρόνο για την ολοκλήρωση της διεργασίας:

$$T(x) = \frac{W_0}{1 - \rho} + x \quad (5.1.1)$$

Για την ουρά M/G/1 ισχύει:

$$W_0 = \frac{\lambda \bar{x}^2}{2} \quad (5.1.2)$$

Παρατηρούμε ότι ο χρόνος αναμονής είναι πλήρως ανεξάρτητος από το μήκος της εργασίας. Συνεπώς, η μαζική επεξεργασία δεν διακρίνει τις εργασίες με βάση το μέγεθός τους. Αυτό αποτελεί μειονέκτημα, διότι οι μικρές εργασίες δεν λαμβάνουν προτεραιότητα έναντι των μεγάλων και για το λόγο αυτό η μαζική επεξεργασία δεν είναι κατάλληλη για ένα χρονο-μοιραζόμενο σύστημα.

Από τις σχέσεις (5.1.1) και (5.1.2) λαμβάνουμε την τελική σχέση:

$$T(x) = \frac{\lambda \bar{x}^2}{2(1 - \lambda \bar{x})} + x \quad (5.1.3)$$

4.2. Ουρές σε τράπεζα

4.2.1. Εισαγωγικά

Κατά κανόνα οι άνθρωποι απεχθάνονται τις ουρές στις τράπεζες. Για να βελτιώσει το επίπεδο των υπηρεσιών της τράπεζας, ο διευθυντής ενδιαφέρεται να βρει:

1. Το μέσο αριθμό ανθρώπων που περιμένουν στην τράπεζα, δηλαδή το μήκος της ουράς
2. Το ποσοστό του χρόνου στον οποίο ένας υπάλληλος δεν είναι απασχολημένος

Για την καλύτερη κατανόηση της περίπτωσης, χρησιμοποιούμε ένα αριθμητικό παράδειγμα θεωρώντας ότι ο αριθμός των υπαλλήλων είναι από έναν έως πέντε.

4.2.2. Υποθέσεις και λεπτομέρειες του μοντέλου

- Η κατανομή του χρόνου που χρειάζεται ένας υπάλληλος για να διεκπεραιώσει μία εργασία είναι εκθετική, με μέση τιμή 2 λεπτά και τυπική απόκλιση 5/4 λεπτά
- Εικονικά δεν υπάρχει περιορισμός στο μήκος της ουράς, δεδομένου ότι η τράπεζα έχει αρκετό χώρο αναμονής
- Οι πελάτες φθάνουν με κατανομή Poisson, με μέση τιμή 25 πελάτες ανά ώρα
- Η εξυπηρέτηση γίνεται σύμφωνα με ουρά FCFS
- Το μοντέλο είναι ουρά M/M/c, όπου $1 \leq c \leq 5$

4.2.3. Σύμβολα

Στον παρακάτω πίνακα φαίνονται τα σύμβολα που χρησιμοποιούνται, καθώς και οι τιμές τους αν υπάρχουν από την υπόθεση:

Σύμβολα	Επεξήγηση και τιμή (εάν δίνεται)
c	Αριθμός υπαλλήλων (εξυπηρετητών)
L_q	Μέσο μήκος της ουράς
W	Ποσοστό του χρόνου του υπαλλήλου για το οποίο δεν είναι απασχολημένος
λ	Μέσος ρυθμός αφίξεων (25 ανά ώρα)
μ	1/μέσος χρόνος εξυπηρέτησης πελατών (30 ανά ώρα)
p	Μέση εργασία για κάθε εξυπηρετητή ανά ώρα

4.2.4. Υπολογισμοί

Από τις παραπάνω υποθέσεις μπορούμε να υπολογίσουμε τα μεγέθη:

$$r = \frac{\lambda}{\mu} = \frac{25}{30} = \frac{5}{6}$$

$$p = \frac{\lambda}{c\mu} = \frac{r}{c} = \frac{5}{6c}$$

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!(1-p)} \right)^{-1}$$

$$L_q = \frac{r^c p p_0}{c!(1-p)^2} = \frac{5^{c+1} p_0}{6^{c+1} c! c \left(1 - \frac{5}{6c}\right)^2}$$

$$W = \frac{1}{\mu} + \frac{L_q}{\lambda} = \frac{1}{30} + \frac{L_q}{25}$$

$$\text{Χρόνος αδράνειας} = 1 - p = 1 - \frac{5}{6c}$$

Τα αποτελέσματα των εξισώσεων αυτών για έναν έως πέντε υπαλλήλους είναι:

Υπάλληλοι	p	$r^n / n!$	p_0	L_q	W	Αδράνεια
-----------	---	------------	-------	-------	---	----------

1	0.833	0.833	0.167	4.167	0.2	0.167
2	0.417	0.347	0.412	0.175	0.04	0.583
3	0.278	0.096	0.432	0.022	0.034	0.722
4	0.208	0.02	0.434	0.003	0.033	0.792
5	0.167	0.003	0.435	0.000	0.033	0.833

Από τον παραπάνω πίνακα φαίνεται ότι ο διευθυντής μπορεί να προσλάβει έναν ή δυο υπαλλήλους ώστε να έχει ικανοποιητικά αποτελέσματα. Από τρεις υπαλλήλους και πάνω το μέσω μήκος της ουράς L_q μειώνεται όμως αυξάνεται ο χρόνος αδράνειας. Έτσι για να επιτευχθεί μία μέση λύση χωρίς να αυξάνεται το μήκος της ουράς και ο χρόνος αδράνειας των υπαλλήλων επιλέγονται μέχρι δύο υπάλληλοι.

4.3. Ουρές που δημιουργούνται κατά την ώρα αιχμής

4.3.1. Εισαγωγικά

Σε αυτό το παράδειγμα φαίνεται πώς είναι δυνατό να μοντελοποιηθεί η συγκέντρωση οχημάτων κατά την ώρα αιχμής, και συνεπώς να φανεί πότε σχηματίζονται ουρές σε έναν ίσιο δρόμο χωρίς διασταυρώσεις, ανάλογα με το ρυθμό αφίξεων των αυτοκινήτων.

4.3.2. Σύμβολα

Σύμβολο	Ερμηνεία
Q(t)	Μήκος ουράς την στιγμή t
A(t)	Συγκεντρωτικές αφίξεις στην ουρά τη στιγμή t
D(t)	Συγκεντρωτικές αναχωρήσεις από την ουρά τη στιγμή t
μ	Μέγιστος ρυθμός εξυπηρέτησης
μ(t)	Ρυθμός εξυπηρέτησης τη στιγμή t
λ(t)	Ρυθμός αφίξεων τη στιγμή t

4.3.3. Μοντελοποίηση της ουράς

Την ώρα αιχμής ο ρυθμός άφιξης, $\lambda(t)$, θα αυξηθεί μέχρι να φτάσει το μέγιστο ρυθμό, δηλαδή τη στιγμή που οι περισσότεροι άνθρωποι θα βρίσκονται στο δρόμο για να

πάνε στις δουλειές τους. Μετά τη στιγμή αυτή ο χρόνος αύξησης θα μειωθεί. Υπάρχουν τέσσερα βασικά σημεία σε αυτή τη ροή, όπως φαίνεται παρακάτω:

Χρονική στιγμή	Σημασία
t_0	Το πρώτο σημείο στο οποίο το $\lambda(t)$ ισούται με το μέγιστο ρυθμό εξυπηρέτησης μ
t_1	Το σημείο όπου το $\lambda(t)$ είναι μέγιστο
t_2	Το σημείο όπου το $\lambda(t)$ θα μειωθεί πίσω στο μ
t_3	Το σημείο όπου η ουρά έχει εξαλειφθεί και άρα ισχύει $\mu(t)=\lambda(t)$

Το $\lambda(t)$ έχει ανάπτυγμα Taylor στο σημείο t_1 :

$$\lambda(t) = \lambda(t_1) - \beta(t - t_1)^2$$

με σταθερά β :

$$\beta = -\frac{1}{2} \frac{d^2\lambda(t)}{dt^2} \Big|_{t=t_1}$$

Ισχύει για μικρά $t - t_1$, δηλαδή είναι έγκυρο στο διάστημα $t_0 < t < t_3$.

Αυτό σημαίνει ότι οι στιγμές t_0 και t_2 μπορούν να υπολογιστούν με αντικατάσταση:

$$\mu = \lambda(t_0) = \lambda(t_1) - \beta(t_0 - t_1)^2$$

$$t_0 = t_1 - \left(\frac{\lambda(t_1) - \mu}{\beta} \right)^2$$

$$t_2 = t_1 + \left(\frac{\lambda(t_1) - \mu}{\beta} \right)^2$$

Το μήκος της ουράς είναι η περιοχή ανάμεσα στις καμπύλες των $\lambda(t)$ και $\mu(t)$, δηλαδή:

$$Q(t) = A(t) - D(t) = \int_{t_0}^t [\lambda(t) - \mu(t)] dt$$

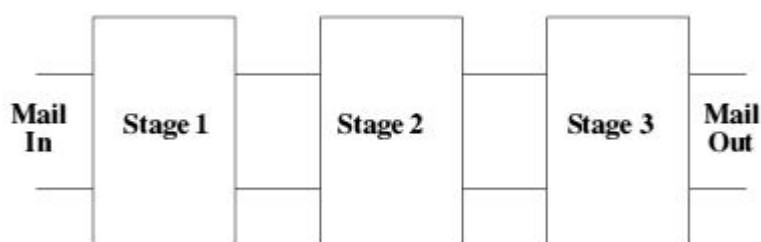
Με χρήση απλών αντικαταστάσεων παρόμοιων με των προηγούμενων, η σχέση μπορεί να απλοποιηθεί:

$$Q(t) = \int_{t_0}^t [\lambda(t) - \mu(t)] dt = \beta(t - t_0)^2 \left(\frac{t_2 - t_0}{2} - \frac{t - t_0}{3} \right)$$

4.4. Γραφείο ταξινόμησης γραμμάτων

4.4.1. Εισαγωγικά

Σε ένα γραφείο ταξινόμησης γραμμάτων συχνά τα γράμματα που πρέπει να ταξινομηθούν έρχονται μαζικά και είναι σημαντικό οι υπάλληλοι να οργανωθούν καλά με τέτοιο τρόπο ώστε να τα ταξινομούν όσο πιο γρήγορα γίνεται. Πολλές εταιρίες που δραστηριοποιούνται σε αυτό τον τομέα συνήθιζαν να μεταφέρουν τους υπαλλήλους τους μαζί με τα γράμματα, δηλαδή όλοι οι υπάλληλοι αρχικά βοηθούσαν στην αρχή της γραμμής, για να μειώσουν το απόθεμα, και σταδιακά μεταφέρονταν κατά μήκος της γραμμής, όσο το απόθεμα κινούταν μαζί της. Ωστόσο η θεωρία των συστημάτων εξυπηρέτησης μπορεί να χρησιμοποιηθεί για να εμφανίσει ένα πιο γρήγορο τρόπο για την ταξινόμηση των γραμμάτων. Για αυτή την εφαρμογή το επίπεδο (stage) αναφέρεται στο επίπεδο της ταξινόμησης, και ο εξυπηρετητής (server) στον υπάλληλο που ταξινομεί τα γράμματα σε κάθε στάδιο.



Εικόνα 22 Επίπεδα ταξινόμησης γραμμάτων

4.4.2. Υποθέσεις και απλοποιήσεις

Γίνονται οι ακόλουθες υποθέσεις για απλοποίηση:

1. Υπάρχει σταθερός αριθμός υπαλλήλων
2. Όλοι οι υπάλληλοι ταξινομούν με σταθερό, συνεχή τρόπο που είναι ανεξάρτητος από το στάδιο στο οποίο βρίσκονται
3. Η συγκεντρωτική είσοδος στο πρώτο από τα n στάδια καταγράφεται
4. Ο χρόνος μεταφοράς των γραμμάτων από και προς τα στάδια αγνοείται, δηλαδή ο χρόνος για τον οποίο τα γράμματα βρίσκονται στο σύστημα εξαρτάται από την αναμονή ανάμεσα στα στάδια
5. Ο αριθμός των γραμμάτων στο σύστημα δεν επηρεάζει την απόδοσή του
6. Υπάρχουν τόσοι πολλοί υπάλληλοι που μπορούμε να αγνοήσουμε ότι είναι διακριτοί εξυπηρετητές

4.4.3. Σύμβολα

Χρησιμοποιούνται τα ακόλουθα σύμβολα:

Σύμβολο	Ερμηνεία
n	Ο αριθμός των σταδίων, τα οποία είναι αριθμημένα από το 1 μέχρι το n
A_k(t)	Συγκεντρωτικός αριθμός γραμμάτων που έχουν φτάσει στο k-οστό στάδιο τη στιγμή t
D_{ks}(t)	Συγκεντρωτικός αριθμός γραμμάτων που έχουν αναχωρήσει από το k-οστό στάδιο για την ουρά του επόμενου τη στιγμή t
D_{kq}(t)	Συγκεντρωτικός αριθμός γραμμάτων που έχουν αναχωρήσει από την ουρά για το k-οστό στάδιο τη στιγμή t
D_k(t)	Συγκεντρωτικός αριθμός γραμμάτων που έχουν αναχωρήσει από το k-οστό στάδιο τη στιγμή t. Είναι ισοδύναμο με το D _{kq} (t) διότι όλα τα γράμματα που μπαίνουν σε έναν εξυπηρετητή πρέπει επίσης να τον αφήσουν και από τις υποθέσεις μας, τα γράμματα μένουν αμελητέο χρόνο στον εξυπηρετητή
μ	Ο ρυθμός εξυπηρέτησης για όλους τους εξυπηρετητές συγκεντρωτικά
μ_k	Ο ρυθμός εξυπηρέτησης για το k-οστό στάδιο

4.4.4. Τοποθέτηση των εργαζομένων

Από το γεγονός ότι μ_k είναι ο ρυθμός εξυπηρέτησης για το k-οστό στάδιο ισχύει:

$$\sum_{j=1}^n \mu_j(t) = \mu$$

Χρειάζεται είτε:

- να ελαχιστοποιήσουμε τη συνολική καθυστέρηση στο σύστημα, δηλαδή κάθε γράμμα να μένει στο σύστημα για ελάχιστο χρόνο ή

- b. να μεγιστοποιήσουμε τον αριθμό των γραμμάτων που έχουν αναχωρήσει από το σύστημα μια δεδομένη χρονική στιγμή.

Και οι δύο αυτοί στόχοι μπορούν να επιτευχθούν με τη μεγιστοποίηση του $D_n(t)$. Αυτό περιορίζεται από τις τιμές του $D_k(t)$ για τα προηγούμενα στάδια, αφού το τελικό αποτέλεσμα δεν μπορεί να είναι περισσότερο από το αποτέλεσμα για τα προηγούμενα στάδια. Συνεπώς:

$$D_n(t) \leq D_{n-1}(t) \leq \dots \leq D_1(t) \leq A_1(t)$$

Για να επιτευχθεί είτε το a είναι το b πρέπει να μεγιστοποιήσουμε το $D_n(t)$, και έτσι:

$$D_n(t) = D_{n-1}(t) = \dots = D_1(t) = A_1(t)$$

Αφού όλα τα στάδια πρέπει να έχουν το υψηλότερο δυνατό ρυθμό εξυπηρέτησης, ο οποίος είναι μ/n , αυτό θα έδινε το υψηλότερο $D_k(t)$ για κάθε k . Δεν πρέπει να ισχύει $\mu(k-1) > \mu k$, διότι αυτό θα δημιουργήσει ουρά ανάμεσα στα στάδια $k-1$ και k , και ο επιπλέον ρυθμός εξυπηρέτησης που χρησιμοποιήθηκε για να δημιουργήσει την ουρά θα μπορούσε να χρησιμοποιηθεί καλύτερα στο επόμενο στάδιο για την αποφυγή δημιουργίας ουράς εξ αρχής.

4.4.5. Περίληψη των αποτελεσμάτων

1. Είναι καλύτερη η κατανομή των υπαλλήλων του ταχυδρομείου ανάμεσα σε όλα τα στάδια
2. Δεν είναι ποτέ καλή η ύπαρξη ενός μk μεγαλύτερου από το επόμενο στάδιο.

4.5. Ουρές που δημιουργούνται σε διασταύρωση δρόμων

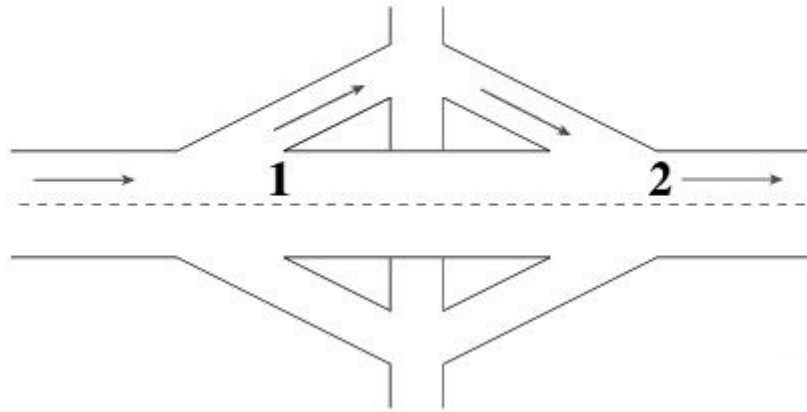
4.5.1. Εισαγωγικά

Αυτό το παράδειγμα αναφέρεται στον τρόπο με τον οποίο μια ουρά που προκαλείται από κίνηση για την είσοδο σε ένα δρόμο μπορεί να προκαλέσει μια δευτερεύουσα ουρά από αυτοκίνητα που προσπαθούν να φύγουν από τον αυτοκινητόδρομο.

Επειδή ενδιαφερόμαστε στο πώς η ουρά για την είσοδο στο δρόμο επηρεάζει την ουρά για την έξοδο από αυτόν, θεωρούμε ότι υπάρχουν στην πραγματικότητα δύο περιοχές στην ουρά που δημιουργείται στο σημείο 2:

1. Η κίνηση που συνεχίζει στο δρόμο πέρα από το σημείο 2

2. Η κίνηση που προσπαθεί να φύγει από το δρόμο στο σημείο 1, αλλά δεν μπορεί λόγω του ότι έχει μπλοκάρει από τη συνεχόμενη κίνηση. Αυτή είναι η δευτερεύουσα ουρά.



Εικόνα 23 Σημεία συμφόρησης στη δημιουργία ουρών στο δρόμο

4.5.2. Υποθέσεις

- Η συμφόρηση πραγματοποιείται στην περιοχή αμέσως μετά από το σημείο συνένωσης.
- Ο δρόμος εισόδου στον κεντρικό δρόμο δεν είναι σημείο συμφόρησης, δηλαδή δεν υπάρχει καθυστέρηση στο δρόμο αυτό
- Ο χρόνος μετάβασης από το σημείο 1 στο σημείο 2 είναι σταθερός, ανεξάρτητα από την κατάσταση της κίνησης. Αυτή είναι μια απαραίτητη υπόθεση ώστε να αποφευχθούν περίπλοκες εκφράσεις.

4.5.3. Σύμβολα

Σύμβολο	Ερμηνεία
m	Αριθμός λωρίδων στο δρόμο
$A_{1m}(t)$	Συγκεντρωτικός αριθμός των μελλοντικών οχημάτων που θα περνούσαν από το σημείο 1 εάν δεν υπήρχε κίνηση
$A_{1s}(t)$	Συγκεντρωτικός αριθμός των υπαρχόντων οχημάτων που θα περνούσαν από το σημείο 1 εάν δεν υπήρχε κίνηση
$A_{2s}(t)$	Συγκεντρωτικός αριθμός των οχημάτων από

	τον δρόμο εξόδου που θα περνούσαν από το σημείο 2 εάν δεν υπήρχε κίνηση
$A_{2m}(t)$	Συγκεντρωτικός αριθμός των οχημάτων του δρόμου που θα περνούσαν από το σημείο 2 εάν δεν υπήρχε κίνηση
$A_2(t)$	Συγκεντρωτικός αριθμός των οχημάτων που θα έφταναν στο σημείο 2 σε χρόνο t εάν δεν υπήρχε κίνηση
$D_{1m}(t)$	Συγκεντρωτικός αριθμός των μελλοντικών οχημάτων που πέρασαν από το σημείο 1
$D_2(t)$	Συγκεντρωτικός αριθμός των οχημάτων που πέρασαν από το σημείο 2
$D_{2s}(t)$	Συγκεντρωτικός αριθμός των οχημάτων που πέρασαν από το σημείο 2 και προέρχονται από το δρόμο εξόδου
$D_{2m}(t)$	Συγκεντρωτικός αριθμός των οχημάτων που πέρασαν από το σημείο 2 και προέρχονται από τον κεντρικό δρόμο
μ	Μέγιστος ρυθμός εξυπηρέτησης του δρόμου στο σημείο συνένωσης, δηλαδή ο μέγιστος αριθμός αυτοκινήτων ανά ώρα που μπορούν να περάσουν το σημείο 2
$\mu(t)$	Ο ρυθμός εξυπηρέτησης στο σημείο 2 τη στιγμή t
μ_s	Ο μέγιστος ρυθμός εξυπηρέτησης στο δρόμο εισόδου
$\mu_s(t)$	Ο μέγιστος ρυθμός εξυπηρέτησης στο δρόμο εισόδου τη στιγμή t
t	Ο χρόνος για τη μετάβαση από το σημείο 1 στο σημείο 2 όταν δεν υπάρχει κίνηση
c	Η χωρητικότητα του δρόμου ανάμεσα στα σημεία 1 και 2, δηλαδή πόσα οχήματα μπορούν να δημιουργήσουν ουρά σε αυτό το σημείο του δρόμου

4.5.4. Μοντελοποίηση της ροής της κίνησης

Η κίνηση στο δρόμο μπορεί να προκληθεί είτε από το δρόμο εισόδου είτε από τον κεντρικό. Συνεπώς μπορούμε να πούμε ότι:

$$A_2(t) = A_{2s}(t) + A_{2m}(t)$$

Με μια πρώτη ματιά ο αριθμός των αυτοκινήτων (και συνεπώς το μήκος της ουράς) φαίνεται να είναι $A_2(t) - D_2(t)$. Ωστόσο, αυτό δεν ισχύει αυστηρά διότι το μήκος της ουράς είναι αυτό, συν ο αριθμός των αυτοκινήτων που θα μπορούσαν να είναι στην περιοχή που καταλαμβάνεται από την ουρά, εάν αυτή δεν υπήρχε. Συνεπώς η περιοχή ανάμεσα στο $D_2(t)$ και $A_2(t)$ στο γράφο της συγκεντρωτικής μέτρησης συναρτήσεως του χρόνου δίνει τη συνολική καθυστέρηση για όλα τα οχήματα που περνούν από το σημείο 2.

Οι περισσότερες διασταυρώσεις δρόμων είναι σχεδιασμένες ώστε εάν υπάρχει μεγάλη καθυστέρηση η κίνηση στο δρόμο εισόδου να εισέρχεται στον αυτοκινητόδρομο εναλλακτικά με την κεντρική κίνηση μέσω της αργής λωρίδας. Συνεπώς:

$$\mu_s = \frac{\mu}{2}$$

και ο ρυθμός εξυπηρέτησης στον αυτοκινητόδρομο για την κεντρική κίνηση είναι $\mu - \mu_s$. Αυτό επίσης σημαίνει ότι η κεντρική κίνηση στον αυτοκινητόδρομο απασχολεί $m-1/2$ λωρίδες σε αυτόν.

Αν η κεντρική κίνηση στον αυτοκινητόδρομο δε χρησιμοποιεί όλο το ρυθμό εξυπηρέτησης τότε η κίνηση του δρόμου εισόδου μπορεί να χρησιμοποιήσει την καθυστέρηση στο σύστημα (έως το μέγιστο όριο του δρόμου εισόδου) και συνεπώς η κίνηση της διασταύρωσης λαμβάνει προτεραιότητα και αντίστροφα, όταν η κίνηση του δρόμου εισόδου δε χρησιμοποιεί όλο το ρυθμό εξυπηρέτησής της τότε η κεντρική κίνηση του αυτοκινητόδρομου μπορεί να χρησιμοποιήσει το περίσσειμα και να λάβει προτεραιότητα. Συχνά η ουρά δημιουργείται στον αυτοκινητόδρομο, αλλά ο δρόμος εισόδου παραμένει ελεύθερος από κίνηση, όσο λειτουργεί υπό την πλήρη χωρητικότητά του μ_s .

Μπορεί να θεωρηθεί ότι:

- $D_{2m}(t) = D_2(t) - D_{2s}(t)$, αφού τα οχήματα που βγαίνουν από το σημείο 2 πρέπει να προέρχονται είτε από τον αυτοκινητόδρομο είτε από το δρόμο εισόδου
- $D_{1m}(t) = D_{2m}(t) + c$

- $A_{1m}(t) = A_{2m}(t + \tau)$, διότι εάν δεν υπήρχε ουρά ο αριθμός των οχημάτων που φθάνουν στο σημείο 2 θα έπρεπε να ήταν ίδιος με τον αριθμό των οχημάτων που φθάνουν στο σημείο 1, τ μονάδες χρόνου πριν, και που θα συνέχιζαν προς το σημείο 2

Εάν η ουρά δεν έχει φτάσει ακόμα το σημείο 1 τότε $A_{1m}(t) - D_{2m}(t)$ είναι ο αριθμός των οχημάτων ανάμεσα στα σημεία 1 και 2, είτε σε μια ουρά είτε εν κινήσει κατά μήκος του αυτοκινητοδρόμου. Συνεπώς $A_{1m}(t) - D_{2m}(t) - c$ είναι ο αριθμός των διερχόμενων αυτοκινήτων που έχουν καθυστερήσει και δεν έχουν φτάσει ακόμα στο σημείο 1.

Εάν υποθέσουμε ότι δεν υπάρχει μια λωρίδα που προορίζεται αποκλειστικά για τη χρήση της εξερχόμενης κίνησης τότε η κίνηση αυτή θα πρέπει να μοιραστεί μια λωρίδα μαζί με τη διερχόμενη κίνηση. Έτσι η καθυστέρηση για την εξερχόμενη κίνηση στο σημείο 1 σε μία χρονική στιγμή T θα είναι η ίδια για ένα διερχόμενο όχημα που φθάνει στο σημείο 1 τη στιγμή T .

Η καθυστέρηση για όλα τα διερχόμενα οχήματα θα είναι ίση με την περιοχή ανάμεσα στην καμπύλη του $A_{1m}(t)$ (η οποία ισούται με $A_{2m}(t + \tau)$) και του $D_{1m}(t)$ (η οποία ισούται με $D_{2m}(t) + c$). Συνεπώς η καθυστέρηση για ένα μοναδικό όχημα είναι το παρακάτω ολοκλήρωμα:

$$\text{Καθυστερήση ανά όχημα} = \frac{\int_0^T [A_{2m}(t + \tau)] dt - \int_0^T [D_{2m}(t + \tau) + c] dt}{A_{1m}(T) - D_{2m}(T)}$$

Για να μην υπάρχει επίδραση στην κίνηση που βγαίνει από τον αυτοκινητόδρομο θα έπρεπε να υπάρχει μια ειδική λωρίδα για την εξερχόμενη κίνηση στο σημείο 1. Στην περίπτωση αυτή:

- Η διασταύρωση εξόδου δε θα επηρέαζε το χρόνο καθυστέρησης της ουράς
- Η ουρά θα ήταν μεγαλύτερη και η ταχύτητα της κίνησης σε αυτή θα ήταν υψηλότερη
- Η ουρά δε θα επηρέαζε την έξοδο από τον αυτοκινητόδρομο

Αυτή η λύση χρησιμοποιείται σε πολλά σημεία αυτοκινητοδρόμων, όπου υπάρχει αρκετός χώρος.

Κεφάλαιο 5: Υλοποίηση τραπεζικού συστήματος εξυπηρέτησης

5.1. Περιγραφή

Στο κεφάλαιο αυτό γίνεται η παρουσίαση του προγραμματιστικού κώδικα σε γλώσσα προγραμματισμού Java για την υλοποίηση ενός τραπεζικού συστήματος εξυπηρέτησης. Θεωρούμε ότι υπάρχουν τρεις κατηγορίες πελατών: οι κανονικοί, οι VIP και οι εξπρές. Κάθε κατηγορία πελάτη εξυπηρετείται με διαφορετικό τρόπο από την τράπεζα. Με την είσοδο του κάθε πελάτη στην τράπεζα του χορηγείται ένα εισιτήριο από το μηχάνημα έκδοσης εισιτηρίων ανάλογα με τον τύπο του. Ο πελάτης στη συνέχεια εισέρχεται στην ουρά αναμονής. Έχοντας υπόψη τον τύπο του πελάτη γίνεται η εξυπηρέτησή του στον γκισέ. Αφού γίνει αυτό του αφαιρείται το εισιτήριο και εξέρχεται από την τράπεζα.

5.2. Κλάσεις

5.2.1. BankQueue

Η κλάση αυτή αποτελεί την κύρια είσοδο του προγράμματος. Σε αυτή την υλοποίηση δημιουργούνται τέσσερις κανονικές υπηρεσίες, μια υπηρεσία εξπρές και μια VIP υπηρεσία. Κάθε ένα δευτερόλεπτο δημιουργείται ένας κανονικός πελάτης που προστίθεται στην ουρά, κάθε 6 δευτερόλεπτα δημιουργείται ένας VIP πελάτης και κάθε 3 δευτερόλεπτα δημιουργείται ένας πελάτης εξπρές.

```
package org.epdo.sim;

import java.util.concurrent.Executors;
import java.util.concurrent.ScheduledExecutorService;
import java.util.concurrent.TimeUnit;

public class BankQueue {
    public static void main(String args[]) {
        // Generate 4 normal service windows.
        ServiceWindow normalSW1 = new ServiceWindow();
        normalSW1.setType(TicketType.COMMON);
        Thread normalTd1 = new Thread(normalSW1);
        ServiceWindow normalSW2 = new ServiceWindow();
        normalSW2.setType(TicketType.COMMON);
        Thread normalTd2 = new Thread(normalSW2);
        ServiceWindow normalSW3 = new ServiceWindow();
        normalSW3.setType(TicketType.COMMON);
        Thread normalTd3 = new Thread(normalSW3);
```



```

ServiceWindow normalSW4 = new ServiceWindow();
normalSW4.setType(TicketType.COMMON);
Thread normalTd4 = new Thread(normalSW4);

// Generate one express service window.
ServiceWindow expressSW = new ServiceWindow();
expressSW.setType(TicketType.EXPRESS);
Thread expressTd = new Thread(expressSW);

// Generate one VIP service window.
ServiceWindow vipSW = new ServiceWindow();
vipSW.setType(TicketType.VIP);
Thread vipTd = new Thread(vipSW);

// Set the Service window Thread names. The names will be
printed out in
// console.
normalTd1.setName("Normal TD1");
normalTd1.start();
normalTd2.setName("Normal TD2");
normalTd2.start();
normalTd3.setName("Normal TD3");
normalTd3.start();
normalTd4.setName("Normal TD4");
normalTd4.start();

expressTd.setName("express TD");
expressTd.start();

vipTd.setName("VIP TD");
vipTd.start();

// Generate a normal customer every second.
ScheduledExecutorService commonCustomers = Executors
    .newScheduledThreadPool(1);
commonCustomers.scheduleAtFixedRate(new Runnable() {
    public void run() {
        TicketMachine tm = TicketMachine.getInstance();
        Ticket t = tm.generateTicket(TicketType.COMMON);
        System.out.println("Common customer " +
t.getSequenceNo()
                                + " is waiting.");
    }
}, 0, Constants.TIME_INTERVAL, TimeUnit.SECONDS);

// Generate a express customer every 3 seconds.
ScheduledExecutorService expressCustomers = Executors
    .newScheduledThreadPool(1);
expressCustomers.scheduleAtFixedRate(new Runnable() {
    public void run() {
        TicketMachine tm = TicketMachine.getInstance();
        Ticket t = tm.generateTicket(TicketType.EXPRESS);
        System.out.println("Express customer " +
t.getSequenceNo()
                                + " is waiting.");
    }
}, 0, Constants.TIME_INTERVAL * 3, TimeUnit.SECONDS);

// Generate a VIP customer every 6 seconds.

```

```

        ScheduledExecutorService VIPCustomers = Executors
            .newScheduledThreadPool(1);
        VIPCustomers.scheduleAtFixedRate(new Runnable() {
            public void run() {
                TicketMachine tm = TicketMachine.getInstance();
                Ticket t = tm.generateTicket(TicketType.VIP);
                System.out.println("VIP customer " +
t.getSequenceNo()
                    + " is waiting.");
            }
        }, 0, Constants.TIME_INTERVAL * 6, TimeUnit.SECONDS);
    }
}

```

5.2.2. Constants

Σε αυτήν την κλάση ορίζονται οι σταθερές του συστήματος.

```

package org.epdo.sim;

public class Constants {
    public static int MAX_PROCESSTIME = 5000; // takes 5 seconds maximum
to
    // process a customer.
    public static int MIN_PROCESSTIME = 1000; // takes 1 second minimum
to
    // process a customer.

    // the time interval of generating a customer.
    public static int TIME_INTERVAL = 1;
}

```

5.2.3. ServiceWindow

Η κλάση αυτή υλοποιεί την υπηρεσία της τράπεζας, την κύρια επιχειρηματική λογική. Η υπηρεσία θα εξυπηρετήσει πρώτα τους VIP και τους εξπρές πελάτες της. Αν δεν υπάρχουν VIP και εξπρές πελάτες θα εξυπηρετηθούν οι κανονικοί πελάτες.

```

package org.epdo.sim;

import java.util.Random;

public class ServiceWindow implements Runnable {
    // Service Window type is the same as ticket type. Use ticket type
as
    // service window type.
    private TicketType wType;
    // Ticket Machine is used to get the first customer waiting in the
queue.
    private TicketMachine tm = TicketMachine.getInstance();

    // Process customer who is in the queue.
    public void processCustomer(Ticket t) {

```

```

        System.out.println(Thread.currentThread().getName()
            + "start to process ticket:" + t.getSequenceNo()
            + ". Ticket type:" + t.getTicketType());
        if (t.getTicketType() != null) { // customer must have a ticket
with
                                                    // ticket
type.
            // Process common customers
            if (t.getTicketType().equals(TicketType.COMMON)) {
                try {
second to 5 seconds
                    // It will take a random time between 1
                    // to process normal customer.
                    Random r = new Random();
                    int random =
r.nextInt(Constants.MAX_PROCESSTIME
                        - Constants.MIN_PROCESSTIME);

                    System.out.println(Thread.currentThread().getName()
                        + " process " +
t.getSequenceNo() + " for "
                        + random + " milliseconds.");
                    Thread.sleep(random);
                } catch (InterruptedException e) {
                    // TODO Auto-generated catch block
                    e.printStackTrace();
                }
            }

            // process express customers
            if (t.getTicketType().equals(TicketType.EXPRESS)) {
express customer.
                try {
                    // It will take 1 second to process
                    Thread.sleep(Constants.MIN_PROCESSTIME);

                    System.out.println(Thread.currentThread().getName()
                        + " process " +
t.getSequenceNo() + " for "
                        + Constants.MIN_PROCESSTIME +
" milliseconds.");
                } catch (InterruptedException e) {
                    // TODO Auto-generated catch block
                    e.printStackTrace();
                }
            }

            // process VIP customers
            if (t.getTicketType().equals(TicketType.VIP)) {
customer.
                try {
                    // It will take 5 seconds to process VIP
                    Thread.sleep(Constants.MAX_PROCESSTIME);

                    System.out.println(Thread.currentThread().getName()
                        + " process " +
t.getSequenceNo() + " for "
                        + Constants.MAX_PROCESSTIME +
" milliseconds.");

```

```

        } catch (InterruptedException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }
    }
    System.out.println(Thread.currentThread().getName()
        + " finish processing ticket:" +
t.getSequenceNo()
        + ". Ticket type:" + t.getTicketType());
    }
}

public void setType(TicketType wType) {
    this.wType = wType;
}

public TicketType getType() {
    return wType;
}

@Override
public void run() {
    while (true) {
        Ticket t = null;
        switch (wType) {
            case COMMON:
                t = tm.fetchTicketWithType(TicketType.COMMON); //
                // customer
                break;
            case EXPRESS:
                t = tm.fetchTicketWithType(TicketType.EXPRESS); //
                // express
                // customer
                if (null == t) // if there is no express customer,
                // customer
                t =
                tm.fetchTicketWithType(TicketType.COMMON);
                break;
            case VIP:
                t = tm.fetchTicketWithType(TicketType.VIP); //
                // customer
                if (null == t) // if there is no VIP customer,
                // customer
                t =
                tm.fetchTicketWithType(TicketType.COMMON);
                break;
        }
        if (null == t) { // If there is no customer waiting at
            all, wait a
                // second and check again

```

```

after waiting.

    System.out.println(Thread.currentThread().getName()
                        + ": no customers waiting now. Wait
another second.");
        try {
            Thread.sleep(1000);
        } catch (InterruptedException e) {
            e.printStackTrace();
        }
    } else
        processCustomer(t);
    }
}

```

5.2.4. Ticket

Η κλάση Ticket μοντελοποιεί την έννοια του εισιτηρίου. Κάθε εισιτήριο έχει έναν αύξοντα αριθμό και έναν τύπο που μπορεί να είναι κανονικός, εξπρές και VIP.

```

package org.epdo.sim;

public class Ticket {
    private int sequenceNo;
    private TicketType ticketType;

    public void setSequenceNo(int no) {
        this.sequenceNo = no;
    }

    public int getSequenceNo() {
        return this.sequenceNo;
    }

    public void setTicketType(TicketType t) {
        this.ticketType = t;
    }

    public TicketType getTicketType() {
        return this.ticketType;
    }
}

```

5.2.5. Ticket machine

Η κλάση αυτή είναι η μηχανή παραγωγής εισιτηρίων της τράπεζας. Κάθε φορά που καταφθάνει στην τράπεζα ένας πελάτης του χορηγεί ένα εισιτήριο ανάλογα με τον τύπο του. Όταν ο πελάτης εξυπηρετείται του αφαιρείται το εισιτήριο.

```

package org.epdo.sim;

```

```

import java.util. ArrayList;

public class TicketMachine {
    private static int sequence;
    private static ArrayList<Ticket> commonTickets = new
ArrayList<Ticket>(); // common

// customer
// waiting
// queue

    private static ArrayList<Ticket> vipTickets = new
ArrayList<Ticket>(); // vip

// customer
// waiting
// queue

    private static ArrayList<Ticket> expressTickets = new
ArrayList<Ticket>(); // express

// customer
// waiting
// queue

    // Use a private static class to load this TicketMachine class.
    private static final class A {
        public static final TicketMachine t = new TicketMachine();
    }

    private TicketMachine() {
        System.out.println("creating the ticket machine class");
    }

    // When customer come in, generate a ticket for the customer. Must
be
    // synchronized to avoid thread conflicts.
    public synchronized Ticket generateTicket(TicketType type) {
        Ticket t = new Ticket();
        t.setSequenceNo(sequence++);
        t.setTicketType(type);
        switch (type) {
            case COMMON:
                commonTickets.add(t);
                break;
            case EXPRESS:
                expressTickets.add(t);
                break;
            case VIP:
                vipTickets.add(t);
                break;
        }
        return t;
    }
}

```

```

// When customer is called, return the first available ticket and
delete it
// from the waiting queue.
public synchronized Ticket fetchTicketWithType(TicketType type) {
    switch (type) {
        case COMMON:
            if (commonTickets.size() > 0)
                return commonTickets.remove(0);
            break;
        case EXPRESS:
            if (expressTickets.size() > 0)
                return expressTickets.remove(0);
            break;
        case VIP:
            if (vipTickets.size() > 0)
                return vipTickets.remove(0);
            break;
    }
    return null;
}

// load the current singleton class
public static TicketMachine getInstance() {
    return A.t;
}
}

```

5.2.5. TicketType

Το TicketType είναι μια απαρίθμηση (enumeration) που παρουσιάζει τους τύπους των εισιτηρίων.

```

package org.epdo.sim;

public enum TicketType {
    COMMON, VIP, EXPRESS;
    public String toString() {
        String name = null;
        switch (this) {
            case COMMON:
                name = "common";
                break;
            case VIP:
                name = "VIP";
                break;
            case EXPRESS:
                name = "express";
                break;
        }
        return name;
    }
}
}

```

Παρακάτω, βλέπουμε την ενδεικτική έξοδο που δίνει το πρόγραμμα μετά την εκτέλεσή του με το λογισμικό Netbeans IDE.

- Αρχικά δεν υπάρχουν πελάτες οπότε τα νήματα/threads εκτυπώνουν το μήνυμα: ("Όνομα νήματος": no customers waiting now. Wait another second.)
- Όταν εισέλθει στην ουρά αναμονής ένας πελάτης, τότε εκτυπώνεται από την κλάση BankQueue.java το μήνυμα: ("Τύπος πελάτη" "αριθμός πελάτη" is waiting.).
- Στη συνέχεια, γίνεται ενεργοποίηση του κατάλληλου thread (ανάλογα με τον τύπο του πελάτη) και ξεκινά η εξυπηρέτησή του για ένα συγκεκριμένο χρονικό διάστημα. Έτσι εκτυπώνεται από το εκάστοτε thread το μήνυμα: ("Όνομα νήματος" process ticket:"Αριθμός εισιτηρίου". Ticket type: "Τύπος εισιτηρίου".).
- Στη συνέχεια, εκτυπώνεται ο χρόνος που χρειάστηκε το thread για να εξυπηρετήσει τον πελάτη και εκτυπώνεται το μήνυμα: ("Όνομα νήματος" process "Αριθμός εισιτηρίου" for "Χρόνος εξυπηρέτησης" milliseconds.).
- Τελικά, αφού ολοκληρωθεί η εξυπηρέτηση του πελάτη από το συγκεκριμένο νήμα, εκτυπώνεται το μήνυμα: ("Όνομα νήματος" finish processing ticket:"Αριθμός εισιτηρίου". Ticket type: "Τύπος εισιτηρίου".).

```
creating the ticket machine class
Normal TD1: no customers waiting now. Wait another second.
Normal TD4: no customers waiting now. Wait another second.
express TD: no customers waiting now. Wait another second.
Normal TD2: no customers waiting now. Wait another second.
Normal TD3: no customers waiting now. Wait another second.
VIP TD: no customers waiting now. Wait another second.
Common customer 0 is waiting.
Express customer 1 is waiting.
VIP customer 2 is waiting.
Normal TD3: no customers waiting now. Wait another second.
Common customer 3 is waiting.
Normal TD2: no customers waiting now. Wait another second.
express TDstart to process ticket:1. Ticket type:express
VIP TDstart to process ticket:2. Ticket type:VIP
Normal TD1start to process ticket:3. Ticket type:common
Normal TD4start to process ticket:0. Ticket type:common
Normal TD1 process 3 for 1137 milliseconds.
Normal TD4 process 0 for 1367 milliseconds.
Common customer 4 is waiting.
express TD process 1 for 1000 milliseconds.
Normal TD3start to process ticket:4. Ticket type:common
Normal TD2: no customers waiting now. Wait another second.
Normal TD3 process 4 for 1604 milliseconds.
express TD finish processing ticket:1. Ticket type:express
express TD: no customers waiting now. Wait another second.
Normal TD1 finish processing ticket:3. Ticket type:common
Normal TD1: no customers waiting now. Wait another second.
Normal TD4 finish processing ticket:0. Ticket type:common
Normal TD4: no customers waiting now. Wait another second.
Common customer 5 is waiting.
Express customer 6 is waiting.
```


Κεφάλαιο 6: Συμπεράσματα

Η θεωρία των συστημάτων εξυπηρέτησης εφαρμόζεται σε πολλούς και ποικίλους τομείς των επιστημών αλλά και της καθημερινότητας. Οι υποθέσεις όμως της κλασσικής θεωρίας αναμονής μπορεί να είναι πολύ δεσμευτικές για να είναι ικανές να μοντελοποιήσουν καταστάσεις της πραγματικής ζωής με μεγάλη ακρίβεια. Η πολυπλοκότητα των γραμμών παραγωγής και χαρακτηριστικά ειδικά για κάθε προϊόν δεν μπορεί να αντιμετωπισθεί με τα απλά μοντέλα που αναλύθηκαν στην παρούσα εργασία. Για το λόγο αυτό έχουν αναπτυχθεί ειδικευμένα εργαλεία για την προσομοίωση, την ανάλυση, την οπτικοποίηση και τη βελτιστοποίηση της χρονοδυναμικής συμπεριφοράς των γραμμών αυτών.

Οι περιορισμοί της καθολικής εφαρμογής της θεωρίας αναμονής γίνονται εμφανείς αν εξετάσουμε αυτό το παράδειγμα: στα μαθηματικά μοντέλα υποθέτουμε συχνά άπειρο αριθμό πελατών, άπειρη χωρητικότητα ή χρόνους αφίξεων και εξυπηρέτησης χωρίς κανένα περιορισμό. Είναι προφανές ότι πρέπει να υπάρχουν περιορισμοί στην πραγματικότητα. Συχνά, παρά το γεγονός ότι υπάρχουν όρια και περιορισμοί, μπορούν να αγνοηθούν με ασφάλεια επειδή οι διαφορές ανάμεσα στον πραγματικό κόσμο και τα μοντέλα δεν είναι στατιστικώς σημαντικές, αφού η πιθανότητα να εμφανιστούν τέτοιες περιπτώσεις περιορισμών είναι πολύ χαμηλή σε σχέση με την αναμενόμενη συνήθη περίπτωση. Επιπρόσθετα, αρκετές έρευνες έχουν αποδείξει την ευρωστία των μοντέλων αναμονής εκτός των περιοριστικών τους υποθέσεων. Σε άλλες περιπτώσεις οι θεωρητικές λύσεις μπορεί να αποδειχθούν δισεπίλυτες ή ελλιπείς σε πληροφορίες.

Εναλλακτικά μέσα ανάλυσης χρησιμοποιούνται σήμερα για την παροχή διορατικότητας στο πρόβλημα που δεν αποκλείουν πολύ από τη σκοπιά της θεωρίας αναμονής. Οι διαφορές τους είναι ότι είναι συχνά ειδικευμένες στα σενάρια του κάθε προβλήματος διότι αποτελούνται γενικά από προσομοιώσεις μέσω υπολογιστών και αναλύσεις εμπειρικών δεδομένων.

Βιβλιογραφία

Ελληνική Βιβλιογραφία

- Γεωργίου, Α. (2011). *Θεωρία Γραμμών Αναμονής*. Θεσσαλονίκη: Πανεπιστήμιο Μακεδονίας.
- Δάρας, Τ. & Σύψας, Π. (2003). *Στοχαστικές Ανελίξεις, Θεωρία & Εφαρμογές*. Θεσσαλονίκη: Ζήτη.
- Ευσταθίου, Δ. & Τσίτσος, Σ. (2011). Δίκτυα Τηλεπικοινωνιών Μετάδοσης. Διαθέσιμο στο: ftp://teiser.gr/pliroforiki/DIKTIA_TILEPIKOINONION/Askiseis_2010_2011/Mathima_3.pdf, Πρόσβαση: 30/07/2013.
- Κούτρα, Μ. (2004). *Εισαγωγή στις Πιθανότητες. Θεωρία και Εφαρμογές*. Αθήνα Σταμούλης.
- Κουτρουβέλη, Ι.Α (1999). *Εφαρμοσμένες Πιθανότητες*. Πάτρα:Συμμετρία.
- Μάγκλαρης, Β., & Παπαβασιλείου, Σ. (2011). *Συστήματα Αναμονής*. Αθήνα: Εθνικό Μετσόβιο Πολυτεχνείο.
- Χρυσ αφίνου, Ο. (2004). *Εισαγωγή στις Στοχαστικές Ανελίξεις*. Θεσσαλονίκη: Σοφία

Ξένη Βιβλιογραφία

- Ford, J. L. (1980). *Two Applications of Queuing Theory*. Ithaca: Cornell University.
- Ivo Adan, J. R. (2001). *Queueing Theory*. Eindhoven, The Netherlands: Department of Mathematics and Computing Science, Eindhoven University of Technology.
- Morgan, H. (1999, Ιούνιος 14). *Andrew Ferrier*. Ανάκτηση από Applications Of Queueing Theory: http://www.andrewferrier.com/oldpages/queueing_theory/Henry/index.html
- Yue, W., Takahashi, Y., & Takagi, H. (2009). *Advances in Queueing Theory and Network Applications*. Kyoto: Springer.